

# Rapid Capture:

# Faster Throughput in Digitization of Special Collections

Ricky Erway

Senior Program Officer  
OCLC Research



A publication of OCLC Research

Rapid Capture: Faster Throughput in Digitization of Special Collections  
Ricky Erway, for OCLC Research

© 2011 OCLC Online Computer Library Center, Inc.

Reuse of this document is permitted as long as it is consistent with the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 (USA) license (CC-BY-NC-SA):

<http://creativecommons.org/licenses/by-nc-sa/3.0/>.

April 2011

OCLC Research

Dublin, Ohio 43017 USA

[www.oclc.org](http://www.oclc.org)

ISBN: 1-55653-398-5 (978-1-55653-398-3)

OCLC (WorldCat): 712406431

Please direct correspondence to:

Ricky Erway

Senior Program Officer

[erwayr@oclc.org](mailto:erwayr@oclc.org)

Suggested citation:

Erway, Ricky. 2011. *Rapid Capture: Faster Throughput in Digitization of Special Collections*.

Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2011/2011-04.pdf>.



## Table of Contents

Introduction .....	5
Archives of American Art, Smithsonian Institution .....	6
Indiana University .....	7
Library of Congress .....	9
The Bancroft Library, UC Berkeley .....	10
University of California, Los Angeles .....	11
University of Minnesota: Green Revolution .....	12
University of Minnesota: University Archives .....	14
University of North Carolina .....	15
The Walters Art Museum .....	16
So, what to make of this? .....	17
Appendix .....	20
References .....	23

## Introduction

Back in the day, we used to think digitizing pictures was easy; digitizing books was difficult. Books have parts and order and hierarchy—and text! Then Google (and Internet Archive, and Kirtas, and...) showed us how quick and easy digitizing books could be. Suddenly it seems as though we're all saying digitizing books is easy; digitizing everything else is difficult.

OCLC Research has been chipping away at the question of why digitizing special collections is so, well, special. How could it be made a bit more ordinary, routine, affordable, and effective? In 2007, we published *Shifting Gears* (Erway and Schaffner 2007), which took the position that for special collections we're going to preserve the original, so why be obsessed with preservation-quality digitization? Digitize for access! Better isn't more. More is better.

Last year, OCLC Research took on some of the rights issues related to special collections, with the *Undue Diligence* event (2010a) and the subsequent community of practice (2010) that has formed around the *Well-intentioned Practice* document (2010b), which argues that we should be less risk-averse. We shouldn't be the ones saying "No" to digitizing unpublished works. If the copyright holders come forward, give them a chance to say it. (They rarely do.)

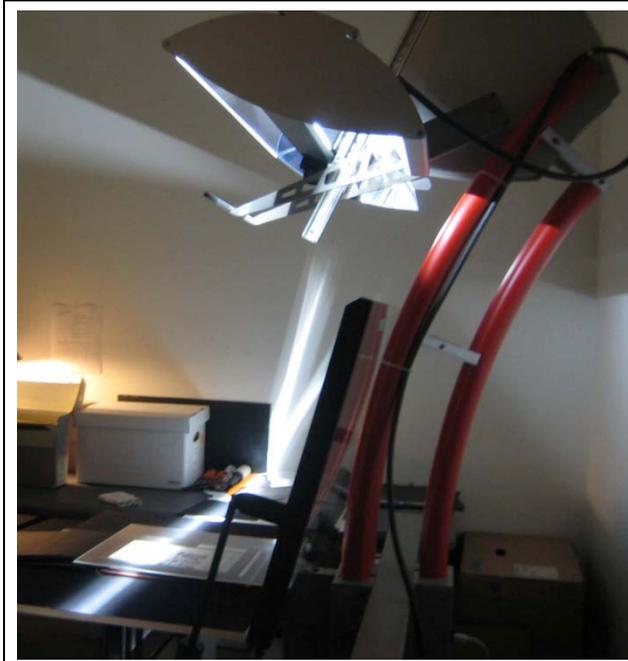
In between we have made other forays to help increase the scale of digitization of special collections. (See *Supply and Demand* [Erway 2008] on users' changing expectations, *Good Terms* [Kaufman and Ubois 2007] on commercial partnerships for digitization, *Capture and Release* [Miller 2010] on digital cameras in the reading room, and *Scan and Deliver* [Schaffner, Snyder, and Supple 2011] on patron-initiated digitization.)

In this study, we look at the actual digital capture part of the process. Google et al. figured out how to maximize throughput for book digitization; have there been any similar innovations for the digitization of other formats? In hardware? In workflow? In attitude? Our inquiring minds wanted to know.

So in an extremely casual survey, we asked some of our colleagues in libraries, archives, and museums to identify initiatives where non-book digitization was being done "at scale." We didn't define "at scale," because we thought we'd know it when we saw it. It wasn't always so easy. We heard from lots of places: some with minor achievements, others making a big difference. We selected a few of the latter stories to share with you here. Each one starts with a picture of the equipment in use at the site along with a sample from the collection being digitized. Additional details can be found in the appendix.

We hope these vignettes will encourage others to consider whether any of these approaches might be applied to their own projects to increase the scale of their digitization efforts. Ultimately, we hope these gleanings will contribute to increasing access to special collections.

## Archives of American Art, Smithsonian Institution



Series 4: Photographs and Scrapbooks.  
Reproductions of Artwork. Installation Views, 1946-1951. (Box 26, Folder 4, No. 46) Walt Kuhn Family papers and Armory Show records, 1859-1978. Archives of American Art, Smithsonian Institution.

