

# Archived Documentation for the UF Digital Library Center (from approx. 2012):

Average Times for Digitization Activities  
Average File Sizes  
Project Planning Resources

## Average Times for Digitization Activities

Below is a list of the component activities in digitization offered by the DLC with estimates of average times per component. All digitization complies with national standards.

See the average file sizes and project planning sections below for more resources for planning projects.

---

Estimating pages: "A cubic foot of records comprises about 2,000 pages."  
(<http://www.archives.gov/foia/ufos.html>). The average archive box is 5 inches.

---

## Calculating Costs

In consultation with the DLC, use this [spreadsheet to calculate costs](#).

- Labor: unless otherwise specified, labor is calculated for the salary and benefits of a Library Associate 2. (Current [fringe rates are linked here](#))
- Overhead: when applicable, added automatically on the workbook as shown on the "Totals\_All" sheet (and can be removed as applicable)
- Server Costs: server costs are calculated per annual web and archival costs.

Abstracted, simplified chart, assuming other supports in place, based on the example projects detailed below:

Bound books	7 hours
Disbound books	2 hours
archival/photos	11 pages / hour
large format	2.5 hours
born digital	50 pages / hour

print newspapers	40 pages / hour
vended digitization, newspapers on microfilm, NDNP-compliant	210 pages / hour
vended digitization, newspapers on microfilm, non-compliant	29 pages / hour
Oral history files, 30 minutes; born digital; with PDF transcript	1 set (audio and PDF) / hour

*Bound book:* assumes average of 200 pages; however, cost is based on volume; average is 4-11 hours, assuming average of 7 hours

*Disbound book:* cost is based on item size; dissertations, theses; anything of at least 200 pages that can go through the high speed scanner

*Archival/photographs:* all print photographs that are not oversized; aerials, regular photographs; manuscripts and archival materials where the physical collections have already been processed

*Large format:* times are similar for A/V items

*Print newspapers:* for broadsheet newspapers that must be cut

*Born digital:* includes ingest of vendor materials; harvest and processing UF serials; FTP receipt/harvest and processing newspapers; partner CD/DVDs

*Other formats:* other formats may require specialized staff skills and should be estimated based on actual materials. For instance, materials in the round require skilled staff time for set up, capture, and post processing (a minimum of 8 hours for hat-sized and smaller objects) with additional time required for travel and set up for imaging conducted off site and for larger objects.

#### Material Type by Equipment and File Size

Material Type	Unit	Equipment	File Size in TB	Total Hours
Bound books	1 book	CopiBook	0.00579357147 (20.25MB/page as average of rgb/bw; 300 pages)	7
Disbound books	1 book	HighSpeed	0.00579357147	2
archival/photos	1 page	Flatbed	0.0000193119049072265625	0.09
large format	1 page	large format	0.0002384185791015625	2.5
print newspapers	1 page	CopiBook	0.00006103515625	0.025
vended digitization, newspapers on microfilm, NDNP-compliant	1 reel (1,000 pages)	workstation-only	0.00006103515625	5
vended digitization, newspapers on microfilm, non-compliant	1 reel (1,000 pages)	workstation-only	0.00006103515625	34.5

#### Example Projects:

Overview	Details
Catalog records created by	DLC handled digitization (imaging, image processing, QC with structural

<p>Cataloging</p> <p>Copyright status already known to be public domain</p> <p>Physical material prep. and post-proc. by Preservation</p> <p>DLC digitization total average time for a 200 page book: <b>4 1/6 - 11 hours</b></p>	<p>metadata, OCR, loading, and archiving) for 2,500 books over 2 years, or 1,250 books per year. For each of the two grant years, dedicated staff time for cost share in the DLC: 2.15 FTE</p> <p>Total of 2,500 volumes, or 500,000 pages over a two year period 1,250 volumes per year 6,000 pages/week; approximately 30 volumes of 200 pages each</p> <p>Scanning &amp; Initial image processing (deskew, crop)</p> <ul style="list-style-type: none"> <li>• Kodak DCS 24n megapixel DSLR camera: 3 min/page x 200 pg = 600 min/60 = <b>10 hrs/volume</b></li> <li>• Copibook scanner: .60 min/page x 200 pg = <b>3 1/3 hr/volume</b></li> <li>• Flatbed scanners: 3 min/page x 200 pg = 600 min/60= <b>10 hrs/volume</b></li> </ul> <p>Pre-processing, QC and preliminary XML creation (derive jpgs from master tiff images, create table of content images to use in XML creation, check for missing and/or unacceptable images, assign page numbers, division names, and chapter titles). From numbers recorded in previous two phases, approximately 3/4 of the volumes imaged have no errors necessitating rescanning; 1/4 of the volumes have errors <b>40 min/volume for imaged volumes with no errors</b> <b>60 min/volume for imaged volumes with errors</b></p> <p>Mark-up (metadata review and revision; text review): <b>10 min/volume</b></p> <p>The full grant proposal is online <a href="#">here</a>.</p>
---	---

