Internet Archive (IA)
Book Digitization and Quality Assurance Processes

Robert Miller

May 2007

Rev B

Confidential Document for use by
Internet Archive (IA), LPs (LP) and Digitization Sponsors (DS)

Overview- Technical and Operational details.

This section details in general, of the technical specs and the operation of how the equipment works.

1. Image capture-Color

2. Output formats
    a. Color images in JPEG2000 format in pixels per inch listed below.
    b. OCR in 2 XML formats: ABBYY and DJVU formats. ABBYY 6.0 is used, with its quality. As new versions and alternative vendors become available, a review will be coordinated between LP and DS before implementation. OCR XML character format is UTF-8.
    c. XML for metadata from MARC.
    d. XML for operational metadata collected during scanning.
    e. Searchable PDF.
    f. XML structural metadata for monographs include page numbers when apparent on the pages that is checked by the scanner operator.
    These formats will be delivered from the Internet Archive servers the Internet via HTTP, FTP, RSYNC, or OAI.

3. DPI vs. Size
    a. Example of DPI vs. size, chosen to optimally image a given size book.

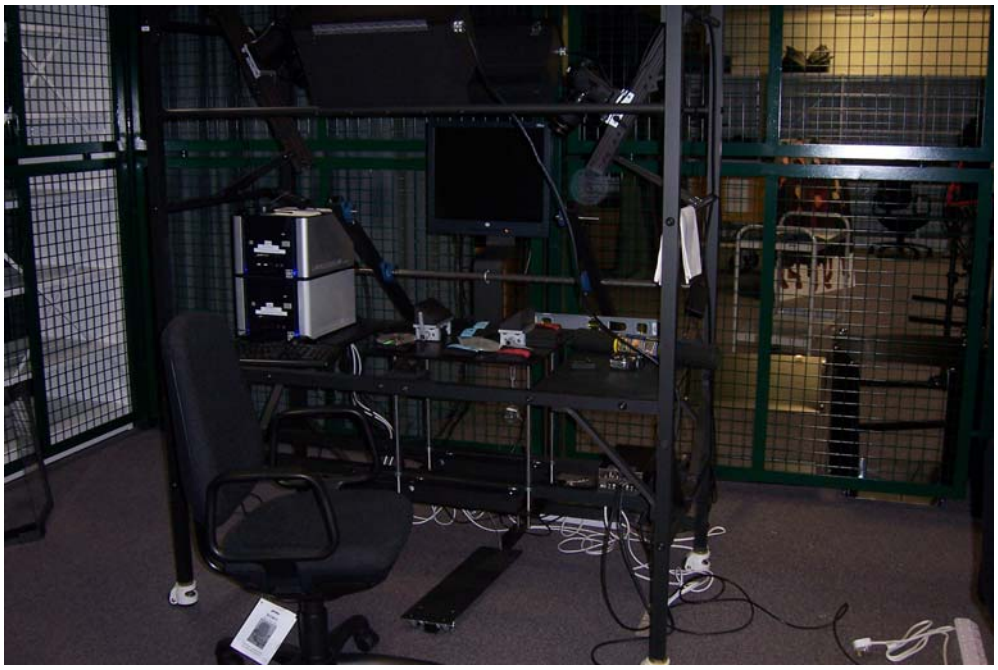
DPI	Height (inch)	Width (inch)
300	14.6	9.7
400	10.9	12.5
500	8.7	6.58
600	7.3	4.9

Scanning Equipment
For Bound and non-bound materials
"Scribe"

Background-
The Internet Archive has tested and evaluated many commercially available scanning devices, but felt that due to the great variety of paper types, binding types and collections to be digitized, an in-house developed Scanning solution would provide the safest and, ultimately, a cost effective way of scanning books. The equipment shown below has been field tested and has successfully scanned millions of pages with virtually no damage caused by the equipment to materials being digitized.

Non-Destructive Scanning Station –

The Scribe workstation*- which is comprised of a frame that holds two cameras on a rail, to capture both the verso and recto pages of the book to be digitized, a cradle that the book sits in (the spring supported cradle is 'v' shaped so there is minimal stress put on the book), a glass platen that is raised and lowered by means of a foot pedal to allow for the pages to be turned, two banks of lights that illuminate the book and two small computers to run the cameras and pre-process the images. The captured images are, upon the completion of the book being scanned and the QA process being completed, are uploaded via RSYNC to processing computers located in California.



* Mechanical/Electrical parameters- 7 amps per Scribe, 800 watts of heat generated per Scribe, standard UK/USA/ voltage, appx 100 sq feet per Scribe of work area, Scribe footprint is 68" long x 37" wide x 79 high; (Dimensions for shipping with out crate- 60" long-rails removed, x 32"wide- remove monitor/arm x 79in high; Dimensions for shipping with crate- 68" long x 42" wide x 86" high, 609 pounds)- NOTE check all doors for access- the weight is appx 300 lbs. For a scanning center of 10 Scribes appx 1000 sq feet of space is required.

Workflow-
Image capture

1. The Scribe scanner currently captures page images using a pair of digital single-lens-reflex (DSLR) cameras, either a 16.7 mega-pixel Canon 1DS-Mark II or a Canon EO 5D, 12.8 mega-pixel, with a Canon EF 100mm f 2.8 macro lens (http://www.usa.canon.com/app/pdf/lens/EFLensChart.pdf). IA is always evaluating new cameras and if a better solution comes along, after coordinating with LP and DS.

2. The lighting system for illuminating the target books consists of eight (8) 5000 Kelvin, 36 degree, 35 watt museum-grade Solux bulbs, and provides a smooth daylight spectrum with a high color-rendering index.

3. Lighting compensation program- To help make the lighting even across images being scanned.

4. Reference targets- A color target (ColorChecker 24) and a white card is shot with each book for reference, which can be used for ICC-based color management.

5. Image transfer- Page images are downloaded in real time to a scanner management and image processing computer which also run the camera management software that releases the camera shutters.

6. Equipment Calibration
    1. Cameras are calibrated per Manufacturer's spec. Cameras out of spec or standard performance will be sent back to Manufacturer for repair.
    2. Lights used in scanning process are replaced as is necessary. Light comp algorithms are run daily on each unit of scanning equipment.
    3. Scribe stations are calibrated and aligned before being used.

Overview- Process steps between Library and Internet Archive (IA)

This document outlines the general flow of information and sequence of steps between the IA and the library. At the end of this document is the QA section that details how the process is checked, maintained and updated to ensure compliance.

Process steps-

A. Meta Data- for all new collections that require attribution or at the beginning of a scanning center set up.

1. Meta data and set up form- includes Contributing Library, Digitzation Sponsor, Collection Name, Contact details, etc. to be filled out by Library and sent to Jae@archive.org .
2. Meta data set up is then incorporated into IA scanning/loading screens. This is to ensure proper attribution and organization of the materials is being completed.
3. If the Z39.50 set up is not being used to locate Library catalog records, alternate confirmation of how records are to be located must be agreed to by LP and IA. A test pick list of at least 50 records, see below, and should be generated to test that IA can locate the proper MARC records.

Collection	BOOK ID	BIB ID	Title	Author	Vol.	Item Barcode	Call	Item Location	tatu	re
BANF 4.11.07	historicaldictio00wilebost	2368353	istorical dictionary of Jainism	ey, Kristi L., 19		3.9999E+13	BL1303.W45 2004	BPL1ST	i	
BANF 4.11.07	padaindexreverse00yamabost	1523716	ada index to early Jain canon	mazaki, Morii		3.7777E+13	L1310.5 .Y36 1995	BPL1ST	i	
BANF 4.11.07	studienzumsuyaga00bollbost	560193	Jainas u.d. anderen Weltans	ollee, Willem B		3.7777E+13	BL1312.3.S886 B6	BPL1ST	i	
BANF 4.11.07	thisisreality00potabost	1686912	This is reality	Potanjali		3.7777E+13	.132.Y6 P267 196	BPL1ST	i	
BANF 4.11.07	Assemblyoflisten00carrbost	1003270	sembly of listeners : Jains in			3.7777E+13	BL1320 .A87 1991	BPL1ST	i	

4. Any changes or requests by the Library that would impact the Meta data must be in writing; for example, a new collection, a new funding source or a new sub-collection.

B. Preservation and handling-

1. Library preservation personnel will meet with IA to establish and agree on how to handle materials to be scanned, how to deal with obvious rejects, how to flag and tag materials not able to be scanned, agreement on error codes and the like. Deviations from this process will

be in writing and where possible all steps are documented with a visual example for reference.

2. A review between IA/LP on what materials can/can't be scanned is conducted. Questionable materials may be tested or tried before being put into the scanning plan. All selection of materials will follow these guidelines**,** See section F.
3. Rare materials are defined as materials that would normally not be in circulation or should not be included in the general scanning population.

C. Retrieval, packing and shipment of Materials by LP-

1. Materials, meeting IA specifications, see F, will be delivered to the Internet Archive.
2. Materials will be packed on cart and wrapped for transport and shipping, unless otherwise agreed to. Any special procedures will be determined in advance.
3. A paper and a digital (excel) copy are to travel with each shipment of materials delivered to Scanning Center.
   (Not applicable at University of Toronto)

D. Receipt of shipment by IA prior to scanning

1. IA will match the count on the pick list to the materials on the cart. If count matches, books will move to the loading area. If the book count does not match- IA will alert Library.

E. Inspection of materials to be scanned by IA prior to scanning

1. IA, as it loads each item to be scanned, will inspect each item for possible factors that would impact it's ability to be scanned. Rejected materials will be marked in a pre-agreed format and returned with books scanned. These items may be scanned at a later date as new processes become available or a different cost structure is put in place.

F. Digitization Criteria to be used by Library and IA- see addendum

1. Criteria to be used for determining if the materials may be digitized are listed below and will include, but not necessarily be limited to the following:
    1. Any special preservation standards that came on cart from the LP.
    2. Materials that have multiple titles/physical volume (eg, 63 vols, 20-40 pamphlets per vol, of "Forestry Pamphlets" without analytics) will be reviewed to ensure all proper meta data is understood.
    3. Materials will be screened for size- materials not fitting the requirements shown below will be returned unscanned.