The Archives of American Art is doing large-scale digitization of entire manuscript collections from originals and from microfilm, linking the resulting images to the finding aids that describe the collections. The collections contain manuscripts, photographs, slides, sketchbooks, scrapbooks, diaries, postcards, posters, and rare published materials. A 2005 grant from the Terra Foundation of American Art allowed them to carry this work forward. They have digitized 110 collections representing 1,015 linear feet of manuscript collections and 442 reels of microfilm—over 1.5 million images.

The Archives uses two Zeitschel overhead planetary scanners acquired from The Crowley Company. In late 2010, they received Phase One equipment from Digital Transitions, Inc. The Zeitschel takes 7-10 seconds to capture grayscale at 300 dpi. The Phase One takes approximately one second per scan in color. With the Phase One camera, they are now scanning everything in color and at a significantly faster speed. The Archives has developed a remarkable software system to track the process from beginning to end, coordinating 16 steps (processing, scanning, quality control, creation of derivatives, approvals, etc.) and linking the digital images to the finding aids.

**Outsourcing:** The Archives has outsourced some microfilm scanning to Crowley and some to Modern Microimaging. Capture from originals was done exclusively in house until 2010, when

Crowley began digitizing selected collections from originals. In the past 18 months Crowley completed 206 linear feet of originals. Since they can have more than one capture stream working on different sections of the collection, they can complete far more than can be scanned in-house.

**Throughput:** At the Archives, one operator can scan 1 ½-2 linear feet per week, sometimes as many as 500 images per day.

**Bottlenecks in other parts of the process:** Standard appraisal processes are inherently time-consuming. Processing takes about the same amount of time as capture.

*Of note: The Archives of American Art was able to attempt large-scale digitization due to its history of microfilming entire collections ever since its founding in 1954 and continuously up until 2005, when they replaced their microfilm operation with digital capture. Microfilming involved creating 7 sets of film for regional AAA offices and affiliates, required a painstaking quality-control process, making labels, supporting ILL—they've shown significant benefits in workflow and accessibility by switching to a digital workflow.*

## Indiana University



*Recording Formats. Kevin Atkins. Courtesy of Indiana University.*

The Archives of Traditional Music is exploring issues related to the digitization of audio for preservation and access. Through grant funding and ongoing conversion of various at-risk media, their research focuses on automating as much of the process as possible and working on fast, accurate means of transfer to digital form. While they work with many types of media, we focused on their work with audio recordings on cassette tapes.

Indiana University has undertaken research to develop safe practices for parallel transfer, which is the digitization of multiple recordings simultaneously. They can process three cassette tapes at the same time. They have two banks of three cassette decks and, as the first set is being transferred, the second set is being prepared. An audio engineer monitors the three decks via software that automatically brings each of the three sources into primary focus (higher volume by 12-15 db) for a short amount of time (10-15 seconds). The technician

can always hear all three recordings, so has continuity, but focuses on one at a time. This approach allows comparison from one source to another (batches consist of similar materials), so the technician can make constant comparisons to listen for subtle shifts in fidelity.

**Throughput:** Although audio must be digitized at playback speed, the three-source approach is *more* than three times more efficient than doing one at a time. They have also automated much of the post-processing activity to run in overnight batches so that metadata is embedded in files and entered into a database, derivatives are created, and files are copied and checksums verified by software applications.

**Outsourcing:** While Indiana University is fortunate to have a wide range of in-house expertise and equipment, most sound archives do not. It may be more efficient to outsource to an experienced service provider with expert engineers, skilled technicians, and an array of software and hardware devices. Service providers have streamlined workflows, can handle multiple streams, and are often far less expensive than establishing in-house capabilities.

**Bottlenecks in the capture process:** An important part of the process is selection. They have to select stable, suitable materials for parallel transfer, while directing unsuitable recordings to a custom 1:1 transfer process. If there's a problem with one of the tapes, they have to stop and put it in the cue for special handling and investigation.

*Of note: Unlike other types of materials, when it comes to sound (and video) recording, there are no long-lived analog formats. Everything must be converted to digital for preservation and access. Content that is on obsolete media or that is degrading is especially at risk.*

## Library of Congress



*"Fiddlin" Bill Henseley. Asheville, North Carolina. Photographer: Ben Shahn. Library of Congress, Prints & Photographs Division, FSA-OWI Collection, [LC-USF33-6258-M1].*

The Library of Congress is digitizing 23,000 Farm Security Administration 35mm nitrate film negatives. The work was outsourced to CSC, Inc. who used equipment from Stokes Imaging, Inc. The equipment consisted of a digital camera, transmissive lighting, a table top sufficiently anchored to safely handle materials and limit vibration, a negative carrier adapted to hold strips of film flat (the film is manually loaded but the carrier is automatically controlled), and an enclosed workspace with filtered air to reduce dust.

**Throughput:** Three staff (one production manager and two technicians), work eight-hour workdays and capture about 500 frames per day.

**Outsourcing:** Digital conversion efforts are entrusted to experts who can focus solely on that task. Vendors can quickly upgrade or change technology as requirements change from project to project, or as capture technologies rapidly improve.