Overview	Details
<p>Metadata and material prep and post proc.: Map Library</p> <p>DLC digitization: 1,390 hours for 13,418 images: <b>9.6 photos/pages per hour</b></p> <p><b>Plus:</b> DLC cost share of .23 for one year for ingest of another 7,473 already digital images,</p>	<p>Digitize 13,418 historical aerial photographs and 120 paper indexes</p> <p>Incorporate 7,473 aerial photographs from FDOT</p> <p>In total, link 21,417 aerial photos to georectified images</p> <p>OPS Scanning: 1,125 hours (Five scanning technicians for 15 hrs/week for 15 weeks each)</p> <p>OPS Metadata/quality control student: 225 hours (13,418 images @ 60 images/hr)</p>

<p>and training and supervising students</p>	<p>OPS Digital camera operator: 40 hours (120 paper index images @ 3 paper index images/hr)</p> <p>DLC cost share of .23: for ingest of other 7,473 images, system upgrades, and training and supervising students</p>
Overview	Details
<p>267 architectural drawings/blueprints</p> <p>OPS time: 654 hours</p> <p>Average pages per hour, without factoring in cost share time: <b>0.40 pages per hour</b></p> <p><b>Plus:</b> DLC cost share, years 1 and 2</p>	<p>267 architecture drawings, blueprints and related material</p> <p>OPS time: 654 hours</p> <p>DLC cost share, year one: .10 DLC cost share, year two: .18</p>
Overview	Details
<p>Spreadsheets for metadata by Special Collections</p> <p>Material prep. and post-proc. by Preservation</p> <p><i>Digitization by DLC:</i> Monthly averages 3,216 pages/260 hours = <b>12.369 pages per hour</b></p> <p><b>Plus:</b> DLC cost share, .30FTE for each of the three years</p>	<p>DLC cost share: .30 FTE for each of the three years</p> <ul style="list-style-type: none"> <li>• 99,690 pages (90,400 pages; 9,040 letterbook pages; and 250 photo prints/ negatives) in 31 months; average of 3,216 pages per month</li> <li>• Overall average of 3,216 pages per month; actual production per month will vary for the letterbooks and photos, which are more time-consuming</li> <li>• OPS: 1.25 FTE for each of the three years; or 60 hours per week for 31 months; 4.33 weeks per month, or 260 hours per month</li> <li>• Monthly: 3,216 pages/260 hours = <b>12.369 pages per hour</b></li> </ul> <p>Based on experience with test sets, we're building in a 10% reshoot rate for pages, 15% reshoot for letterbooks, and 15% for photos. Adjusted estimates are: 99,440 pages 10,396 letterbook pages 288 photographic materials.</p>

	<p>This estimate assumes use of CopiBook scanner with white sheet backing for letterbooks, and, use of flatbed scanners for all photographic materials and other pages. Some individual sheets may withstand sheet feed scanner, based on experience with similar collections, but we will not count on it. All pages images will be 300 dpi color (24-bit) images. All photographic materials will be 600 dpi grey-scale (8-bit) images.</p> <p>Student labor, no staff costs, archival pages:</p> <ul style="list-style-type: none"> <li>* \$0.25/page scanning +</li> <li>* \$0.25/page image correction/QC +</li> <li>* \$0.03/page mounting/archiving +</li> <li>* \$0.01/page media +</li> <li>* \$0.02/data-logging each file</li> </ul> <p>Subtotal: \$0.56/page + \$0.06 (10% error correction) Each page unit = \$0.62/page</p> <p>Student labor, no staff costs, photographic materials:</p> <ul style="list-style-type: none"> <li>* \$0.40/page scanning +</li> <li>* \$0.25/page image correction/QC +</li> <li>* \$0.03/page mounting/archiving +</li> <li>* \$0.01/page media +</li> <li>* \$0.29/data-logging each file</li> </ul> <p>Subtotal: \$0.98/item + \$0.09 (10% error correction) Total each photo unit = \$1.07/image</p>
--	---