4. Size-
    - 9.7 wide by 14.5 high at the max. Books as small as 3 inches x 3 inches may possibly be scanned.
    - less than 3 inches thick are ok, greater than 3 inches will need to be reviewed.
    - Books should on average should be 200 pages or larger. If a collection is mostly under 100 pages a review should be undertaken between IA and DS to ensure the quoted price per scanned page can be met.

Comment [MC1]: Robert will determine Cost for special production channel in June 07

5. Book Style-
    1. Side bound Monographs, no single sheets, no top bound books.
    2. Rebound books need to be checked for how tight the gutter or binding is, or if the text runs outside the margin (it will show the cradle).
    3. When there are more than one title within a bound book; each of these has to be clearly marked with a paper strip; each of these will be counted as a separate book. Deviations from this must be approved between IA/LP.
    4. Fold outs are rejected now, but will be included as part of a special production channel.
    5. Soft cover books are ok if they are bound.
    6. Covers that are almost separated from book and appear too fragile may be rejected unless agreed to in advance by LP.
    7. Material condition-Materials should be of similar condition or quality to what would be put into circulation. Materials deemed not robust to go into circulation should be reviewed with Scanning Center Coordinator before scanning.
    8. Tight bindings that will not lay open for digitization per IA specification limits will be rejected.

Comment [MC2]: Robert will determine Cost for special production channel in June 07

6. Paper style-
    1. Most paper styles can be scanned, except highly acidic paper that disintegrates to the touch. Note that if a hard-to-scan paper is to be digitized; a review of time to scan versus any 'damage' will be undertaken.
    2. All pages should be pre-cut. Unless otherwise instructed, books with uncut pages will not be scanned.
    3. Pages should be able to be lifted and turned with normal effort. Sticky pages or the like will not be scanned.

4. Pages should not be excessively dusty, have excessive mildew or be moldy.
5. 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. If the microfilm reproductions have more than one page on each leaf these will be rejected.

7. Gutters/Margins-
   1. Any book where the text is less than a quarter inch off the gutter, on an approximate 75-degree angle will be unscannable.
   2. Text that runs to the edge of the page or margin can be scanned but the presentation will be poor, as the cradle will show. The LP must approve this.

8. Bibliographic data-
   1. Multiple books or multi volume set-The usual problem with bibliographic info is that there is only one bib record for a set. The LP will make a decision as to how this should be handled.
   2. Books that are out of approved copyright range. These will be set aside, unless otherwise agreed to by LP and IA.
   3. A book will be set aside when a Marc record can't be located when a Call ID/Book ID or equivalent is inputted.

9. Rejection codes to be sent back to Library are:

Code	Definition
BI	Fragile or no binding (includes items in clam shells or phase boxes)
CAT	Cataloging error
DAM	Damaged
DAT	still in copyright
FO	Foldouts
LG	too large
MAR	margins too tight
MIS	Missing pages
MUL	multiple titles bound together
NA	not available
LAN	Outside language parameters
LIST	Picklist error
LINK	unsuccessful link to metadata
NOS	not on shelf – missing/lost
OUT	not on shelf – checked out
PAG	Pagination problems: section(s) bound out of order or upside down
PAP	Brittle paper, tissue paper

SKW	Skewed text – to point of being unreadable
SM	too small
UNC	Uncut pages (more than 5)
FOR	non-book format
VEL	Vellum
WD	Withdrawn
SPH	requires special handling
DUP	exact duplicate of another on list

G. IA loading process-

1. The book id is loaded into IA Screen to located appropriate MARC record- records not found will be cause for reject of book.
2. If IA receives a series that is cataloged under one bib record, without volume numbers, IA will either add the year as a volume number or add the volume number in IA's set up screen. This would be analogous to, in the physical world, where the series are shelved together, a person locating the book(s) from the call number and then scanning the items on the shelf for the volume they want. IA will not in a series, delete any information from the descriptions field or add to the descriptions field on the MARC record.
3. QA check is done to ensure book and MARC record match.
4. Unique identifier is created and MARC record is attached to that identifier.
5. Book is placed in queue for scanning.
6. As mentioned above, any books determined not to be scannable are set aside and a rejection form is attached.
7. Each book is also given a color-coded flag that shows the Book identifier.
8. Any special scanning instructions are included with book.

H. IA scanning process-

1. Materials to be scanned are placed in queue for scanners, typically on book carts.
2. The flag inside the material to be scanned is matched to digital file to ensure a proper match.
3. Images are scanned into appropriate digital file.
4. Images are QA'd following adjustments to ensure proper preservation or presentation.
5. Digital file is closed and uploaded to IA processing center.

I. IA processing

1. Uploaded images are processed to create storage files and access files.

2. File Formats, are covered in the Technical Spec and Equipment section above.

3. Files delivered for download to LP are listed above.

4. Detailed list of files
    a. ID.pdf
    b. ID_jp2.zip - will not be accessible, this is only for long term preservation
        i. zipped folder of the book without bookplate and watermark is specific to the sponsor, contributor; e.g. this will vary for each sponsor/library.
        ii. [ID]_nnnn.jp2 (where first image index number is the front cover and the last scan # is the back cover)
    c. ID_lib_jp2.zip
        i. zipped folder of the book with bookplate and watermark
        ii. [ID]_lib_nnnn.jp2 (where image index the first number is front cover and the last scan # is the back cover)
    d. ID_marc.xml
    e. ID_meta.mrc
    f. ID_meta.xml
    g. ID_metasource.xml
    h. ID_raw_jp2.zip , unprocessed storage format, no watermark/book plate
    i. Scandata.zip

5. Metadata_will reside in meta.xml file, and will include the following required fields for the library:
    a. Identifier
    b. ARK (begins ark:/13960/*) ; this is an experimental field for California Digital Library
    c. Collection-Library (from pick list); this applies only for UC Libraries
    d. Identifier-bib (unique identifier -- local catalog number; from pick list), this only applies to UC libraries
    e. Contributor
    f. Title
    g. Volume
    h. Creator (if in MARC record)
    i. Publisher (if in MARC record)
    j. Collection (possibly multiple collection fields)
    k. Operator
    l. Scanner
    m. Scandate

n. Identifier-access (URL for accessing this book)

6. **Processing Background-**
   The digitized image is captured initially as a camera raw file (CR2). This is run through a JPG 2000 compression to generate a raw JPG 2000 for storage. The raw JPG 2000 is then turned into a processed master which is used to generate the access formats.

   i. Storage format- raw JPG 2000 is a compressed, lossy, uncropped, non-rotated, non-deskewed, non-light comp'd JPG 2000 file; which is the storage file. Image sizes vary depending on the complexity of the page, but are typically in the 900 KB range, yielding an approximate compression ratio of 15:1 relative to the camera raw image (CR2 is appx 15MB/image.)
   ii. Processed master- lossy, cropped, rotated, de-skewed, light comp'd JPG 2000. Image sizes may vary depending on complexity of the page, but are typically in the 800 KB range, yielding an approximate compression ratio of 15:1 relative to the camera raw image (CR2 is appx 15MB/image).
   iii. Access format- the processed JPG 2000 masters are compressed in a JPG 2000 format which feeds into the OCR and book generation tools. Image sizes may vary depending on the complexity of the page, but are typically in the 760 KB range, yielding an approximate compression ratio of 20:1; relative to the camera raw image (CR2 is appx 15MB/image). PDF and DjVu; both of which are OCR'd.

   b. Quality settings will vary based on vendor tools used. For example a quality setting of 50 on a scale of 1-100 was used for the Luratech. This setting was determined based on user surveys.

K. Turnaround for processing by IA- typically 72 hours from arrival to return of book cart.

1. The goal is to derive and upload a book within 24 hours after scanning.
2. An internal IA QA step is performed inside the scanning center. Criteria for QA are outlined below in the Quality Section.
3. If the scanning lot is rejected, then the process outlined in the Quality Section K is undertaken-
4. A scanned item is then published online within 48 hours after scanning.

5. Materials scanned are then 'curated' by IA and are available for downloading after that by the LP.
6. Approved Materials having been scanned are then ready to be checked out and returned to the LP.

L. IA Scanning Center Check-Out Process

1. Scanning coordinator packs book into shipping cart/container per guidelines established between LP and IA.
2. Creates and attaches the report communicating books rejected for scanning and identify failure.
3. Books transferred to LP.

QA Plan-

Overview- There are four major phases to the QA process:

1) Before the materials are uploaded-At the book loading and scanning station; the scanners looks for; amongst other things, missing pages, crop/deskew problems, page marking (title page, front/back cover, tissue paper, first page of table of contents and notes any defects in the book (i.e. Missing/torn pages).

2) After the images are uploaded, derived and available via an URL- a statistical sampling and QA is conducted within the Scanning Center. Per ANSI z1.4 1993 Table 1, General Level 2. Details; see below.

3) Before the curation and bill is generated- An internal random audit is conducted outside the scanning center before the final curation approval and bill is generated.

4) After the materials are received by the LP and the DS. Errors brought to IA's attention will be dealt with in a timely (appx monthly) basis. A decision will be made by IA as to whether it is best to rescan the material or fix it post-derive. Rescanning is to be avoided as it requires generating a new URL and is usually the most expensive solution. The timeframe for the library to identify errors that will be fixed by the IA at no-charge shall be detailed in the digitization plan.

Scanning Center QA: IA uses ANSI z1.4 1993 Table 1, General Level 2 http://www.proqc.com/dl/aql.pdf

Each day the scanning center will review a set of books from the previous days scanning. The number of books to QA depends on the total number of books in the set.

Books in set	9-15	16-25	26-50	51-90	91-150	151-280
number to QA	3	5	8	13	20	32

The scanning center coordinator is responsible for choosing a representative set to reflect a mix of scanners/scribes and conforming to the statistical chart.

A. QA Process steps

Books are inspected for the criteria below (see Freeze codes B1-B 3) ; on-line, using the relevant files for each coded; found in the digital book record including pdfs to look for errors or defects. Errors or defects, if found, are noted and added to the IA meta manager form. An automatic scoring is then performed and a "pass/fail" grade is assigned to the lot.

bin	min	max	level	sample	major_accept	major_reject	minor_accept	minor_reject
1	9	15	2	3	0	1	0	1
2	16	25	2	5	0	1	0	1
3	26	50	2	8	0	1	1	2
4	51	90	2	13	1	2	1	2
5	91	150	2	20	1	2	2	3
6	151	280	2	32	2	3	2	3
7	281	1,000,000	2	99	7	10	7	10

Explanation- If 125 books were scanned in a period to be inspected, bin 5 would be selected. According to the truth table above, if there were 1 Major error or less and 2 Minor errors or less, the lot is passed. If there are 2 or more major defects or 3 or more Minor rejects, the lot fails. See below for what happens after this. The major/minor detail is show below in B5. Note: for major defect found during QA, they will be repaired on that book even if the lot passes.