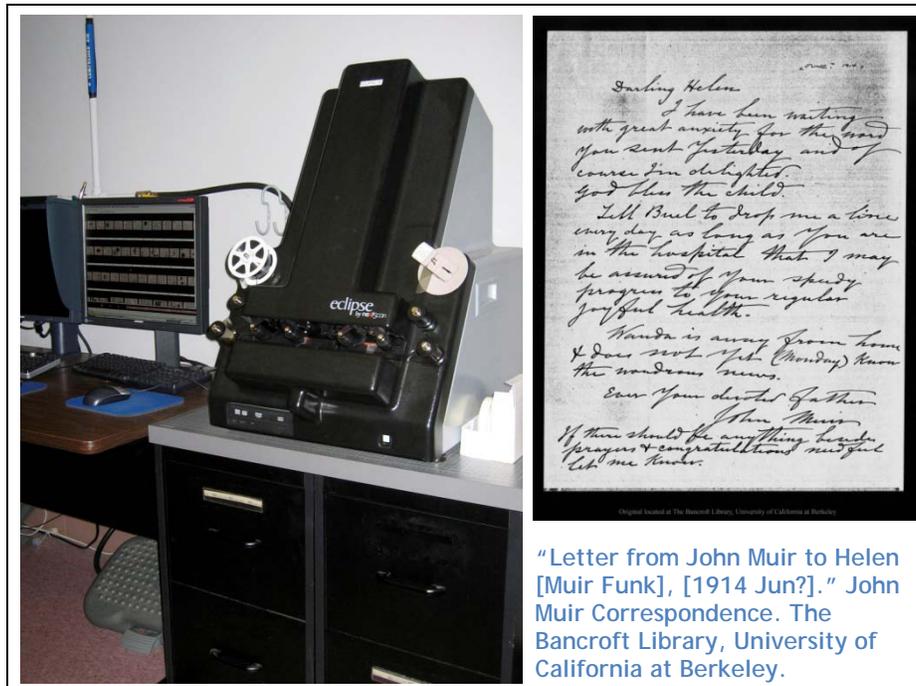
**Bottlenecks in the capture process:** The startup phase is the most time-consuming as technical specifications are confirmed and aesthetic expectations are honed. Some problems occur with film curl, film stuck to housing, and mismatches between collection inventory and materials in hand.

**Bottlenecks in other parts of the process:** Handling physical media deliveries, post-scan gamma/tonality adjustments, and delivering large file sizes and quantities all take time. Despite mechanical and technical conveniences, a good eye for quality, an understanding of how photographs should "look," and overall visual literacy skills are required. Setting and calibrating aesthetic quality expectations can be difficult and slow, but are worth the up-front investment.

*Of note: Concurrent with a process to make preservation film copies of the deteriorating negatives, these photos were originally digitized in the early 1990s as a by-product of*

creating a videodisc. The low-resolution images have proved so popular as to warrant re-scanning at far superior quality.

## The Bancroft Library, UC Berkeley



"Letter from John Muir to Helen [Muir Funk], [1914 Jun?]." John Muir Correspondence. The Bancroft Library, University of California at Berkeley.

The Bancroft Library and the University of the Pacific Library's John Muir Correspondence project, funded by The Library Services and Technology Act, outsourced digitization of a 1985 microfilm set from high quality, duplicate positive microfilm made from the original negatives. The 22 reels (containing 6,500 letters) were scanned by OCLC Preservation Services (since acquired by Backstage Library Works) resulting in 24,800 total image files. The vendor used two NextScan Eclipse scanners. A technician made the initial settings for each reel to get the best output per reel. Once a test scan was accepted, the frames on the reel were scanned automatically. After image capture, images were checked in the NextScan Audit software and then auto-cropped and de-skewed in Photoshop.

**Throughput:** Approximately six 100 foot reels per day per scanner—about 4800 images per day.

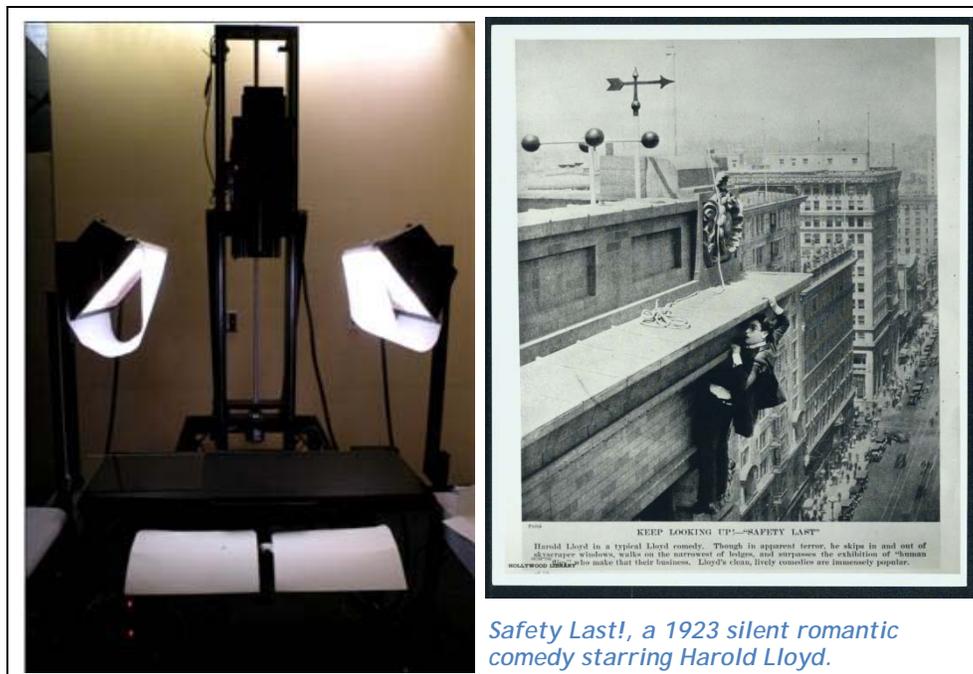
**Outsourcing:** Microfilm scanning can usually be done most cost-effectively using a vendor. Reels are not original materials, so can be shipped off-site without the high cost (insurance, specialized shipping) of sending original materials. The microfilmed materials can be scanned for a fraction of the cost of doing the work in-house from the originals.