### Time Requirements by Workflow Component

	<b>Catalog record available</b>	DLC evaluates existing record, ingests, and massages records as needed.	<b>Average time:</b> 1 - 5 minutes per item
	<b>Spreadsheet available and accurate</b>	DLC reviews, enhances, imports, and verifies.	<b>Average time:</b> 40 minutes - 2 hours per spreadsheet; average spreadsheet has 200 items

			<p>Longer for extensive spreadsheets or those with new mappings or categories.</p> <p><i>Note:</i> this is only for the import process. The DLC trains others on what information is needed and assists in creating spreadsheet until the creator is comfortable doing so alone.</p>
	<p><b>Spreadsheet available, but incomplete or inaccurate</b></p>	<p><i>Example:</i> a Word file with a table with a single line listing titles, authors, and dates without any consistent separation (no columns, tabs, or commas that can be used to create tabular data).</p> <p>DLC finds a way to separate the rows into tabular data if possible, or copies and pastes all information into a spreadsheet in the correct format. Then, DLC sends the spreadsheet to the selector with any recommendations for added fields and asks for feedback.</p>	<p><b>Average time:</b> 1 minute per item to create the spreadsheet item</p> <p><b>Additional time required:</b> 40 minutes - 2 hours for the completed spreadsheet</p>
	<p><b>No catalog record, spreadsheet, inventory, finding aid, etc. &amp; Materials can be determined.</b></p>	<p><i>Example:</i> a box of only books with no other information.</p> <p>DLC reviews materials, sorting and creating metadata as possible. DLC offers training for future spreadsheet and metadata creation.</p>	<p><b>Average time:</b> 10 minutes per item.</p> <p><b>Additional time required:</b> 40 minutes - 2 hours for the completed spreadsheet</p> <p>OR</p> <p><b>Average time:</b> one or more 1</p>

	OR  For items needing actual catalog records in a traditional format, DLC sets a meeting with Cataloging and together they establish a workflow to have the items cataloged in Cataloging and then returned to the DLC for digitization.	hour meetings + Cataloging time to catalog materials.
<b>No catalog record, spreadsheet, inventory, finding aid, etc. &amp; Materials cannot be determined.</b>	DLC reviews materials, sorting and creating metadata as possible. After sorting and review, DLC staff create a brief spreadsheet. If a Collection Manager is available, DLC staff send the spreadsheet and ask the Collection Manager for feedback. If no Collection Manager is available, DLC staff attempt to work using the newly created spreadsheet.	<b>Average time:</b> varies and can only be determined on a case by case basis
<b>Permissions cleared</b>	Permissions status clearly documented and provided when physical materials received.	<b>Average time:</b> 0 - 1 minute to check documentation in files and update if needed.
<b>Officially Published in US pre-1923, Clear Public Domain</b>	Information is available in a published document. No requirements to consult documentation on length of copyright by year or country; no requirements to consult book copyright renewal database.	<b>Average time:</b> 1 minute to read and verify information to verify status as cleared.
<b>Archival, permissions status communicated after inquiry</b>		<b>Average time:</b> 1 - 3 minutes to call or email to check and update documentation.
<b>Permissions not cleared, but permissions status and the need for DARK archiving clearly documented and provided when</b>	Dark Archiving, if identified as such, requires no additional research.	<b>Average time:</b> 0 - 1 minute to check documentation in files and update if needed.

