

LYRASIS Mass Digitization Collaborative Procedures: Internet Archive Book/Print Materials

June 21, 2011

1. **Request Estimate/Place Order:** request an estimate for your digitization project via the LYRASIS website. Simply return your signed agreement to start the project!
2. **Select:** review the selection guidelines and select items for digitization
3. **Metadata:** prepare metadata spreadsheet
4. **Ship:** prepare items for transportation to regional IA scanning site
5. **Digitization:** IA regional scanning center staff receives, digitizes and returns items
6. **Review:** review and download scanned images

Detailed information for each step is provided below.

1. Request Estimate/Place Order

Visit the LyraSIS website <http://www.lyrasis.org/Products-and-Services/Digital-Services/Mass-Digitization-Collaborative/Request-an-Estimate.aspx> to request estimate.

Library provides the estimated number of pages for the estimate. Please note that every page is digitized (including covers, blank pages). There are also 2 additional shots for each item for camera calibration purposes (shots of a color card at the beginning of each book). Final billing is based on final number of pages digitized.

Return the signed form to submit order. You will receive confirmation, access to collaborative forms, and an invitation to set up an orientation meeting for the program.

2. Select

Institutions should select items according to selection criteria detailed below.

a. Content

Institutions are free to choose materials on any subject. Many have asked for suggestions. LYRASIS encourages unique items such as local histories, special collections, city directories, college catalogs and yearbooks. The process is not recommended for rare books.

For a state example, see Pennsylvania Advisory Committee for Collaborative Digitization (PACCD)'s Collection Development Working Group's *Building the Pennsylvania Digital Library Guidelines* available at (<http://padl.pbwiki.com>).

Note on potential duplication: LYRASIS is willing to accept duplication. Members frequently choose to digitize an item that is already digitized, especially when their copy is unique or previous digitization was incomplete.

Note: All materials digitized for the program must be made available online. If you have concerns about copyright or privacy you should resolve them before digitization. If you want to

redact specific names or pages, please cover those sections before sending for digitization (with white paper/post it note, etc). You should consider preparing a note that the name/info was redacted for privacy.

b. Copyright

Materials must be in the public domain or have copyright holder's permission. Acceptable materials include:

- Published in US before 1923 (public domain)
- Published in US between 1923 - 1963 without a copyright notice (public domain)
- Published in US but out of print (with or without copyright)
- Permission from copyright owner (such as college catalogs)

Additional copyright resources:

- *When Works Pass Into the Public Domain: Copyright Term for Archivists*
www.copyright.cornell.edu/training/Hirtle_Public_Domain.htm
- *Copyright Advisory Office, Columbia University includes Fair Use Checklist and model permissions*
<http://copyright.columbia.edu/copyright/copyright-in-general/>
- *Permission Letters University of California*
<http://www.universityofcalifornia.edu/copyright/permissionletters.html>
- *Crash Course in Copyright* www.utsystem.edu/ogc/intellectualproperty/cprtindx.htm
- *Library of Congress Copyright Office* www.loc.gov/copyright/

c. Metadata – Bibliographic Data

All items must have metadata. Multiple books or multi-volume sets may be digitized. Include volume detail in packing list to ensure details are captured.

d. Size/Style/Paper

Some restrictions are based on equipment. If material can fit under the glass, sit on the platen and pages can be turned, item can be digitized. Limitations are:

- Largest: 9.25" page width x 16" page height
 - 8.75" (21.5cm) x 5.5" (14cm) - 500 PPI or greater
 - 10.75" (27cm) x 8.75" (17cm) - 400 PPI or greater
 - 16" (40cm) x 9.25" (23cm) - 300 PPI or greater
- Smallest: 3" x 3"
- Books greater than 3" thick will need to be reviewed
- Books need to be 150 pages or longer to qualify for lowest pricing. This is a production goal as set up time for large or small books is the same. For items of shorter length, higher costs may be charged.
- Please review materials to send the most complete copy. If pages are missing/damaged, it is problematic to include pages later.
- Foldouts: Many books contain foldouts or items larger than the standard pages in an item. If they are too large, they will be image captured in a separate piece of equipment and the images can be digitally stitched back into proper sequence.