**Bottlenecks in the capture process:** Some of the more time-consuming tasks included determining how to handle all the varied “assets” on the film (metadata and targets), filenaming protocols, and dealing with unsprocketed film and inconsistent frame sizes.

**Bottlenecks in other parts of the process:** It takes time to duplicate the microfilm, clear rights and transcribe text.

*Of note: Anytime scanner throughput can be increased, costs are reduced. Doing work in quantity and grouping materials by size so that equipment adjustments are not necessary reduces the overall cost of capture. The Bancroft Library has successfully reduced costs on high-end capture by 80% and more using this methodology.*

## University of California, Los Angeles



The Collection of Motion Picture Stills consists primarily of American motion picture stills dating from the 1920s to the present. The collection includes black-and-white prints and a small amount of color prints and slides. The photographs include set and scene shots as well as publicity stills. The collection represents films produced by various companies including Columbia Pictures, 20th Century Fox, MGM, Paramount, Republic, RKO Universal, United Artists, Warner Brothers and more. The stills are well-used and are stored off site; digitization will increase access to the collection.

Stokes Imaging installed a newly designed scanning station at the Southern Regional Library Facility (SRLF) for testing and evaluation as a mass-digitization option. Plans and preparation began in 2008/09. With funding from the UCLA Library, two new workstations for image

capture and processing were provided at the SRLF, along with expanded storage capacity, 1.25 FTE staff, and support from UCLA's Digital Library and Cataloging/Metadata staff. In addition to digitizing the Motion Picture Stills Collection, project staff have also scanned the Aldous Huxley Collection and will continue through 2011 with additional photo and manuscript collections.

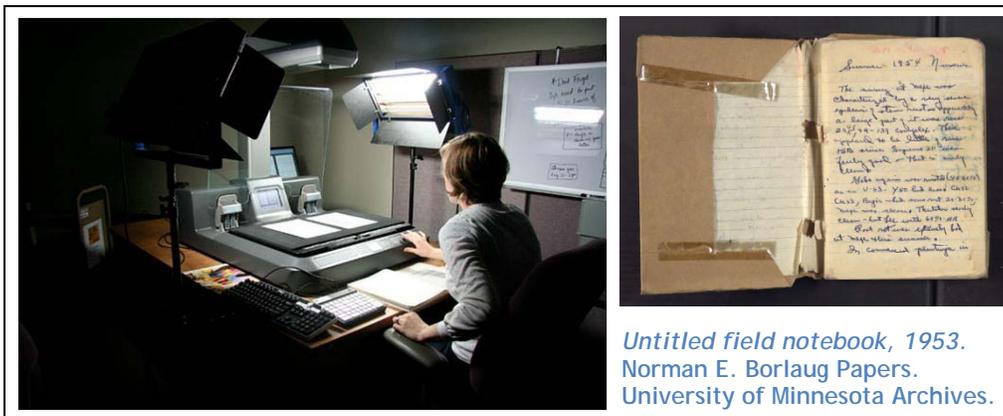
**Throughput:** About 400 images per day, with 5-6 hours of scanning and 2-3 hours of processing. Image resolution ranges from 300-2,400 ppi. Throughput rate can be as high as 165 images per hour at lower resolution.

**Bottlenecks in the capture process:** The new setup had a lot of startup delays (installation, training, and several months of working with the new Stokes Imaging Workflow Solution software). Photographs that are old and poorly preserved can be problematic and glossy photos can cause glare in images.

**Bottlenecks in other parts of the process:** Moving digital files (physical media delivery and data storage on server directory) causes some delays.

*Of note: UCLA has now purchased the Stokes Imaging equipment and in 2011 may add a second operator to improve/increase productivity.*

## University of Minnesota: Green Revolution



*Untitled field notebook, 1953.  
Norman E. Borlaug Papers.  
University of Minnesota Archives.*

In the eleven-month, grant-funded, Green Revolution Scanning Project, Archives and Special Collections estimates it will scan 198,350 selected documents and photographs (notebooks, loose sheets, photo albums, news clippings) from personal and alumni papers and departmental records. Documents that are under copyright, such as publications, and materials that are routine in nature, such as invitations and cards, were not scanned.

The Green Project uses an i2S CopiBook HD (from French Firm i2S Innovative Imaging Solutions; US vendor Image Retrieval) and two free-standing fluorescent lights. Four students do the bulk of the scanning in two daily four-hour shifts. Before scanning, the documents are

prepped by student staff, removing paperclips and staples. The students scanning the collections also prep the documents.