	<p>physical materials received;</p> <p>Or</p> <p>Permissions status easy to ascertain</p>		
	<p>Permissions not cleared, but wanted and permissions status clearly documented and provided when physical materials received;</p> <p>Or</p> <p>Permissions status easy to ascertain</p>	<p>Requesting permissions</p>	<p><b>Average time:</b> 20 minutes</p> <p>Average process includes checking all pertinent copyright rules, searching for copyright holder, sending permissions request to copyright holder; updating documentation in files that permissions request was sent and noting the information found on the copyright holder. When applicable, scheduling for follow-up inquiry.</p> <p><b>Note:</b> Some materials are significant enough for the allocation of additional resources for pursuing permissions. Those are a case by case basis and normally require <b>at least 2 hours</b>. At least 30 minutes of this time is normally in meetings with collection managers where the necessary background is communicated on how to possibly locate the rights holder and why the particular materials are significant.</p>
	<p><b>Unclear Copyright Status, Holder, etc.</b></p>	<p>Copyright research, and requesting permissions.</p>	<p><b>Average time:</b> 10-20 minutes for copyright research</p> <p>Copyright research consists of searching for information on the materials and copyright holder. If information can't be located quickly, the item is deferred unless it warrants additional resources.</p> <p><b>Additional average time</b></p>

			<b>required:</b> 20 minutes to request permissions. <i>Only required if copyright holder is located.</i>
Disbinding a book	Also includes any clean-up of physical materials, placing in folders and boxes that are labeled and placing those on appropriate book trucks or shelves to be reviewed for appropriate imaging technology	<b>Average time for disbinding a book:</b> 8 minutes per book	
Cutting newspaper pages (normal newspaper size*)	Includes placing in boxes that are labeled and/or placing those on appropriate book trucks or shelves to be queued for imaging  *Some newspapers (i.e.; <i>Iguana;Justice</i> ) are 8 1/2 X 11 and are cut using a paper cutter, and then go through the high speed scanner.	<b>Average time:</b> 20 minutes per inch of newspaper  One month of newspapers from August 2008, with no born digital titles, is 16 inches.  One month of newspapers from October 2009, with 37 newspapers born digital (total of 72 newspaper titles in the Florida newspaper queue), would be under 1/2 of this or under 8 inches.	
Preparing archival files	Sorting, separating, unfolding, flattening, removing staples, paperclips, debris, etc.	<b>Average time:</b> varies and can only be determined on a case by case basis	
Collating, de-duping	Breakdown  Separating out / checking title: 5 sec/title  Collating for input into tracking: 2 secs for monthly, 30 secs for daily  Inputting into tracking (calling	<b>Average time:</b> for new and non-organized or inventoried collections, varies and can only be determined on a case by case basis  <b>Average time for collating newspapers:</b> 47 seconds for one month of a monthly	

		<p>up tracking, inputting, printing tracking sheet, placing on shelf, record in xls for physical tracking): 40 secs for monthly and 3:10 for daily</p>	<p>newspaper; 3:45 for one month of a daily newspaper</p> <p><b>Average time for de-duping:</b>varies, but close to collation time after initial physical material ingest, inventory, and review; duplicates do add an additional time component if they cannot be discarded or returned and must be arranged and kept for an unknown length of time</p>
	<p><b>Books</b></p>	<p>Disbound, and can go through the highspeed scanner</p>	<p><b>Average time:</b> 10 - 15 minutes for 300 normal pages (300dpi grayscale, time increases if many color pages)</p> <p><b>Average time, brittle:</b> 45 - 60 minutes for 300 pages</p> <p>*Time level varies if the scanner has to be cleaned. Brittle pages must be scanned at a slower rate to help prevent rips and jamming.</p>
	<p><b>Books</b></p>	<p>Bound, average book</p>	<p><b>Average time, scanned on a copibook:</b> 90 - 110 minutes for 300 normal pages (no foldouts, tip-ins or oversized pages)</p> <p><b>Average time, if oversized:</b>use times listed for maps and oversized items</p>