- If the lot passes, the Scanning Center will approve all books (Curate).
- If a "fail' is generated, the Scanning Center Coordinator will review the errors/defects to ascertain if the errors were generated from outside the Scanning Center (for example a missing access file error would be sent to engineering for review) or from within the Scanning Center (for example a missing page).
    - If the error was generated from within the Scanning Center the Coordinator would follow a pre-determined set of process steps ultimately culminating in a recommendation to deviate or approve the lot or a portion of the lot with appropriate corrective actions identified. At this stage the Book's Director or the Headquarters QA staff person is involved and must approve a deviation. A corrective action report will be generated for rejected lots. This will be reviewed with management for longer-term solutions or corrective action. This is done daily.

B. Codes used and shown on the QA report are:

B1. FREEZE CODES, part I. (100-113, and 130-138 will ultimately be rescanned and the original URL made dark (not publicly available), a new URL generated and this will be communicated to the Library by IA via email)
Books in general that can't be corrected post-derive, will have to be rescanned. If it eventually turns out that IA can't rescan or fix a book, it will not be billed.

-- Formats --
101 Test book
102 DjVu is missing or corrupt
103 PDF is missing or corrupt
104 Flip book is missing or corrupt

Resolution of errors found here- Material is rederived, if that doesn't correct the problem, the material is rescanned.

-- Uploading or piping problems --
110 Truncated file(s)
111 Book deleted from scribe before upload completed
112 Missing files(s)
113 Cr2.tar file is malformed

Resolution of errors found here- Material is rescanned.

-- Metadata –
120 Book is not in public domain
121 Date is 1923 or later
122 Date is unclear
123 Date is 1923 or later-

Resolution of errors found here- If material is in copyright, item is removed from Search engine. If material is in question, Library is consulted and appropriate action is taken.

-- Images --
130 Cropped text
131 Blurred page(s)
132 Missing page(s), goal is zero pages missing per book. IA will note if pages are missing in the books to be scanned.
133 Front cover missing

134 Back cover missing
135 Book was scanned twice; identified copy is darkened and removed from search engine.
136 Text is washed out or overly dark (bad light-comp)
137 Evidence of scanner (fingers/shadows/etc) visible on page
138 Glass not centered in gutter; text is distorted or cropped

Resolution of errors found here- For items 130-134, material is rescanned. For items 136-138 a review is made by IA to deviate, accept or rescan.

B2. ***FREEZE CODES, part II.***
Use these codes for books that have fixable problems, but are not yet in billable condition.
140 Book and metadata do not match
141 n/a
142 Tissue pages marked incorrectly
143 Anomaly in image format is under investigation
144 Left/right pages are reversed

Resolution of errors found here- For items 140 and 142, post derive correction is attempted. For item 144, if post derive correction won't work, material is rescanned.

B3. *INFORMATIONAL CODES*

150 Bibliographic data missing:
151 De-commissioned
152 Copyright evidence reported incorrectly. Info corrected in QA.
153 Bibliographic record from library is truncated
154 Possible error in bibliographic record from library
155 Foreign language character encoding is incorrect
156 Incorrect of missing collection-library or bib-id

160 Light/dark pages (intermittent)
161 Light/dark pages (throughout),
162 Pages skewed
163 Color cards show in access formats
164 White cards show in access formats
165 Both white cards and color cards show in access formats
166 Image of cradle is visible at front or back
167 Different crop-box sizes in same spread

168 Bad crop at page edges
169 Duplicate page spreads scanned

170 Page types not marked or marked incorrectly
171 Title page not marked b/c book does not have title page
172 Scan factors not marked or noted

198 This would be a good display book
199 Approved with no problems noted

Resolution of errors found here- For items 150, 152, 155 and 156, post derive correction is possible. For errors, 153 and 154, errors must be reviewed with Library. For errors, 160, 161, 162, 166, 167, 168 a review is conducted to see if the material is acceptable for OCR'able texts. This is accomplished by using the OCR function on the word in question. If the material can't be OCR'd based on the IA software being used, then the book would be rejected For errors, (only IA internal requirements) 166, 170 and 171 a post derive correction is attempted, if unsuccessful materials must be rescanned.

Error resolution falls into three forms of correction; post-derive, rescanning or a consultation with the library. If materials can't be corrected with either of these methods, than the book is rejected.

The following errors may be attempted to be corrected by post-derive treatment:
110, 111, 112, 113, 140, 142, 144, 150 ,152, 155, 156, 160, 161, 162, 166, 167, 168, 170, 171

The following errors may require rescanning:
130, 131, 132, 133, 134, 136, 137, 138

The following errors require consultation with the Library.
120, 121, 122, 123 , 153, 154

B 4- Rescanning process-

1. For materials that are to be rescanned, a request to pull those books requiring rescanning is submitted to the Library; usually once a month. Materials are pulled, scanned and the original item is removed from the Internet Archive search engine and a new URL is assigned. This new URL is sent to the LP and the DS along with the old URL for reference. A bug report could be the means to track this process.

B5- Error Codes & Classes

class/id	Description	type	defects	Totals
Formats				
101	Test book	major	0	
102	DjVu is missing or corrupt	major	0	
103	PDF is missing or corrupt	major	0	
104	Flip book is missing or corrupt	major	0	
			major:	0
			minor:	0
			total:	0
			status:	OK
Uploading or piping problems				
110	Truncated file(s)	major	0	
111	Book deleted from scribe before upload completed	major	0	
112	Missing files(s)	major	0	
113	Cr2.tar file is malformed	major	0	
			major:	0
			minor:	0
			total:	0
			status:	OK
Metadata				
120	Book is not in public domain	major	0	
121	Date is 1923 or later	major	0	
122	Date is unclear	major	0	
123	Date is 1923 or later	major	0	
			major:	0
			minor:	0
			total:	0
			status:	OK
Images				
130	Cropped text	major	0	
131	Blurred page(s)	major	0	
132	Missing page(s)	major	0	
133	Front cover missing	major	0	

134	Back cover missing	major	0	
135	Book was scanned twice; this copy darkened.	minor	0	
136	Washed-out text (bad light-comp)	minor	0	
137	Evidence of scanner (fingers/shadows/etc) visible on page	minor	0	
138	Glass not centered in gutter; text is distorted or cropped	minor	0	
140	Book and metadata do not match	minor	0	
141	n/a	major	0	
142	Tissue pages marked incorrectly	major	0	
143	Anomaly in image format is under investigation	major	0	
144	Left/right pages are reversed	major	0	
			major:	0
			minor:	0
			total:	0
			status:	OK
Bibliographic				
150	Bibliographic data missing: MetaFetch was not run in scanning center (post MF done in QA)	minor	0	
151	Bibliographic data missing: MetaFetch was run but did not merge (post MF done in QA)	minor	0	
152	Copyright evidence with reported incorrectly. Info corrected in QA.	minor	0	
153	Bibliographic record from library is truncated	minor	0	
154	Possible error in bibliographic record from library	minor	0	
155	Foreign language character encoding is incorrect	minor	0	
156	Incorrect or missing collection-library or bib-id	major		0
			minor:	0
			total:	0
			status:	OK
Consistency				
160	Light/dark pages (intermittent)	minor	0	
161	Light/dark pages (throughout)	minor	0	
162	Pages skewed	minor	0	
163	Color cards show in access formats	minor	0	

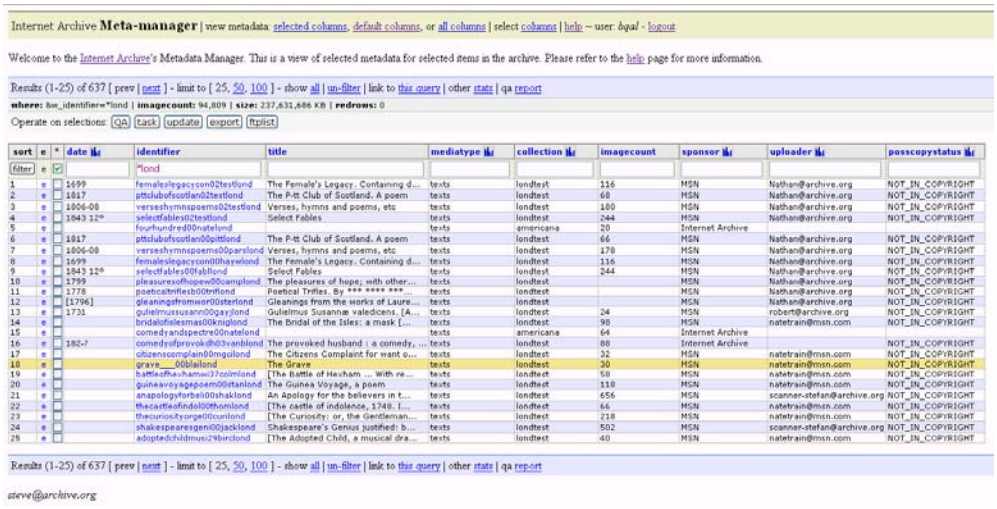
164	White cards show in access formats	minor	0
165	Both white cards and color cards show in access formats	minor	0
166	Image of cradle is visible at front or back	minor	0
167	Different crop-box sizes in same spread	minor	0
168	Bad crop at page edges	minor	0
169	Duplicate page spreads scanned	minor	0
170	Page types not marked or marked incorrectly	minor	0
171	Title page not marked b/c book does not have title page	minor	0
172	Scan factors not marked or noted	minor	0

major:	0
minor:	0
total:	0
status:	OK

2. To track errors found by LP or the DS a bug tracking system could be used. Response times for error resolution will be determined based on type of error and time of response. In general, IA will attempt to resolve errors brought to its attention within 30 days of an error being identified.

C. Meta Manager, the post scanning reporting tool

This is the reporting tool that the Library may use to search and review books that have been scanned, uploaded, QA'd and then curated. The curation stage is the last stage in the IA process where the books are made viewable to the Library. This may happen on a non-scheduled basis but is typically done several times a month.