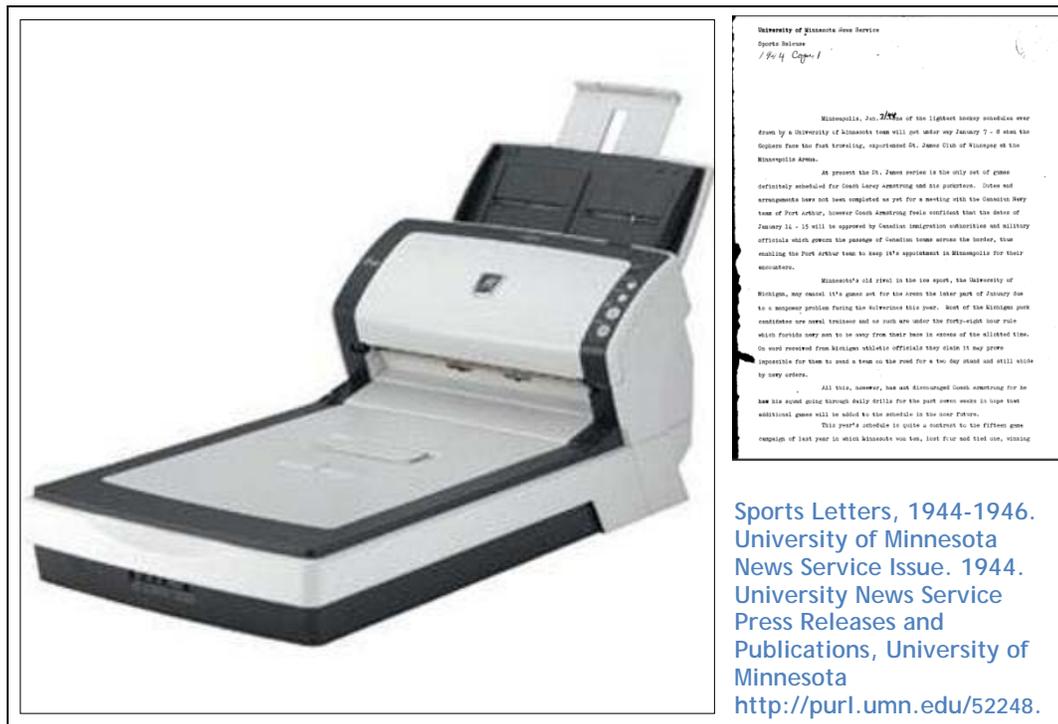
**Throughput:** They average 1536 scans per day.

**Bottlenecks in the capture process:** Actual image capture (i.e., not including any post-scanning quality control and processing) remains the fastest part of the overall process.

**Bottlenecks in other parts of the process:** Quality control continues to provide the largest bottleneck. In order to ensure fidelity of capture and to avoid any missed page scans the sampling size is incredibly large. Other bottlenecks currently are batch upload into the repository and discovery environment. They are working to effectively integrate a batch loader into the system, but even so, loads at this scale are time consuming. They would like to better leverage technology to help shepherd image files through the system. Currently, a number of human interactions need to take place to trigger certain events: JPEG 2000 conversion, OCR, batch upload. In the future they would like to incorporate an improved process that takes care of these hand-offs.

*Of note: They are not creating metadata for the scans, beyond the file names created as part of the process. They are using the folder titles (and in some cases box titles) as the metadata so the existing finding aids link directly to the scanned images.*

## University of Minnesota: University Archives



Sports Letters, 1944-1946.  
University of Minnesota  
News Service Issue. 1944.  
University News Service  
Press Releases and  
Publications, University of  
Minnesota  
<http://purl.umn.edu/52248>.

The University Archives at the University of Minnesota is digitizing legacy university publications. This began as a project and is now a routine function of the University Archives. The selection criteria for scanning focuses predominantly on high-use content or items held in duplicate. The duplicates are primarily bound paper documents and the first step is disbinding. Most of the content scanned had never been cataloged and was therefore completely inaccessible. It's now available online with full-text searching, and discoverable through Google. Use statistics are high. The digital versions are all in standard formats with reasonable preservation metadata, so their normal backup procedures should provide adequate digital preservation. In the most recent year, they have scanned 219,074 pages on a single scanner, a \$1,300 Fujitsu fi-6230 Color Duplex Flatbed with sheet feeder, incorporated into the duties of student workers.

**Throughput:** All scanning work is done by student workers, working an average total of 15 hours per week. They average 500 pages (scan and OCR) per hour.

**Bottlenecks in the capture process:** They've routinized the scanning component to the point where it needs attention to detail but very little decision-making. They get stalled occasionally when something is not what it seems—e.g., bound volumes that turn out to be incomplete.

**Bottlenecks in other parts of the process:** The most time consuming step is working with content that is not routine. Bottlenecks also occur sometimes in the delivery process.

*Of note: The UM Archives is unusual in that they are, for the most part, scanning and discarding this material. Duplicates are scanned and then recycled. Content for which there are no duplicates is usually recycled as well—an example is the dozens of volumes of bound press releases. This content has informational, but no artifactual, value. The 219,074 pages scanned last year resulted in 681 bound volumes being discarded, freeing up hundreds of feet of storage space. Only a very small percentage of content scanned is then foldered and added to the archival and manuscript collections. While there is some risk in destroying originals, there are huge benefits; these include improved access and recovering space for incoming material.*

## University of North Carolina



The Digital Southern Historical Collection currently consists of 285 collections and 140,000 images. Established with Mellon funding, it is now a sustainable on-going program aiming to digitize most of the 16 million items in 4,600 archival collections, primarily manuscript materials on paper. While the selection of collections for digitization currently focuses on small collections related to race relations, about half of what has been digitized to date was done in response to special requests. The digitization is done in-house in the University Library's Digital Production Center.

The equipment consists of two Zeutschel overhead planetary scanners for the majority of traditional paper manuscript materials, a Fujitsu scanner for collection materials that can be placed in the high-speed sheet feeder (e.g., note cards, transcriptions on sturdy paper stock), a Phase One camera for oversize materials such as maps, and flatbed scanners for photographic materials.