		below  <b>Average time with processing:</b> See the post-processing for images section for books for a more accurate assessment of the time for scanning and image processing for a single item. Processing time required is directly related to the imaging technology, so it will vary based on the scanning equipment used.
<b>Maps and Oversize Items</b>	One full capture using the large format camera, not multiple captures and splicing (as is required for many oversize materials)	<b>Average time:</b> 15 minutes for a single capture and processing  <b>Average time for multiple captures and splicing:</b> 30 minutes for two captures (includes processing and splicing), 10 minutes for each additional capture (e.g.; 3 captures=40minutes; 4 captures=50 minutes)
<b>Photos, Loose</b>	Photos, loose and not oversized, are scanned on the flatbed scanners at 600 dpi	<b>Average time:</b> 1 - 3 minutes to scan per photo
<b>Photos, Mounted (scrapbook, etc)</b>	Photos, Mounted (scrapbook, etc)	<b>Average time:</b> 45 - 60 min. for 75pgs
<b>Photos, Aerials</b>	Average time estimate for scanning and image quality control is based on three successful Florida aerials grants.	<b>Average time:</b> 9.6 photos/pages per hour
<b>Slides, 35mm</b>	Color slides are scanned at 4000dpi and with the bulk loader, to scan 24 per hour.	<b>Average time:</b> color slides 4000dpi 24 per 60 min. to scan

	Time increases for older, non plastic mounted slides because they tend to jam the slide scanner.	
<b>4x5 color transparencies</b>	4x5 color transparencies 600dpi	<b>Average time:</b> 3 min. per transparency to scan only
<b>Slides, Glass</b>	Scanning only: 4x5 600dpi 3 min.  4x5 900dpi 3.5 min.	<b>Average time:</b> 3 - 3.5 minutes each
<b>Archival materials</b>	Average times for archival materials vary widely because of: special handling needs and average length. If all of the pages are for the same item and can be handled the same way, the overall time is reduced and overhead from switching to a new item and labeling it is reduced as well.	<b>Average time, scanned on a copibook:</b> 90 - 110 minutes for 300 normal pages (no foldouts, tip-ins or oversized pages; no need for backing; all pages are for the same item)
<b>Newspapers: Current</b>		<b>Average time per page in color:</b> 30 sec  <b>Average per page in black and white:</b> 15 sec
<b>Newspapers: Bound</b>	Additional time depends on: the gutter; whether the paper can be captured 1 up or 2 up; turning odd and even pages; whether a glass plate is required to flatten the pages	<b>Average time per page:</b> at least 3x more than for unbound newspapers
<b>Newspapers: Brittle (requiring large format camera)</b>		<b>Average time:</b> at least 3x more than for normal unbound newspapers, can be even higher
<b>Object, Flat</b>	Using DSLR camera	<b>Average time:</b> set up time can be several hours for a single shot; set up is the largest time component
<b>Object, Rotation</b>	Using DSLR camera	<b>Average time:</b> set up time can be several hours for a single item for 126 images; set

		connected to turntable in DLC.  Additional time is required for equipment packing, traveling to location, setup, and repacking and returning.	up is the largest time component
	<b>Audio:</b> Record Cassette tape Reel to Reel tape  <b>Video: VHS</b>	Record	<b>Average time:</b> Actual digitization time equivalent to length of audio or video file. Thus, 1 hour of audio takes 1 hour to digitize.  Set up time is in addition to this; however, estimate includes set up time within the actual time required because of variances from the degree of supervision needed for the digitization process.  Digitization time may or may not need direct supervision at all times. If it needs to be supervised or not impacts how much other work can be done simultaneously, Other work is most often image post-processing.
	<b>Imaging Ingest: Retro files</b>	Files on CDs, DVDs, portable hard drives, SAN	<b>Average time:</b> varies  <b>Example:</b> ingesting the 94 issues (for v. 1-18) burned to disk for FLMNH bulletins required over 12 staff hours. Time required was to work with the disks (two had cyclic redundancy errors), normalize the file quality, qc the files and notice that pages were missing, locate the missing pages or rescan, reprocess, and then OCR and load.
	<b>Imaging Ingest: Born-Digital IR Materials</b>	Variables include server space available, number of items, size of each item, format of each item (PDF, HTML, AVI, AVI streaming which needs to be ripped or	<b>Average time for 1 volume, new item:</b> 4 min.  Time includes entering item into tracking, downloading