Fields seen by LP and DS in the Meta-Manager view will also be inspected, plus several internal fields. These fields will include:

- identifier
- title
- creator
- collection
- image count
- contributor
- sponsor
- sponsor date
- scandate
- curatenote
- curate date

D. Library card, required to view the Meta Manager

An IA library card and an email are required to view the metamanager (see steps listed below). Here is the process to access the meta data page.

1 Go to www.archive.org
2. Go to Patron info
3. Click on "get a virtual library card"
4. Have Jae Jae@archive.org create the sponsor view and attach the library card info to that view
5. Books may then be viewed that have been curated

Changes to these Processes

Proposed process changes that would impact LP or DS will be communicated for review and discussion prior to being implemented.

Contact List- For IA Scanning Center

Contacts-IA- Note some of these will change based on specific IA scanning center. Contact Robert Miller as primary point of contact for initial program review.

Name	Role	Phone	Email	Fax
Jae Mauthe	Meta Data setup	415 810 5972	jae@archive.org	
See Digitization Plan	Pick List and shipment			
See Digitization Plan	Shipment			
Robert Miller	IT	415 640 1092	Robert@archive.org	
See Digitization Plan	Scanning center manager			
Robert Miller	**Project manager**	**415 640 1092**	**robert@archive.org**	
Marcus	Metamanager	415 561 6767	marcus@archive.org	
Marcus Lucero	Quality	415 561 6767	marcus@archive.org	
Robert Miller	Download	415 640 1092	robert@archive.org	
Marcus Lucero	Metadata	415 561 6767	marcus@archive.org	
Robert Miller	Eng questions	415 640 1092	robert@archive.org	

Addendum- Ramping up a Scanning Center

Each of the segments below is to confirm process, performance and expectations as a scanning center is ramped up. Note: This primarily pertains to new scanning centers versus new collections.

IA Engineering pilot- 5 books scanned by supervisor. Confirm pick list, packing list, z39.50, MARC record, IA deriving/processing/posting with attribution. URLs are QA'd by Jae. URLs are then sent to LP and DS for review.

IA Production pilot- 50 books scanned by scanner (5 books from each Scribe). Confirm shipping method to IA, material handling, pick list, packing list, z39.50, MARC record, deriving/processing/posting with attribution, return shipping procedure, metamanager and OAI. URL's are QA'd by supervisor. URLs are sent to LP and DS for review.

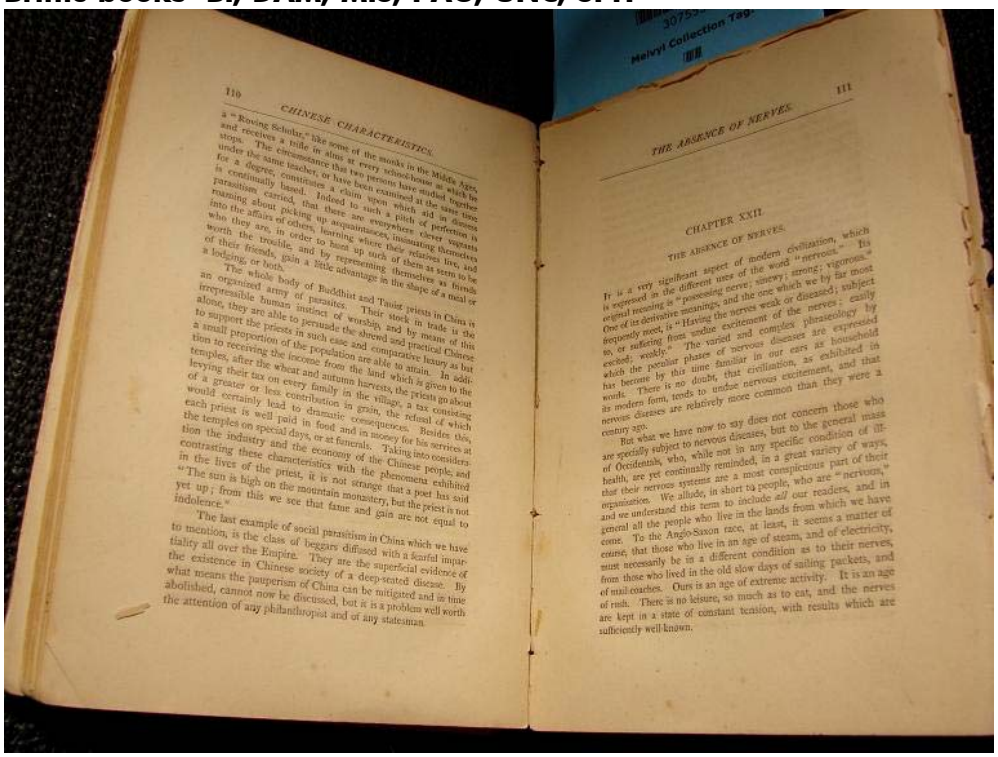
IA Production- establish agreement on quantity and expected turnaround.

Addendum-
Examples of rejected books. Books in question maybe tested for scannability

Code	Definition
BI	Fragile or no binding (includes items in clam shells or phase boxes)
CAT	Cataloging error
DAM	Damaged
DAT	still in copyright
FO	Foldouts
LG	too large
MAR	margins too tight
MIS	Missing pages
MUL	multiple titles bound together
NA	not available
LAN	Outside language parameters
LIST	Picklist error
LINK	unsuccessful link to metadata
NOS	not on shelf – missing/lost
OUT	not on shelf – checked out
PAG	Pagination problems: section(s) bound out of order or upside down
PAP	Brittle paper, tissue paper
SKW	Skewed text – to point of being unreadable
SM	too small
UNC	Uncut pages (more than 5)
FOR	non-book format

VEL	Vellum
WD	Withdrawn
SPH	requires special handling
DUP	exact duplicate of another on list

Brittle books- BI, DAM, MIS, PAG, UNC, SPH



110 CHINESE CHARACTERISTICS.

a "Roving Scholar," like some of the monks in the Middle Ages, and receives a trifle in alms at every schoolhouse at which he stops. The circumstance that two persons have studied together under the same teacher, or have been examined at the same time for a degree, constitutes a claim upon which aid in distress is continually based. Indeed to such a pitch of perfection is parasitism carried, that there are everywhere clever vagrants roaming about picking up acquaintances, insinuating themselves into the affairs of others, learning where their relatives live, and who they are, in order to hunt up such of them as seem to be worth the trouble, and by representing themselves as friends of their friends, gain a little advantage in the shape of a meal or a lodging, or both.

The whole body of Buddhist and Taoist priests in China is an organized army of parasites. Their stock in trade is the irrepressible human instinct of worship, and by means of this alone, they are able to persuade the shrewd and practical Chinese to support the priests in such ease and comparative luxury as but a small proportion of the population are able to attain. In addition to receiving the income from the land which is given to the temples, after the wheat and autumn harvests, the priests go about levying their tax on every family in the village, a tax consisting of a greater or less contribution in grain, the refusal of which would certainly lead to dramatic consequences. Besides this, each priest is well paid in food and in money for his services at the temples on special days, or at funerals. Taking into consideration the industry and the economy of the Chinese people, and contrasting these characteristics with the phenomena exhibited in the lives of the priest, it is not strange that a poet has said "The sun is high on the mountain monastery, but the priest is not yet up; from this we see that fame and gain are not equal to indolence."

The last example of social parasitism in China which we have to mention, is the class of beggars diffused with a fearful impartiality all over the Empire. They are the superficial evidence of the existence in Chinese society of a deep-seated disease. By what means the pauperism of China can be mitigated and in time abolished, cannot now be discussed, but it is a problem well worth the attention of any philanthropist and of any statesman.

111 THE ABSENCE OF NERVES.

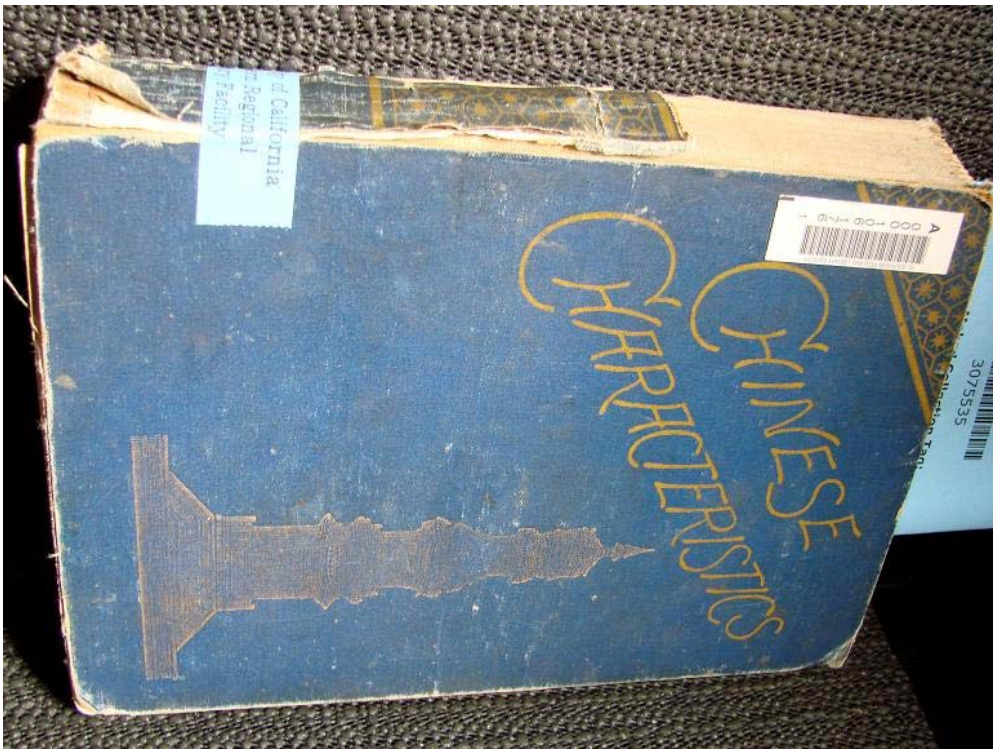
CHAPTER XXII.

THE ABSENCE OF NERVES.