**Throughput:** It is difficult to come up with a daily throughput. There are 70 hours per week of graduate student help, but they do many aspects of the process beyond capture

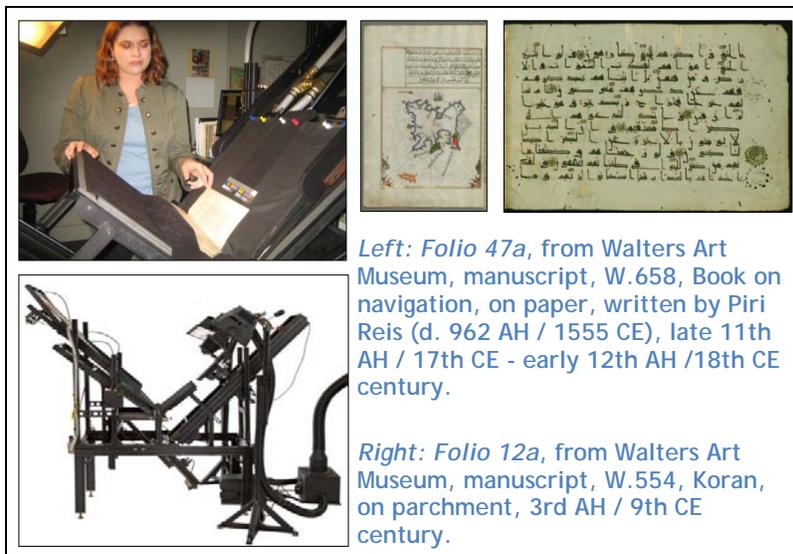
(conversion, metadata, and uploads). Most work four-hour shifts. They can typically scan 90 images per hour.

**Bottlenecks in the capture process:** The condition and complexity of materials cause the biggest bottlenecks in the content capture. Fragile materials naturally require more cautious handling and thus slower capture. Complex documents such as scrapbooks demand careful attention, as well as multiple image captures for each page with foldouts and layers.

**Bottlenecks in other parts of the process:** Time-consuming aspects include file conversion of the digital objects and finalizing a job, digital storage and server constraints, and finding aid revision.

*Of note: A significant feature of this program is the care with which filenames are determined in advance; metadata is extracted from the finding aid to form folder-level metadata to describe all the scans for that folder. The finding aid provides description and enables discovery and links to the images.*

## The Walters Art Museum



*Left: Folio 47a, from Walters Art Museum, manuscript, W.658, Book on navigation, on paper, written by Piri Reis (d. 962 AH / 1555 CE), late 11th AH / 17th CE - early 12th AH / 18th CE century.*

*Right: Folio 12a, from Walters Art Museum, manuscript, W.554, Koran, on parchment, 3rd AH / 9th CE century.*

The Walters is digitizing nearly 100,000 pages of early Islamic and western manuscripts. They are bound, but precious and fragile—not a collection for the Google book scanning approach. With NEH funding, they acquired an imaging apparatus (which they call “Omar”) from Stokes Imaging, Inc. Omar consists of a camera, a stand, an automated digitization cradle apparatus with a vacuum wedge, a stationary board for single leaves; two light stands with five 17-degree tungsten bulbs each, and software to manage the workflow, do automated color-correction, and create derivatives. Because the images are taken from a constant focal plane, the aperture and focus seldom need to be adjusted, so productivity is fairly high. The capture station is in

use 8-10 hours a day. Two full-time staff do all the prep, the capture, and the quality control. The capture takes about the same time as the other processes.

**Throughput:** On smaller books they average 275 images a day, but on larger books it's more like 150 a day. They captured 350 images on their most productive day.

**Bottlenecks in other parts of the process:** Copying back-up files and storing the images takes a surprising amount of time and attention.

*Of note: The Walters is selfless when it comes to making the files available. Though they've put some images up on Flickr and have some of the manuscripts on their Web site, instead of creating a user interface to showcase their collections, they set the files free (under Creative Commons attribution license) in the most reusable form. They make them easily available to those who wish to study them and encourage experts to create sites for the manuscripts, providing transcription and scholarly context.*

## So, what to make of this?

Stanford's Hannah Frost (2008) has described the urge to speed content to the public as determination "to make media more accessible... at any cost, even a low one." That could apply to most of the formats mentioned in this document, but collections vary, and it is important to recognize that comparing throughput rates across the vignettes must take other factors into account. Different materials require different approaches, from the careful handling of precious bound illuminated manuscripts to the sheet feeding of documents that are to be discarded after scanning. Additionally, for some of these projects 300 ppi was adequate resolution; others required as much as 2500 ppi. It can take significantly longer to capture images at higher resolution.

Audiovisual materials have some additional issues. While audio and video *must* be transferred to digital form for preservation purposes and so should be captured at the highest quality feasible, film is quite the opposite. Typically because of the otherwise enormous file sizes, quality is compromised when film is digitized and compressed, so the digital copy is thought of as an access copy and the original film is the preservation copy. Even when the original is on volatile nitrate film, the film is copied onto safety film stock for preservation.

One of the reasons that it is difficult to digitize special collections at the same scale as routine book scanning is that special collections are often heterogeneous in nature. What makes a capture operation efficient is the ability to streamline workflows by setting up equipment and workflows for one set of characteristics and then capturing a mass of similar items, thereby limiting the adjustments done in between captures.

While not all special collections are amenable to capture at scale, many are. If collections of physically uniform materials are chosen—and where reasonable, if uniform materials from one or more collections are grouped for digital capture—some semblance of scale is possible. In

addition to the microfilm, documents, and 8x10 photos in the case studies, other formats such as slides or postcards lend themselves to faster throughput. Every bit of economy that can be achieved on those materials will allow resources to be dedicated to the more deliberate approaches required by more complex collections.

Not all innovations in digitizing special collections are technologically sophisticated; some merely take advantage of what is readily at hand, as illustrated in this Princeton pilot project.