		<p>which requires contacting AT for copies)</p>	<p>item, exporting/converting to TIFF</p> <p><b>Additional volume for serial item already in tracking:</b> 1-4 minutes</p> <p><b>Average time for new groups of materials:</b> varies based on number and type of items</p> <p><b>Example:</b> HPC documents found, required 3 days</p> <p><b>Example:</b> Hard drive from Harn museum for publications and newsletters, required 5 days</p>
	<p><b>Imaging Ingest: Born-Digital Newspapers</b></p>	<p>1-3 minutes per issue covers time to check spreadsheet, add item to tracking with brief data, match new BIBVID to vendor naming structure, and bulk rename, checking data while doing so.</p>	<p><b>Average time:</b> 1-3 minutes per issue, if from a hard drive, not tarred or zipped, do not have errors, and have some human readable title and date identification (in the file name itself, in a spreadsheet or xml file)</p> <p><b>Average time if on CD/DVD:</b> 5-15 minutes per item.*</p> <p>*Includes time to copy files from CD/DVD to a hard drive. Also includes time to recheck copy process because the CD/DVDs have a much higher likelihood of errors.</p>
	<p><b>Imaging Ingest: Vendor Files</b></p>	<p>Variables are:</p>	<p><b>Average time:</b> varies</p>

		<p>File identification and usable structure for batch renaming;</p> <p>Files on drives or decaying disks;</p> <p>Files tarred and zipped have greater frequency of integrity errors;</p>	<p><b>Examples:</b> Vendor digitization of newspapers on microfilm, when digitized to NDNF standards, has an average of 9 hours per reel or 210 pages per hour.</p> <p>Vendor digitization of newspapers on microfilm, when not done to NDNF standards, tends to average a minimum of 70 hours per reel or 29 pages per hour.</p>
	<b>Splitting pages</b>	Required for bulk digitization from microfilm scanned with 2 pages per image.	<b>Average time:</b> varies.
	<b>Splitting separate items</b>	<p>For digitized microfilm, partner files, and retro ingests not separated into items</p> <p>Average reel has 5-10 items. Time required depends on the quality of the film and the accuracy/inclusion of description images (i.e. targets that say the item title &amp; reel position).</p>	<b>Average time:</b> 15 - 30 minutes to split a reel of digitized microfilm into items
	<b>Scan, crop, deskew, levels for Baldwin Books</b>	<p>Scanning &amp; Initial image processing (deskew, crop)</p> <p>Kodak DCS 24n megapixel DSLR camera: 3 min/page x 200 pg = 600 min/60 = <b>10 hrs/volume</b></p> <p>Copibook scanner: .60 min/page x 200 pg = <b>3 1/3 hr/volume</b></p> <p>Flatbed scanners: 3 min/page x 200 pg = 600 min/60= <b>10</b></p>	<p><b>Average time if DSLR camera:</b> 3 min/page</p> <p><b>Average time if scanned with Copibook scanner:</b> .60 min/page</p> <p><b>Average time if scanned on a flatbed scanner:</b> 3 min/page</p> <p>*Please note: in most cases, the times for scanning and image processing are</p>

	<b>hrs/volume</b>	inseparable because the imaging technology used does alter the amount of image processing (deskewing, cropping) required.
<b>Crop, deskew, levels, color correction</b>	disbound volumes	<b>Average time for disbound volumes:</b> 60 - 90 min for 200 pgs
<b>Batching and Copyright blur</b>	Time depends on the amount of material in copyright. Okeechobee News normally requires 1 minute; Miami Times normally requires 30 minutes	<b>Average time:</b> 1-30 minutes
<b>Brief items</b>	Short research items (under ~40 pages) where a table of contents is very unlikely to be used and wouldn't prove of much benefit only have pagination and quality review during QC; no table of contents style metadata is added	<b>Average time:</b> 1-3 minutes per item (item is normally under 40 pages) if no errors
<b>IR</b>		<b>Average time:</b> 1-3 minutes per item (item is normally under 40 pages) if no errors
<b>Newspapers</b>	Sections (A, B, C) and page numbers added, final quality review of item	<b>Average time:</b> 1-3 minutes per item (item is normally under 40 pages) if no errors
<b>Books, Complex</b>	Average time estimate for QC alone is based on the Baldwin Phase III grant time requirements.	<b>Average time:</b> 40 - 60 min/volume (average of 40 for volumes with no errors and 60 for volumes with errors)
<b>Photos, Aerials</b>	Average time estimate for scanning and image quality control is based on two successfully completed grants for Florida aerials.	<b>Average time for scanning and image quality control:</b> 9.6 photos/pages per hour
<b>OCR</b>		<b>Average time:</b> OCR runs constantly against available materials. Average labor time is 15 - 20 minutes per day for all materials to be processed