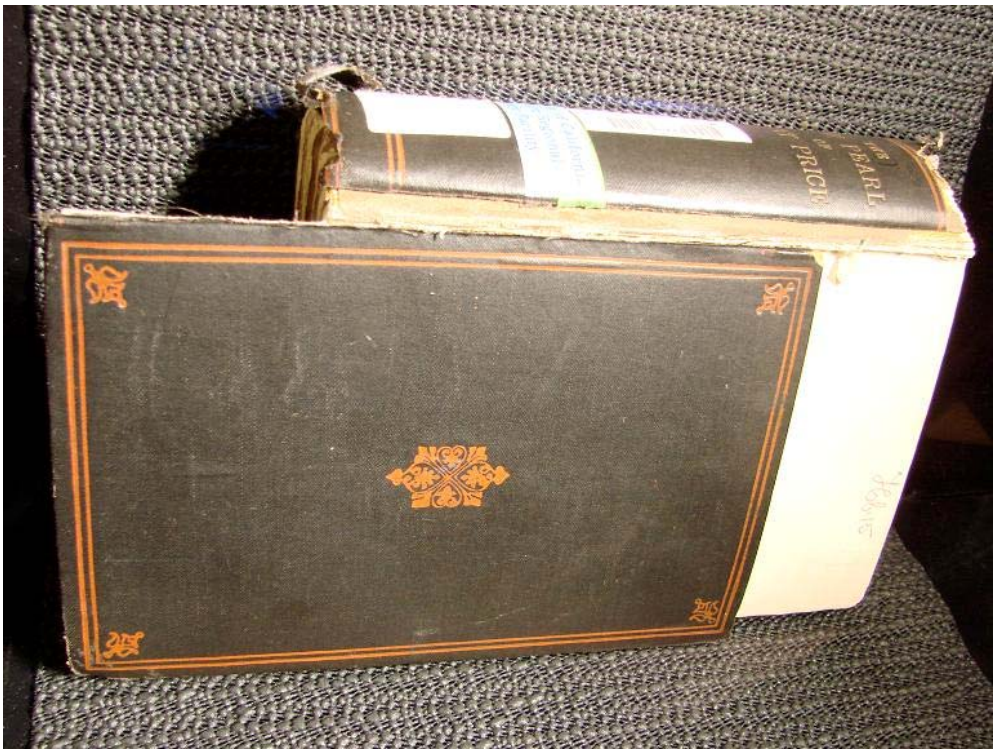
It is a very significant aspect of modern civilization, which is expressed in the different uses of the word "nervous." Its original meaning is "possessing nerve; sinewy; strong; vigorous." One of its derivative meanings, and the one which we by far most frequently meet, is "Having the nerves weak or diseased; subject to, or suffering from undue excitement of the nerves; easily excited; weakly." The varied and complex phraseology by which the peculiar phases of nervous diseases are expressed has become by this time familiar in our ears as household words. There is no doubt, that civilization, as exhibited in its modern form, tends to induce nervous excitement, and that nervous diseases are relatively more common than they were a century ago.

But what we have now to say does not concern those who are specially subject to nervous diseases, but to the general mass of Occidentals, who, while not in any specific condition of ill-health, are yet continually reminded, in a great variety of ways, that their nervous systems are a most conspicuous part of their organization. We allude, in short to all our readers, and in and we understand this term to include general all the people who live in the lands from which we have come. To the Anglo-Saxon race, at least, it seems a matter of course, that those who live in an age of steam, and of electricity, must necessarily be in a different condition as to their nerves, from those who lived in the old slow days of sailing packets, and of mail-coaches. Ours is an age of extreme activity. It is an age of rush. There is no leisure, so much as to eat, and the nerves are kept in a state of constant tension, with results which are sufficiently well-known.