*Using student workers and two photocopiers, the Mudd Manuscript Library at Princeton, as part of a pilot project, has scanned over 20 linear feet of typed and handwritten 20<sup>th</sup> century manuscripts totaling approximately 70,000 images.*

*When students have time between processing or other projects, they take the next box from the list of prioritized collections, go to the photocopiers, and scan folder by folder, using the flat bed or the sheetfeeder as appropriate. Students apply a standard library item barcode to each folder. The barcode serves as a unique identifier and as the file name for each PDF. The resulting PDF file for each folder is stored in Princeton's document management system. A macro script adds the link to the PDF at the appropriate place in the EAD finding aid. Using this approach, they can scan 300 pages per hour, helping Mudd to make more content accessible, using existing metadata and making effective and efficient use of student help.*

*Dan Santamaria, Assistant University Archivist for Technical Services, says, "Where we really save is on training, since we use undergrads and equipment that they are already familiar with. We think this has a much greater impact than other routine tasks that we've typically had students perform."*

*There's been some debate within the University Library about the quality of the scans and about the reliance on description in finding aids (typically at the folder level) rather than item-level description. While some wish for detailed structural and descriptive metadata to go with the digitized documents, and a more selective approach to digitization, the content is made accessible in accordance with archival standards and practices. Santamaria views the work as an iterative process; as more is learned about how the content is used, decisions can be made about where additional metadata or higher quality imaging might be beneficial.*

The University of Montana's Steve McCann has asked, "What's the most efficient method for nearly choking a small digitization shop? Success. That's something we're experiencing . . . and we couldn't be happier" (2009).

One way to become happily choked by success is to establish from the very beginning that the goal is to maximize access—and plan accordingly.

*The Benson Ford Research Center, at The Henry Ford in Dearborn, Michigan, has begun an institution-wide digitization initiative with the vision of capturing and making accessible digital representations of all of their holdings. Key to the success of this Initiative will be*

*the use of a Rapid Capture Process (RCP) for quickly digitizing archival materials. However, Ford staff have found that RCP is straightforward in comparison to the issues involved in integrating the RCP with other workflows, systems, and groups within the institution:*

- *Determining the type, amount, and quality of metadata to be generated for the RCP*
- *Harmonizing the RCP file format, metadata, and image quality output with that coming from three existing imaging streams*
- *Understanding how the collection management system will ingest large volumes of image metadata, and then handle that data within an archival hierarchy*
- *Development of a finding aid presentation method to provide context for images associated with archival collections*
- *Providing file storage and structure that can accommodate the RCP production rate*
- *Creation of a digital “hub” that can deliver images to a variety of contexts, including museum interactives, archival finding aids, and Flickr-like image displays*
- *Education of museum and IT staff on archival terminology, theory and practice*

*The issue of metadata is proving difficult due in large part to their Library-Archive-Museum environment, which requires them to meet archival needs as well as those of the Museum registrars. For example, while the Archive is satisfied with providing an image identifier only, relying on the finding aid to provide context, the Museum requires an image caption at a minimum for both context and search. Expectations of external users with regard to both the amount and the perceived quality of metadata have also been a point of discussion.*

*Brian Wilson, digital processing archivist, noted, “While detail of workflow and equipment made available by other institutions can greatly facilitate RCP adoption, the experience of The Henry Ford indicates that fully integrating the process will require addressing issues that extend beyond the basic operation of the capture workstation.”*

True enough. We can increase the throughput rate—and as digitization of special collections becomes more commonplace, the costs of equipment and services will continue to decrease even as technological improvements speed up hardware and software operations. But speeding up the capture process is only one part of the picture. These significant advances on the capture side will increase pressure to make changes in other parts of the process. And that’s a good thing. Let’s continue to share our stories. We are on the verge of very significant advancements in making special collections more accessible.

## Appendix

Smithsonian Institution	
Unit	Archives of American Art
Name	Weiss, Karen
Title	Information Systems Manager
Email	<a href="mailto:Weissk@si.edu">Weissk@si.edu</a>
Activity	Ongoing digitization program
Format	Manuscript collections
Number	1015 linear ft. and 442 reels of microfilm—over 1.5 million images
Throughput	Up to 500 images/day
Capture staffing	One operator
Equipment	Zeuschel planetary scanners /recently added Phase One cameras
Vendor	The Crowley Company/Digital Transitions
Contact name	Pat Crowley/Peter Siegel
Email	<a href="mailto:pat@thecrowleycompany.com">pat@thecrowleycompany.com</a> <a href="mailto:pes@digitaltransitions.com">pes@digitaltransitions.com</a>

Indiana University	
Unit	Archives of Traditional Music
Name	Mike Casey
Title	Associate Director for Recording Services
Email	<a href="mailto:mcasey@indiana.edu">mcasey@indiana.edu</a>
Activity	Preservation reformatting
Format	Cassette audio tapes
Throughput	24 hours of audio/day
Capture staffing	One audio engineer
Equipment	3 cassette decks, Prism A/D converter, Lynx AES/16 sound card, Wavelab, ProTools

Library of Congress	
Unit	Prints & Photographs Division
Name	Phil Michel
Title	Digital Conversion Coordinator
Email	<a href="mailto:pmic@loc.gov">pmic@loc.gov</a>
Activity	Farm Security Administration Photographs
Format	Nitrate negatives
Number	23,000 frames
Throughput	500 images/day
Capture staffing	2 technicians
Equipment	Digital camera, transmissive lighting, negative carrier
Vendor	Stokes Imaging, Inc.
Contact name	John T. Stokes
Email	<a href="mailto:johnt@stokessoftware.com">johnt@stokessoftware.com</a>
Service provider	CSC, Inc.
Contact name	Jackie Kropman
Email	<a href="mailto:jkropman@csc.com">jkropman@csc.com</a>