- The maximum size for a single-shot foldout is 30" (76cm) x 20" (50cm)
This is required to maintain approximately 300 ppi output.
- Foldouts larger than 30" (76cm) x 20" (50cm) will be shot in sections to preserve at least a 300 ppi resolution. There will be some overlap between each shot to ensure all content has been captured (example at:
<http://www.archive.org/stream/comingstrugglein00weal#page/n705/mode/2up>).
- IA also has the capability to "stitch" a client's image into the leaf images of any digitized item. This image may be virtually any size. In other words, a library partner may use an internal camera to capture an image at a very high resolution and it may be inserted into the IA processed image stack.
- The price for fold outs is \$ 2.00. *This price is per foldout shot. For example, if a single, large foldout was shot in 4 sections, the price for that foldout would be \$8.00 USD (4 images at \$2.00 each).*

Style

- Monographs should be side bound (no single sheets or top bound).
- Rebound books should be checked for tightness of gutter/binding or whether text runs outside the margin.
- Tight bindings that do not allow book to lay open for digitization may be rejected.
- Soft covers are acceptable.
- "Full-bleed" content
 - This is defined as a text block or illustration that is within 1/4" (6mm) from the page edges or that runs to the edge of a page. To effectively capture this material, Internet Archive will employ "over cropping." This allows them to ensure that all content is captured.
 - Process: The crop box will be outside the outer (non gutter sides) of the page. Where appropriate a black, white or neutral background sheet will be placed under the pages being shot. For images or text that runs into the center gutter, the Internet Archive will 'slightly' over crop onto the opposite image to try to ensure text is not lost. An example of this is viewable at:
<http://www.archive.org/stream/taemyongnyulkang03asam#page/n0/mode/2up>
 - Note: All yearbook-like and art/photography-like books will be over cropped as they typically contain full bleed content. *Partners may opt out of this over cropping methodology but additional charges may be assessed*
- Digitization of items on glossy paper may result in images that include glare
- Gutters/Margins
 - Text less than ¼ inches off the gutter, on a 75-degree angle, is unscannable.
- Covers that are fragile or almost separated from monographs may need to be rejected unless agreed upon with the member.
- Some institutions submit materials in poor condition as they wish to capture the item digitally before further degradation. If this is the case, materials might not 'survive' normal page turning done for digitization.
- If there is more than one title within a bound book; each must be delineated with paper strip; each will be counted as separate book (and must meet 150 page minimum for lowest pricing).

- **Note on serials:** bound serials can be treated as monographs. If a bound item is 150+ pages and is presented as a single item, it would be eligible for lowest pricing. If an item is separated and made available as separate digital entities and each is less than 150 pages, these items are subject to higher pricing. For example, if a bound serial has 300 pages (consisting of multiple issues) and is presented as one digital item online, lowest pricing applies. However, if the item is broken into 10 digital issues of 30 pages each, higher pricing applies.

Materials outside of these guidelines might be digitized, but need to be evaluated.

e. Condition

Each item is inspected before scanning; if damage may occur we will contact institution to discuss. It will not be digitized until written permission is received.

- Library preservation personnel will meet with scanning staff to establish and agree on any special handling of materials to be digitized. If special books, rare books or the equivalent are to be digitized, staff must be notified beforehand if special handling processes are to be used.
- Most paper styles can be scanned, except highly acidic paper that disintegrates to touch. If hard-to-scan paper sent, a review of time to scan vs. 'damage' will be undertaken.
- All pages must be pre-cut.
- Pages should be able to be lifted and turned with normal effort.
- Sticky pages will not be scanned.
- Pages should not be excessively dusty, have excessive mildew or be moldy.
- Microfilmed reproductions should be reviewed with scanning center before being scanned. If the paper/image looks like a film negative (Xerox), these won't be scanned.

3. Metadata - All items must have metadata provided before digitization can occur.

a. **Scanning Attribution & Metadata Form:** complete the Scanning Attribution & Metadata/Wonderfetch spreadsheet, send to laurie.gemmill@lyrasis.org

b. **Metadata Transfer/Harvest**

Initial Set up/Projects of < 500 titles

- For the first 20 items or any project of less than 500 titles, metadata can be transferred via Wonderfetch spreadsheet and translated into Dublin Core. Library of Congress records can also be used.
- Information is uploaded from wonderfetch spreadsheet into online record.
- Records are augmented with technical and administrative metadata such as ppi, scan operator, etc., and made available online.

Projects of 500+ titles

- Following initial set up for projects of more than 500 titles, staff will establish a connection with the member library's online catalog via a Z39.50 server.

- Record copies will be retrieved and translated into Dublin Core then augmented with technical and administrative metadata.