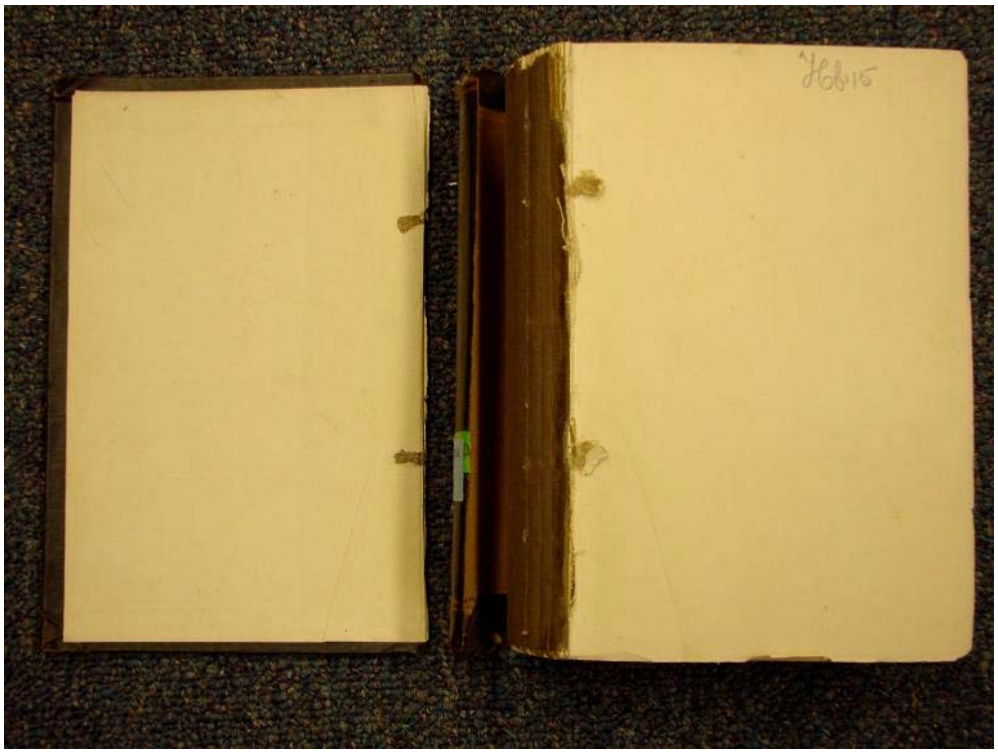
Brittle book- BI, DAM, SPH



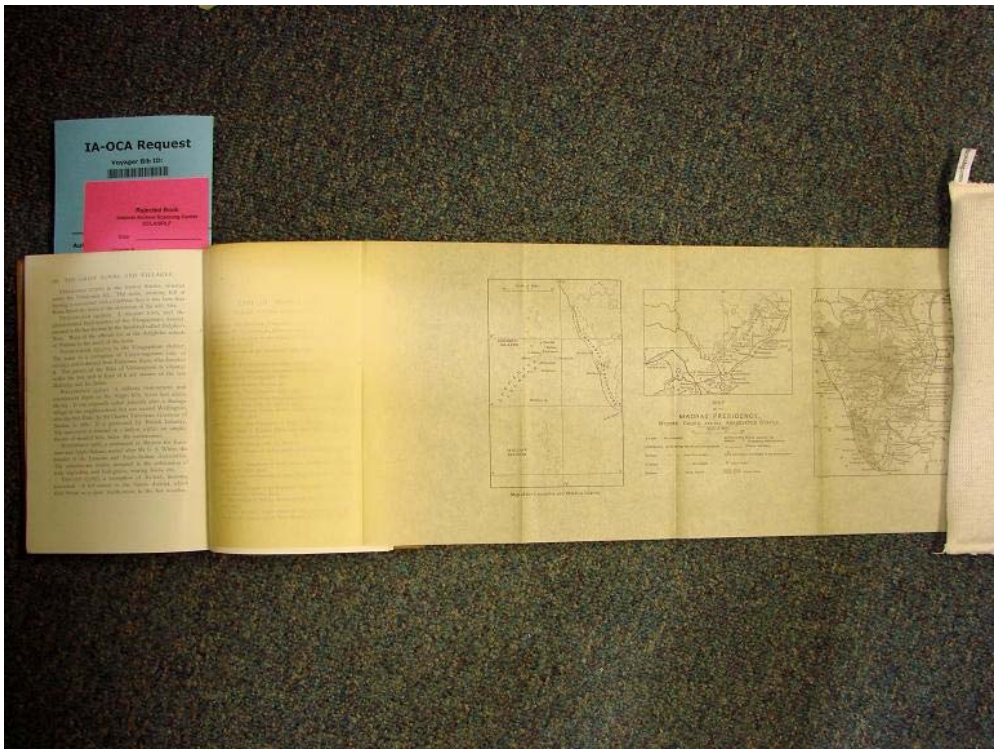
Brittle book- BI, DAM, SPH



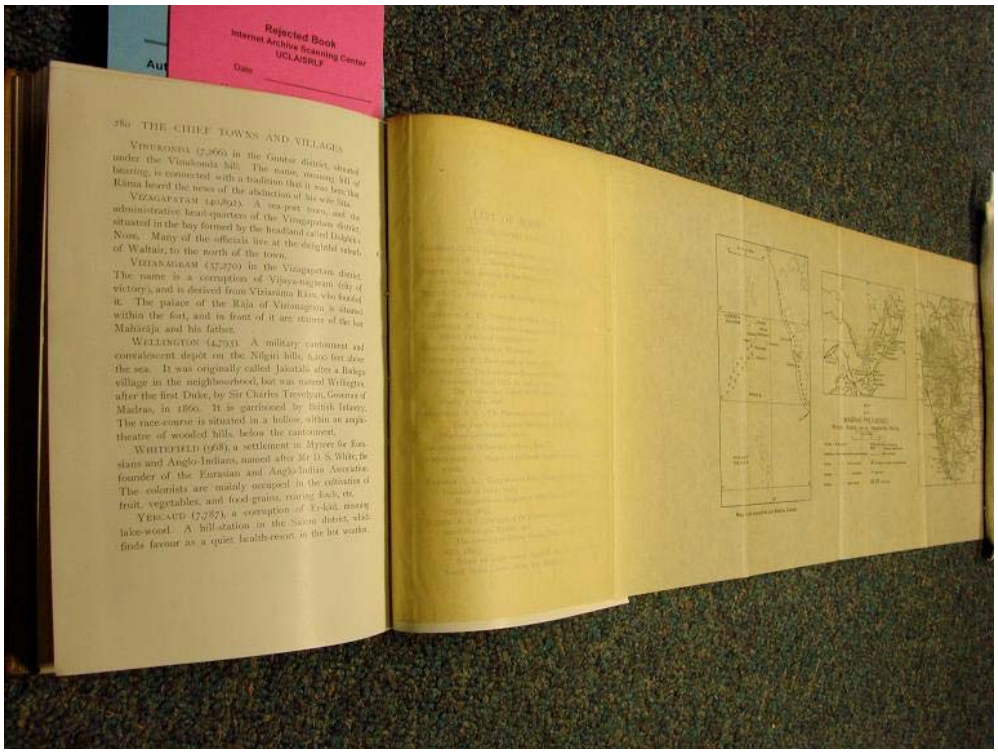
Fold out-FO



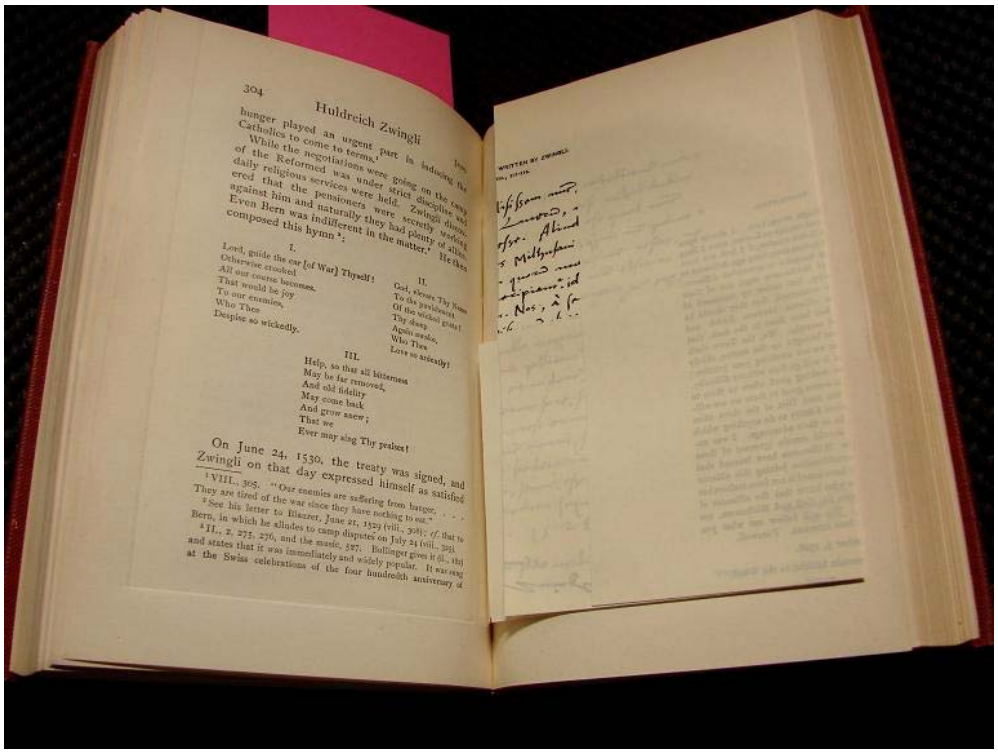
Fold out-FO



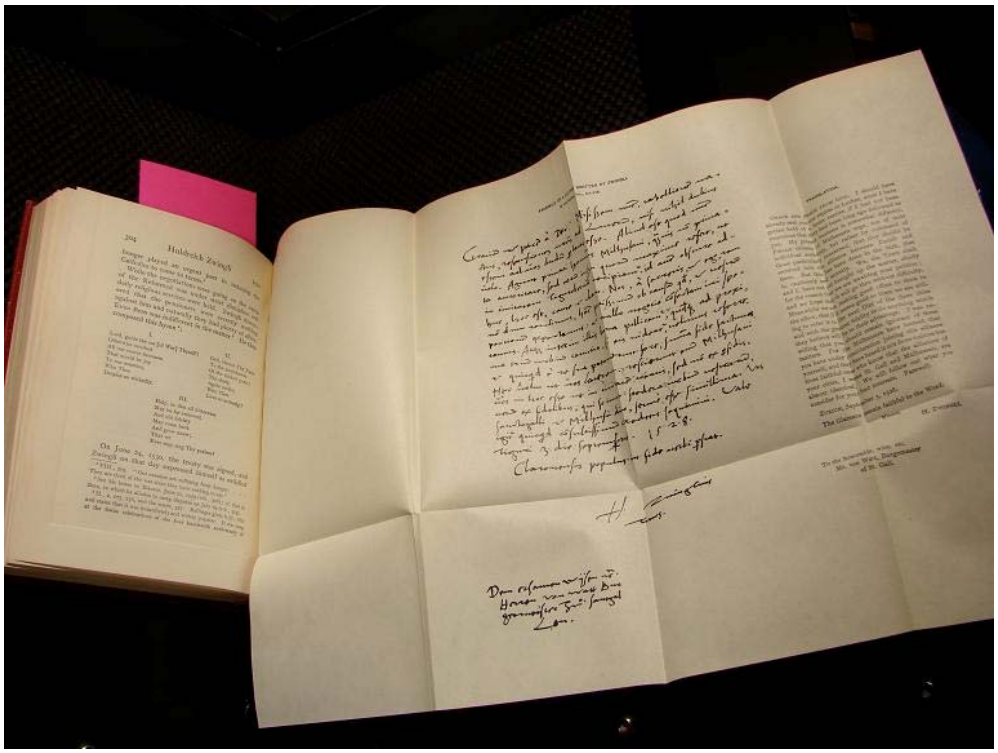
Fold out-FO


Rejected Book
Internet Archive Scanning Center
UCLA/SRLF
Date
280 THE CHIEF TOWNS AND VILLAGES
VINUKONDA (7,266) in the Guntur district, situated under the Vinukonda hill. The name, meaning 'hill of hearing,' is connected with a tradition that it was here that Rāma heard the news of the abduction of his wife Sītā.
VIZAGAPATAM (40,892). A sea-port town, and the administrative head-quarters of the Vizagapatam district, situated in the bay formed by the headland called Dolphin's Nose. Many of the officials live at the delightful suburb of Waltair, to the north of the town.
VIZIANAGRAM (37,270) in the Vizagapatam district. The name is a corruption of Vijaya-nagram (city of victory), and is derived from Vijayārāma Rāzu, who founded it. The palace of the Rāja of Vizianagram is situated within the fort, and in front of it are statues of the late Mahārāja and his father.
WELLINGTON (4,793). A military cantonment and convalescent depôt on the Nīlgiri hills, 6,000 feet above the sea. It was originally called Jakatala after a Baḍaga village in the neighbourhood, but was named Wellington, after the first Duke, by Sir Charles Trevelyan, Governor of Madras, in 1860. It is garrisoned by British infantry. The race-course is situated in a hollow within an amphitheatre of wooded hills, below the cantonment.
WHITEFIELD (908), a settlement in Mysore for Europeans and Anglo-Indians, named after Mr D. S. White, the founder of the Eurasian and Anglo-Indian Association. The colonists are mainly occupied in the cultivation of fruit, vegetables, and food-grains, rearing fowls, etc.
YERCAUD (7,787), a corruption of Er-kād, meaning 'lake-wood.' A hill-station in the Salem district, which finds favour as a quiet health-resort in the hot months.

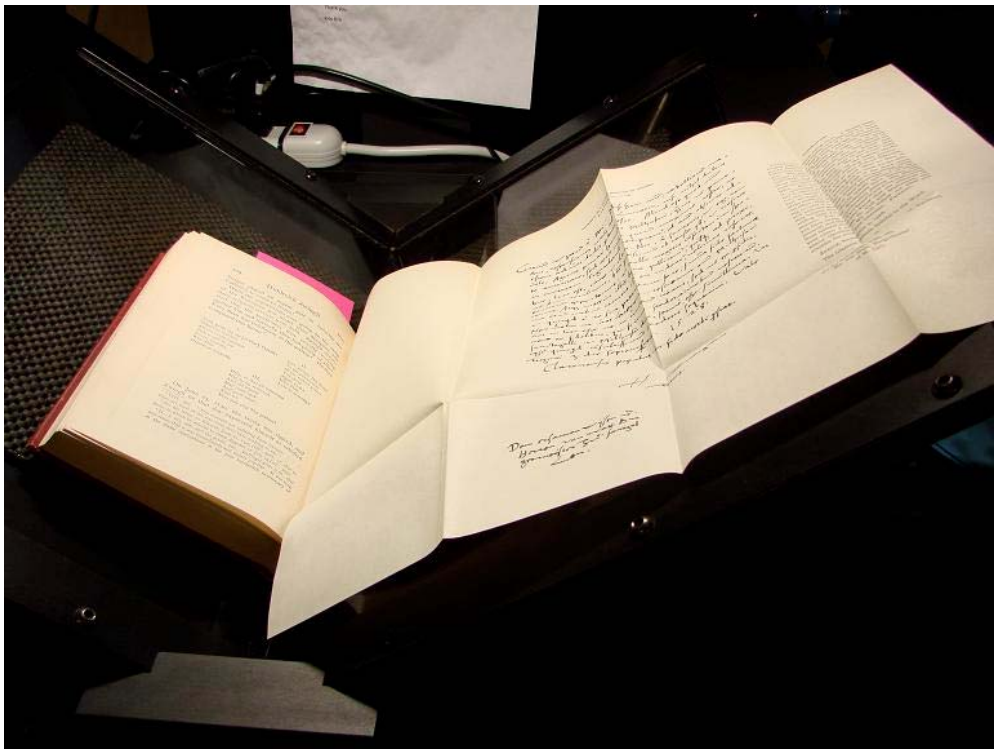
Fold out-FO


304 Huldreich Zwingli
hunger played an urgent part in inducing the
Catholics to come to terms.'
While the negotiations were going on the camp
of the Reformed was under strict discipline and
daily religious services were held. Zwingli directed
that the pensioners were secretly watched
against him and naturally they had plenty of allies.
Even Bern was indifferent in the matter.' He has
composed this hymn ':
I.
Lord, guide the car [of War] Thyself!
Otherwise crooked
All our course becomes.
That would be joy
To our enemies,
Who Thee
Despise so wickedly.
II.
God, Secure Thy Name
To the persistent
Of the wicked grant!
Thy deep
Again arise,
Who Thee
Love so ardently!
III.
Help, so that all bitterness
May be far removed,
And old fidelity
May come back
And grow anew;
That we
Ever may sing Thy praises!
On June 24, 1530, the treaty was signed, and
Zwingli on that day expressed himself as satisfied
1 VIII., 305. "Our enemies are suffering from hunger. . . .
They are tired of the war since they have nothing to eat."
2 See his letter to Blaurer, June 21, 1529 (viii., 308); cf. that to
Bern, in which he alludes to camp disputes on July 24 (viii., 391).
3 II., 2, 275, 276, and the music, 327. Bullinger gives it (ii., 181)
and states that it was immediately and widely popular. It was sung
at the Swiss celebrations of the four hundredth anniversary of