			<p>that day. Time is to check process, refine any jobs as needed, and correct any errors.</p>
	<p><b>Archiving to FDA</b></p>	<p>FTP and loading drives and mailing (forms, error correction, ingest of reports)</p>	<p><b>Average time:</b> 3 hours to set up external hard drive for file transfer and start file transfer (10 minutes), transfer files (varies based on size of all files being transferred; done on a separate machine and does not interfere with other work), and then drive to drop off the drives and drive time to return to work. Average drop off has been 10 hard drives in one trip.</p> <p>Goal is to have FDA catch up on backlog and be able to FTP daily work and have that process easily without the need to use external hard drives.</p>
	<p><b>Archiving internally</b></p>	<p>Required components of burning DVDs:</p> <ol style="list-style-type: none"> <li><b>1. Labels:</b> printed in batches of 100, 5 minutes to renumber and print: .33 seconds printing time for each label</li> <li><b>2. Labeling each DVD:</b> 10 sec</li> <li><b>2. Burning:</b> 7-8 minutes per DVD (4.4GB)</li> <li><b>3. File sort:</b> 20 seconds per DVD</li> <li><b>4. Filing DVD:</b> 10 seconds per DVD</li> </ol>	<p><b>Average time required with burning DVDs:</b> 9 minutes for each DVD (4.4GB)*</p> <p>*Note: DVD time is as though there's no overlap during burning. Normally 3-4 DVDs are being burned simultaneously. Thus, instead of 9 minutes each equally 36 minutes for 4, the time for 4 is closer to 15 minutes with the 9 base plus additional overhead</p>

		<b>5. Transferring files:</b> moving files from the SAN to a local drive to burn locally and not across the network. Done overnight to reduce time delay; otherwise can take 1-3 hours depending on drive availability and system time	and checking time. <b>Average time expected with Tivoli automation:</b> 0; time would be replaced with 100% load verification
	<b>Load and metadata verification</b>		<b>Average time:</b> 1 hour per day for brief validation using only file names and the <a href="#">m=han</a> page; spot checking under 10% of load items
	<b>Returning physical materials, updating holdings records to discharge or withdraw item from DLC</b>	<b>Example:</b> all IFAS documents must be completed before they can be returned, and they must be properly ordered for all issues. This means that the DLC must store all completed items, keep them in order, and must only file newly completed items in the correct order. Once all are done, only then will the holding records be updated in one large batch.	<b>Average time:</b> varies on requirements for returning items.
	<b>Material reclamation</b>	Pulling folders, relabeling boxes	<b>Average time:</b> varies.
	<b>Metadata updates, Manual</b>	Involves updating the metadata of one or more items. Single items are done manually and large projects (including serial hierarchy changes) employ combination of automated and manual methods.	<b>Average time:</b> 10 minutes/title (manual assignment)
	<b>Serial Hierarchy</b>	Prior work required manual updates for each item (10 minutes X 100). With new tool, DLC staff can update serial hierarchy for batches of items. Tool is being refined for optimal performance.	<b>Average time:</b> 10 minutes for 100 items

As Abby Smith notes in the CLIR report on "Strategies for Building Digitized Collections":

Reliable and meaningful cost data about digitization are rare and not often useful in comparative contexts. Costing out the elements of digitizing means beginning with selection and going to physical preparation, cataloging, physical capture, creation of metadata, mounting and managing files, designing and

maintaining the site, providing additional user services, and going through to implementing a long-term preservation strategy. Virtually every step in digitization involves human intervention and skill, and these costs, unlike those of storage, for example, are unlikely to go down. ([Section 4; 2001](#))

---

## Estimated Times from Other Digital Library Centers

### Brown University Library's Estimated Timelines for Production ([online](#))

Estimated duration will be defined per project. Projects should be in the Digital Technologies production queue at **least 4 months** before they are requested to be complete. Submissions made without adequate planning time will be reviewed on a case-by-case basis by ITS representatives and approved only if there are resources available.