4. Ship: prepare items for transportation to regional scanning site

Wonderfetch

Prepare Wonderfetch spreadsheet for **each** shipment. Include:

- all items to be digitized
 - each volume of a multi-volume set
 - multiple items if bound together in one book
- any special handling instructions
- any copyright permissions
- do not delete columns from the spreadsheet as this causes increased time in loading metadata for scanning center staff
- metadata is transferred directly from the spreadsheet including any typos, so please proofread the metadata before sending

Pack Materials

All materials must be packaged for safe transport to scanning site. Institutions can ship books in boxes, using book carts or any method they feel is best. Staff returns materials in the same boxes and repacks the same way they receive them. While no one can control the weather or shipping companies, there are several ways to minimize potential damage while shipping. Suggestions include:

- Use heavy duty shipping boxes (not grocery or records center boxes); books will be returned in same boxes
- Do not make boxes too heavy as they are more likely to be dropped; a good rule of thumb is: the heavier the books, the smaller the box.
- Do not pack boxes overly full or tight with books, they are more likely to break
- Use bubble wrap or other packing materials to fill gaps in box so box is full and items don't shift during transit
- Include layers of bubble wrap or plastic protection on bottom of box, around sides and top (especially to protect against standing water in transit).
- Wrapping individual items in bubble wrap can be beneficial
- Write "fragile" and/or "handle with care" on each side of the box
- Use 2-3 strips of tape to close the box and reinforce the bottom flaps
- Include prepaid, self addressed return labels inside boxes, preferably away from the top
- Number boxes, example 1 of 3 (so it is immediately noticeable if one is not delivered with the others)

Shipments should consist of no more than 75-100 items to facilitate efficient turn around time. You **must** include paper metadata spreadsheet as a packing list with shipment (only main identifying info needs to be included legible: title, author, volume) as they use this to confirm materials sent.

Special handling procedures must be determined in advance. Members can accompany their materials and view the process; visits must be arranged in advance.

Shipping

Libraries can choose to send via UPS, FedEx ground, courier or hand delivery. It is not recommended to send materials via US Mail. However, they can be sent via US Mail but **not** returned that way.

Send electronic metadata spreadsheet (Wonderfetch) to laurie.gemmill@lyrasis.org. Libraries are responsible for shipping costs to and from scanning center. Include pre-paid return shipping labels in each shipment for return of items. Once ready to ship, send electronic copy of packing list to Laurie Gemmill at laurie.gemmill@lyrasis.org. She will coordinate shipment/digitization.

5. Digitization

Once materials are received, each item is inspected for possible factors that impact its scanning ability. Rejected materials will be marked and returned with books scanned.

More detail regarding the specific equipment, digitization workflow and quality assurance info for books are available as a separate document: *Internet Archive Book Digitization and Quality Control Procedures* (available upon request).

6. Review

Once materials are scanned, LYRASIL staff will notify you that images are ready for review. The materials are viewable via the LYRASIL collection on the IA website at: <http://www.archive.org/details/lyrasis> Members can also search by contributor name, library's unique identifier or IA number (provided via updated wonderfetch spreadsheet). Members can have their materials returned immediately or wait until review is complete. Members should review originals upon receipt.

Members should review images and metadata closely especially following the initial shipment so expectations can be set early. Potential areas of review: missing pages, pages in correct order, correct metadata, and image quality. Note there is the potential for some glare to appear in the corner on larger sized items on glossy paper. Requests for rescanning or revision must be made within 45 days.

The compressed PDFs are made available to provide a relatively small, easy loading file, but that compression may result in loss of quality. The JPEG2000 files are the master files and the "Read Online" view uses the JPEG2000 files. Members should review JPEG2000 and/or "Read Online" versions.

LYRASIL encourages members to download their images and use in their own digital projects as well. Digitized images are available almost immediately after digitization but can take 72/94 hours. QA is completed before materials are returned. Wait 30 days before downloading images due to potential updating of files. Members can use the link below for information on downloading multiple files: <http://code.google.com/p/bhl-bits/source/browse/trunk/utilities/grabby/grabby.sh?r=33>.

Please contact us with any questions or concerns.

LYRASIS Contact:

Laurie Gemmill, Mass Digitization Program Manager
1438 W. Peachtree St, NW, Suite 200, Atlanta, GA 30309
laurie.gemmill@lyrasis.org; 800.999.8558 x2908