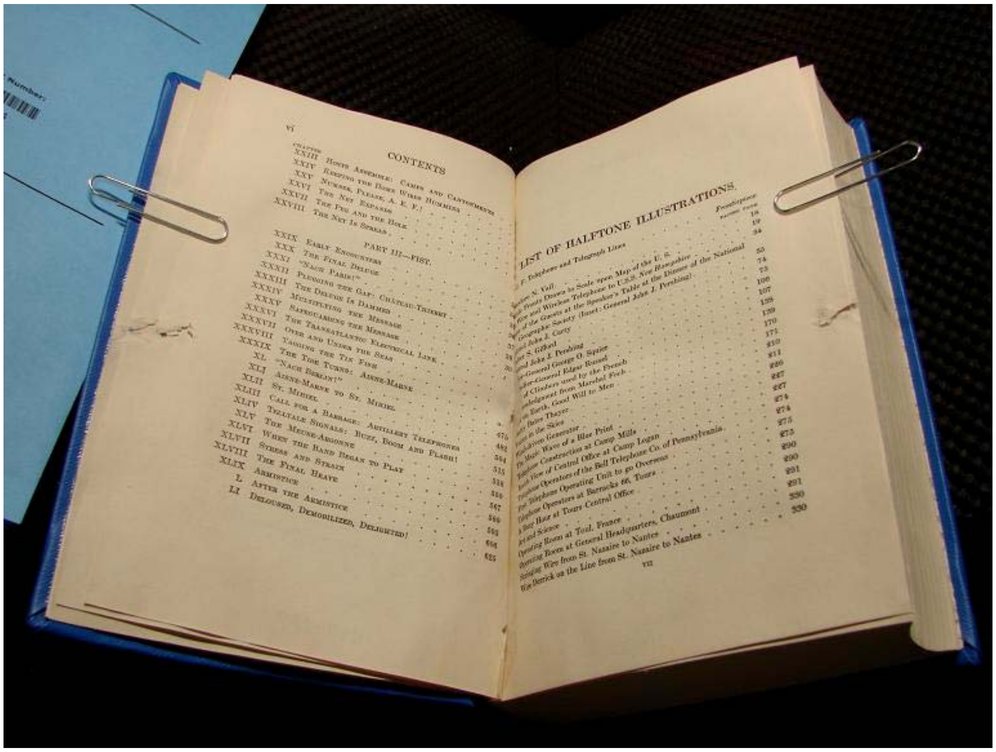
Fold out- FO



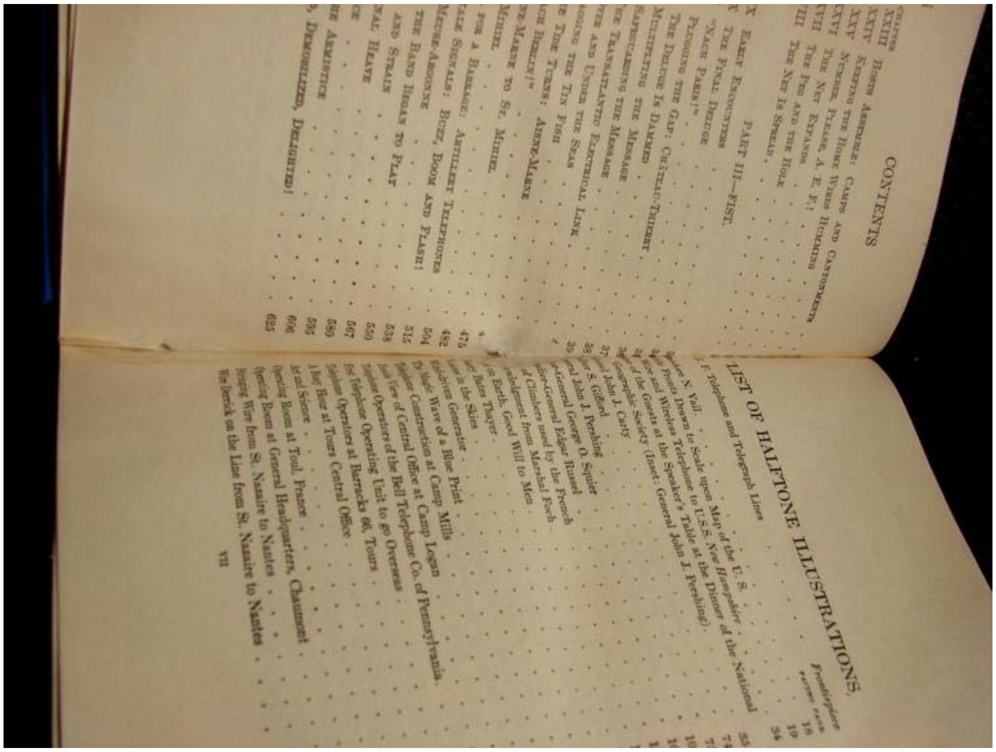
Fold out-FO



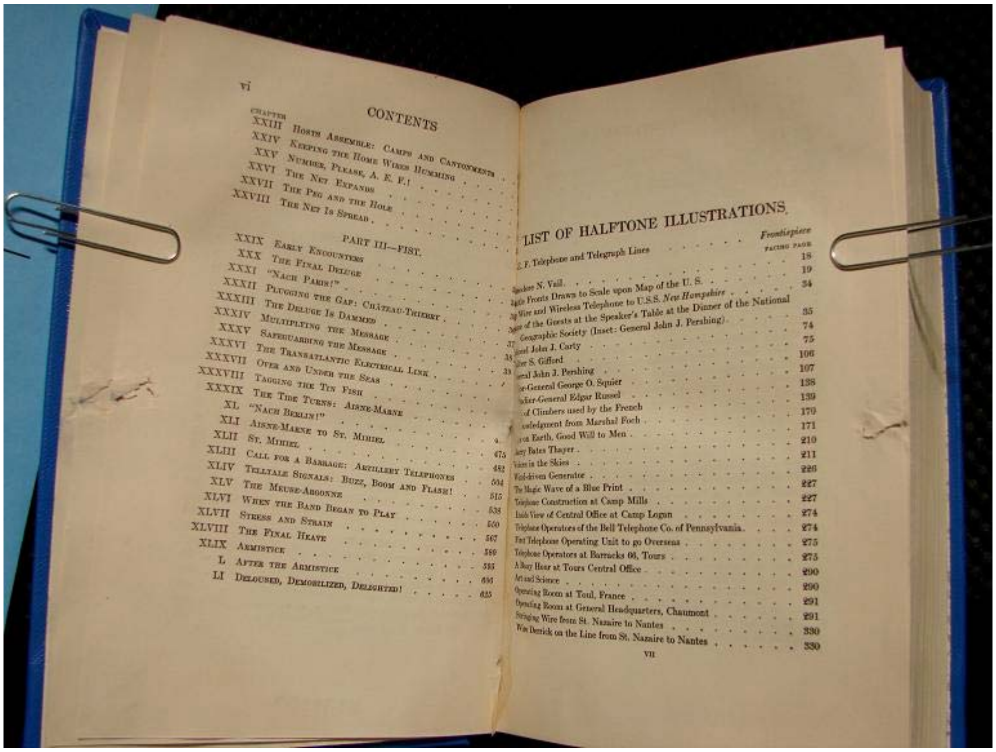
Glued pages- DAM, MIS, PA, PAG, SPH- replace photo with a new shot not using paperclips; paperclips were for photographic purposes only