#### Scans, images for the web, digital projects

- Flatbed scanning (10-14 items an hour)
- Camera Room
  - Book or Broadside with no problems -- 5-6 pages per hour
  - Problematic Book (closely bound, fragile, rare, custom support) -- 3-4 pages per hour
- Metadata -- 10-15 minutes per item
- Quality Control for color, cropping, artifacts, and metadata -- total 5 minutes per item (reshoot if necessary)
- Create submission package
- Project management (coordination, communication) -- 20% of overall project effort

#### Website development

Production time is dependent on the scope of the project. Estimate for a minimum of 4-6 weeks for production of a web project, an estimate which will fluctuate and may in fact be longer if the overall volume of work is high.

#### Types of work ([Glossary](#))

- **Digital collection:** framework to access materials which include digital scans (repository items) plus additional content including history, essays, keyword searching, genre searching
- **Digital exhibit:** smaller groups of materials which showcase selected items; scanned objects are presented with descriptions within a curated, narrative flow
- **Web site:** non-collection driven
- **Maintenance of collections, exhibits, and web sites:** continued development that occurs after the site officially launches. This activity is handled as a different project.
- **Digitization:** finite number of items to be scanned that are not related to a specific collection or exhibit
- **Print materials:** Designing promotional materials for Library-related activities only (i.e., brochures, calendars, posters, catalogs, bookmarks, signs)
- **Application Development:** software that improves efficiency for users and improve efficiency of internal processes

- **Web Services Support:** includes Josiah, and special systems support
- **Technology projects:** such as ABET
- **Event coordination:** including documentation and other promotional activities (OUL requests)
- **Metadata creation:** creation of MODS records for material; if MARC records already exist in Josiah, those may be converted to MODS; if no MARC records, MODS must be created with SR assistance; additional decisions to be made: should there be authority control of headings; should there be subject analysis; abstracts must be provided for materials

## Average File Sizes

Online files are approximately 1/3 of the archival file size.

Type	Average Archival File Size	Online, Derivatives (33%)
photo	27 MB	8.91 MB
archival page, color	27 MB	8.91 MB
archival page, grayscale	13.5 MB	4.45 MB
average archival	20.25 MB	6.68 MB
newspaper page	64MB ( <a href="#">Texas Digital Newspaper Program</a> )	21.12MB
microfilm reel (1,000 frames)	64,000 MB	21,334 MB
Map/large format	250 MB	82.5 MB
<b>30 minute oral history audio &amp; PDF transcript</b> (born digital; multiple derivatives; assumes 500MB audio and 15 page PDF at archival grayscale size)	705 MB (varies)	233 MB (varies)
<b>30 minute video from DVD file</b> (born digital; multiple derivatives; assumes maximum full DVD size of 4.4GB)	3.5 GB (varies; assumes 1GB original file and alternate formats that add an additional 2.5X the size)	varies widely (can even be 0 if streaming via YouTube or similar and full files available only on request)
<b>slide, 35mm (scanned at 4000dpi)</b>	52 MB	17.16 MB
<b>Color transparency (scanned at 900dpi)</b>	52 MB	17.16 MB

**Resources:**

- [MB to GB to TB converter](#)
- Estimating pages: "A cubic foot of records comprises about 2,000 pages." (<http://www.archives.gov/foia/ufos.html>). The average archive box is 5 inches.

## Project Planning Resources for Library Faculty and Staff

A number of variables impact the actual time it takes to complete a project; on average:

- One third of the effort will be project planning, preservation preparation, management, and oversight
- One third of the effort will be archival description and indexing
- One third of the effort will be the actual digitization ([cite](#))

### New Projects

- Projects/Collections
  - [Proposal template for new digital collections/projects](#)
  - See the [Smathers Libraries Copyright Policies](#)
- One-off and small requests with the digitize on demand queue
  - Online form to enter and track requests (old, inactive)
  - Tracking spreadsheet (old, inactive)
- [Draft training for metadata for new projects \(old, inactive\)](#)
- Draft timeline for new projects
- For only one or two items, [contact us](#) to see if the digitize on demand process can meet the request
- [Information for potential partners](#)

### Information in this document originally available:

- [http://digital.uflib.ufl.edu/technologies/documentation/average\\_times.htm](http://digital.uflib.ufl.edu/technologies/documentation/average_times.htm)
- [http://digital.uflib.ufl.edu/technologies/documentation/average\\_filesizes.htm](http://digital.uflib.ufl.edu/technologies/documentation/average_filesizes.htm)
- <http://digital.uflib.ufl.edu/technologies/projectplanning/>