The Bancroft Library, UC Berkeley	
Unit	Digital Collection Unit, Bancroft Library
Name	Mary Elings
Title	Head of Digital Collections
Email	melings@library.berkeley.edu
Activity	John Muir Correspondence
Format	Microfilm
Number	22 reels of microfilm (24,800 image files)
Throughput	4800 images/day
Capture staffing	2 stations with resources assigned as warranted
Equipment	2 NextScan Eclipse scanners, NextScan Audit software, and Photoshop
Service provider	Backstage Library Works
Contact name	Joan Dashiell
Email	dashiell@bslw.com

University of California, Los Angeles	
Unit	Southern Regional Library Facility
Name	Tin Tran
Title	Processing Manager
Email	tintran@library.ucla.edu
Activity	Motion Picture Stills
Format	Photographs
Number	90,000 images
Throughput	Average 400 images/day; up to 165/hour
Capture staffing	1.25 fulltime employees
Equipment	Camera, free standing workframe with lights and conveyor belt
Vendor	Stokes Imaging, Inc.
Contact name	John T. Stokes
Email	johnt@stokessoftware.com

University of Minnesota	
Unit	Archives and Special Collections
Name	Karen Spilman
Title	Processing Coordinator
Email	hoft0003@umn.edu
Activity	Green Revolution Scanning Project
Format	Documents
Number	198,350 pages
Throughput	1536 images/day
Capture staffing	Students 2 four-hour shifts/day
Equipment	i2S CopiBook HD, 2 free standing florescent lights
Vendor	IImage Retrieval Inc.
Contact name	Derek Jenkins
Email	sales@iiri.com

University of Minnesota	
Unit	University Archives
Name	Beth Kaplan
Title	Head, University Archives & Co-Director, University Digital Conservancy
Email	kapla024@umn.edu
Activity	Digitization of legacy university publications
Format	Documents
Number	219,074 pages
Throughput	Average 500 pages/hour
Capture staffing	Students 15 hours/week
Equipment	Fujitsu fi-6230 Color Duplex Flatbed scanner with sheet feeder

University of North Carolina	
Unit	Southern Historical Collection
Name	Laura Clark Brown
Title	Digital SHC Coordinator
Email	ljcb@email.unc.edu
Activity	Digital Southern Historical Collection
Format	Manuscripts and photos
Number	140,000 images
Throughput	90 images/hr
Capture staffing	Students 70 hours/week
Equipment	2 Zeutschel planetary scanners, Fujitsu scanner with sheet feeder, Phase One camera, flatbed scanners
Vendor	The Crowley Company
Email	hardware@thecrowleycompany.com

Walters Art Museum	
Unit	Department of Manuscripts and Rare Books
Name	Diane E. Bockrath
Title	Digitization Specialist
Email	dbockrath@thewalters.org
Activity	Islamic and western manuscripts
Format	Manuscripts (bound)
Number	100,000 images
Throughput	Average 275 images/day
Capture staffing	One fulltime scanner operator
Equipment	Camera with digitization cradle apparatus with vacuum wedge and stationary board for single leaves, 2 light stands
Vendor	Stokes Imaging, Inc.
Contact name	John T. Stokes
Email	johnt@stokessoftware.com

## References

- Erway, Ricky. 2008. "Supply and Demand: Special Collections and Digitisation." *Liber Quarterly* 18 (3/4). <http://liber.library.uu.nl/publish/articles/000263/index.html>.
- Erway, Ricky and Jennifer Schaffner. 2007. *Shifting Gears: Gearing up To Get into the Flow*. Dublin, OH: OCLC Programs and Research. <http://www.oclc.org/research/publications/library/2007/2007-02.pdf>.
- Frost, Hannah. 2008. "Reformatting Moving Images." Presentation given at the annual meeting of the American Library Association, Rare Books and Manuscripts Section, 29 June. [http://www.rbms.info/conferences/conf-docs/2008/Anaheim\\_Frost.ppt](http://www.rbms.info/conferences/conf-docs/2008/Anaheim_Frost.ppt).
- Kaufman, Peter and Jeff Ubois. 2007. "Good Terms—Improving Commercial-Noncommercial Partnerships for Mass Digitization: A Report Prepared by Intelligent Television for RLG Programs, OCLC Programs and Research." *D-Lib Magazine* 13 (11/12). <http://dlib.org/dlib/november07/kaufman/11kaufman.html>.
- McCann, Steve. 2009. "Large(ish)-Scale Digitization at the National Archives." *A Digital Outrigger* (blog), 5 October. <http://librarydigitalprojects.com/2009/10/05/largeish-scale-digitization-at-the-national-archives/>.
- Miller, Lisa, Steven K. Galbraith and the RLG Partnership Working Group on Streamlining Photography and Scanning. 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>.
- OCLC Research. 2010. "Support for Well-Intentioned Practice for Putting Digitized Collections of Unpublished Materials Online." <http://www.oclc.org/research/activities/rights/support.htm>.
- . 2010a. "Undue Diligence: Seeking Low-Risk Strategies for Making Collections of Unpublished Materials More Accessible." <http://www.oclc.org/research/events/2010-03-11.htm>.
- . 2010b. "Well-intentioned Practice for Putting Digitized Collections of Unpublished Materials Online." <http://www.oclc.org/research/activities/rights/practice.pdf>.
- 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>.