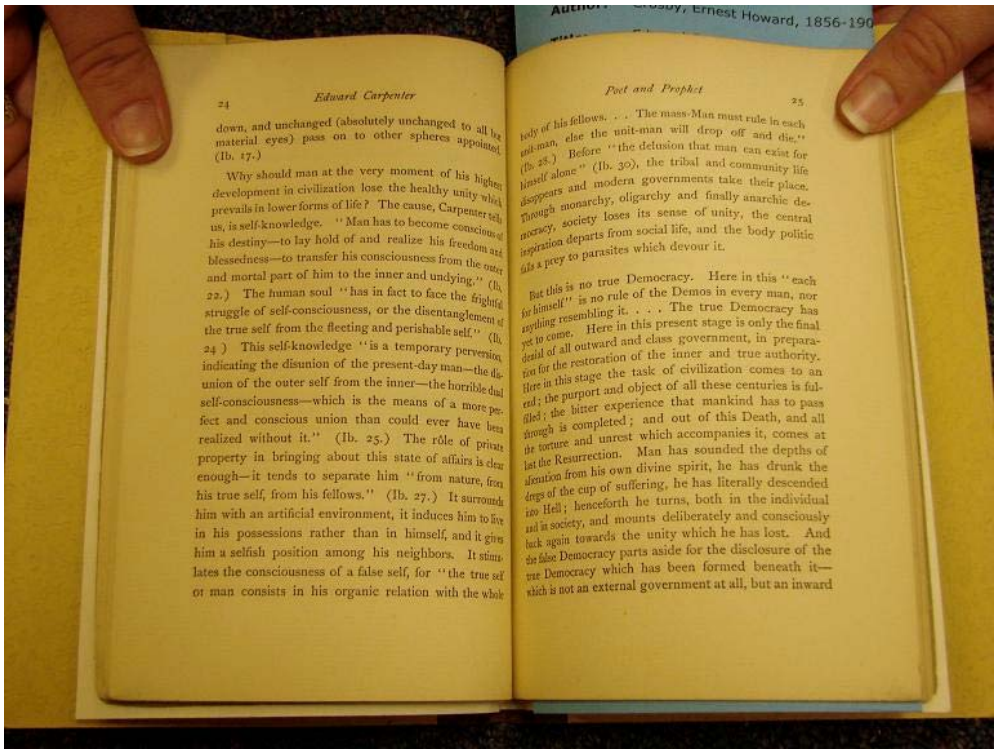
Glued pages- DAM, MIS, PA, PAG, SPH



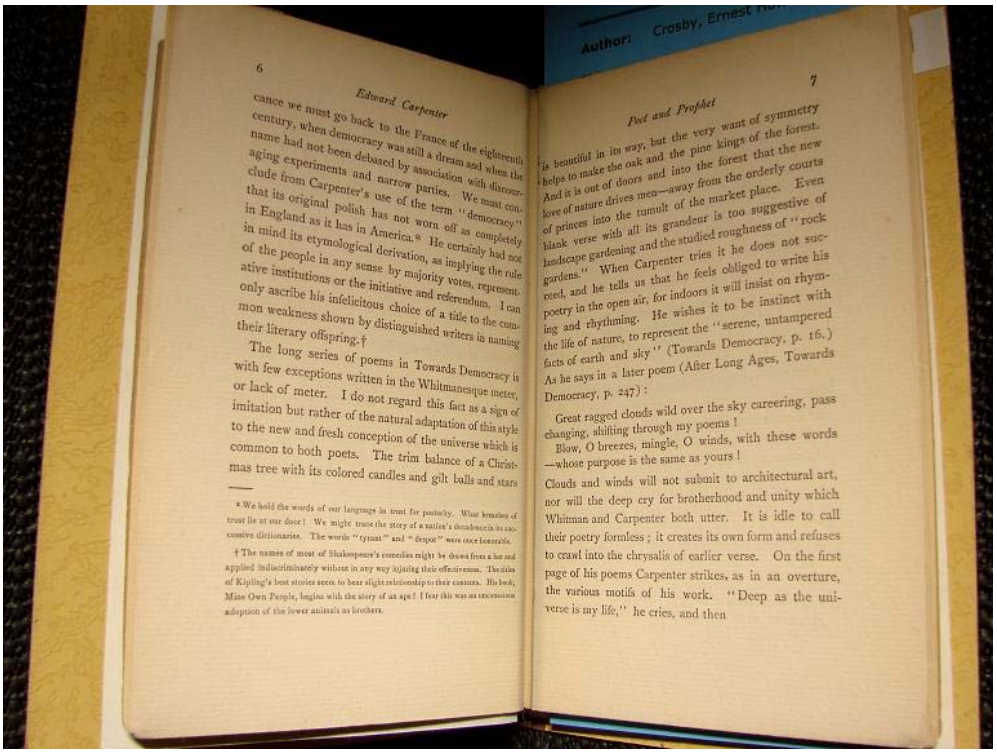
Glued pages- DAM, MIS, PA, PAG, SPH- replace photo with a new shot not using paperclips; paperclips were for photographic purposes only



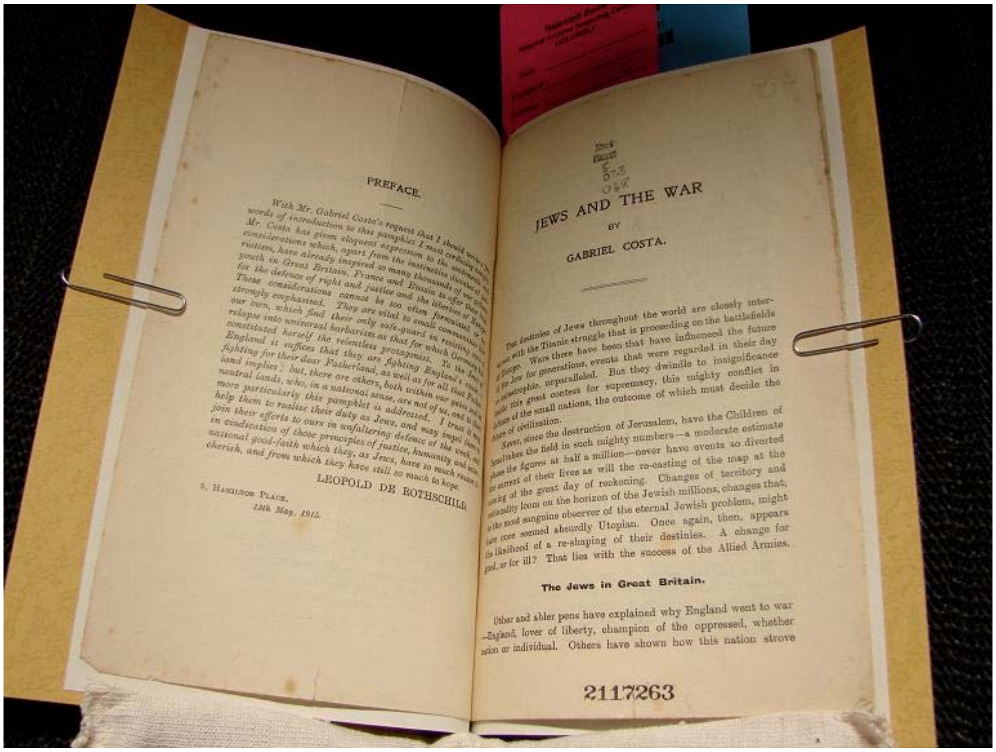
Tight Binding- MAR, PA, PAG, SPH



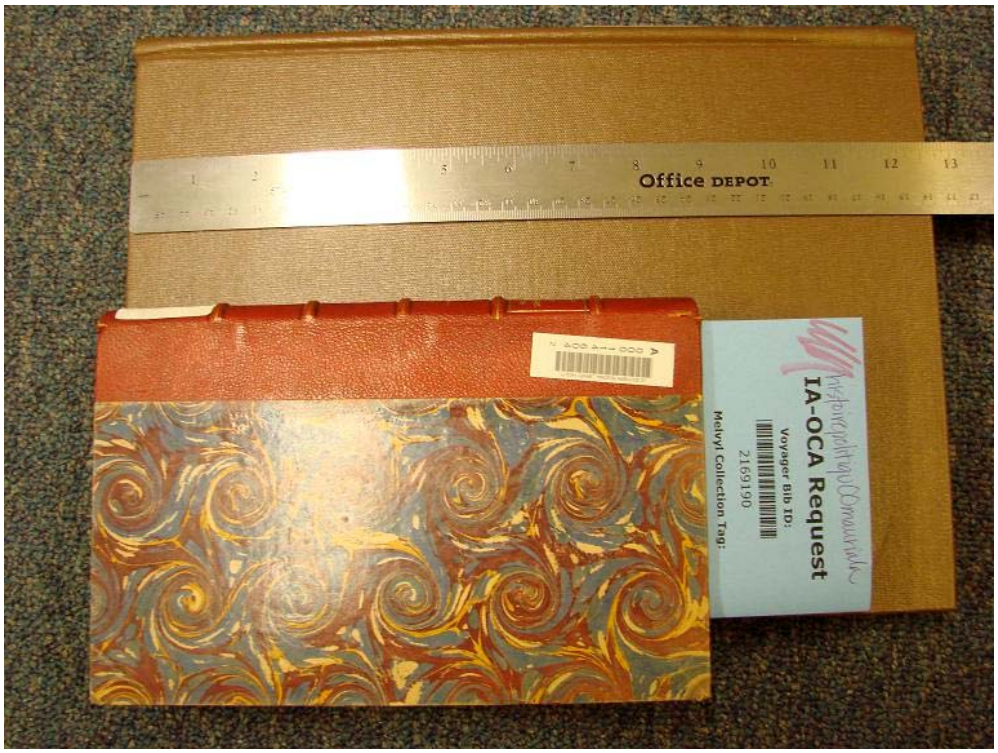
Tight Binding- MAR, MIS, PA, PAG, SPH



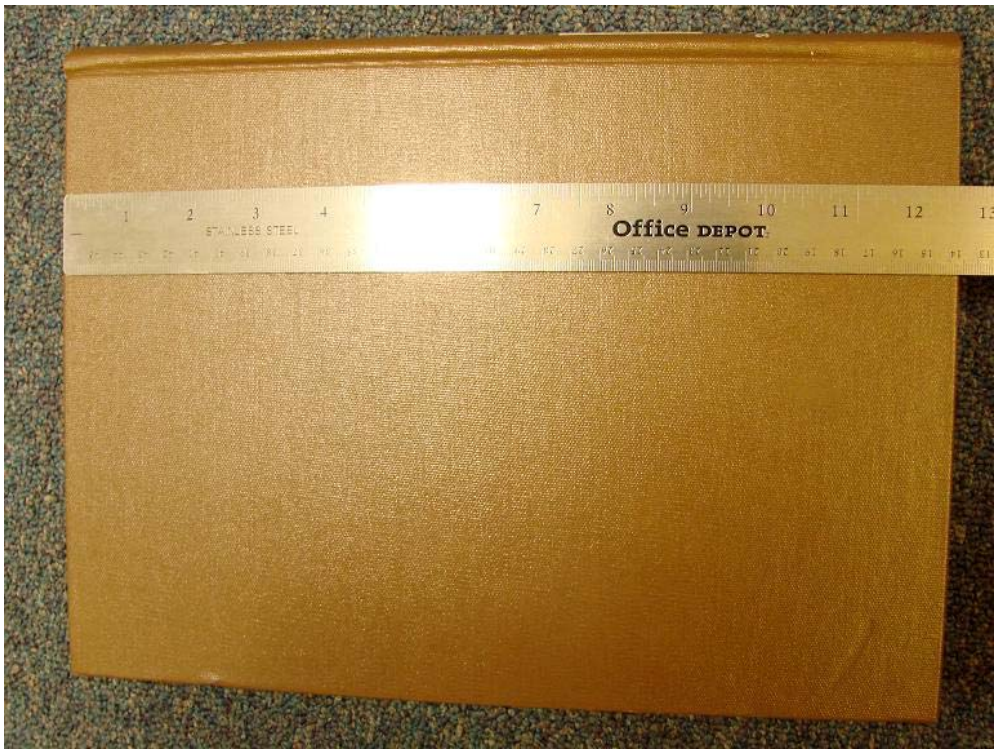
Tight Binding- MAR, MIS, PA, PAG, SPH- replace picture without paperclips (which are for photographics purposes only)



Size too big- LG, SPH



Size too big-LG, SPH



Book to thick- LGH, SPH

Book too thick-LGH, SPH


University of California
Southern Regional
Library Facility
STAINLESS STEEL