

Historic Pittsburgh Full-Text: Page-Level Metadata - Spreadsheet Creation

Revision: August 12, 2002

Please read this entire document before setting up a spreadsheet for the first time

1. General Principles

The completed spreadsheet should represent an accurate record of all the image files that belong to a work. It is created prior to scanning so that each page image created during the scanning process can be assigned a correctly formulated file name. This will enable faithful reproduction in digital form, microform, or in the production of a paper facsimile. Where an item has value as an artifact (for example, a rare book), include any information that identifies the specific copy, or anything which might be of potential importance to a user.

The document structure for Historic Pittsburgh on-line books consists of three basic parts: (1) the production note and front matter, (2) the body of the text, and (3) the back matter. Important elements of the front matter should be identified and appropriately tagged. These usually--though not always--include the title page, the table of contents, a list of illustrations, a list of tables, and a chronology (which may appear in the back matter instead). If the cover contains additional information (such as a title variation) it should be scanned. Important elements included in the back matter are the appendix (es), index (es), bibliography, and glossary. Generally, few books will contain all of these parts, and some may include parts not mentioned. Consult the Digital Production Librarian in doubtful cases. Refer to APPENDIX III. Directory and Document Structure for additional details.

In spite of all precautions and the best of intentions, human errors do occur. Therefore, it is important that someone other than the spreadsheet's author proofread each spreadsheet. If the item is to be scanned in-house, the person doing the scanning should also check the spreadsheet for logic and accuracy. Consult with the author of the spreadsheet if any questions arise. The author's name should appear in the document footer.

2. Initial Setup for a New Spreadsheet

Select a book from the shelf. Ensure that the Digital Products Tracking Sheet included with the book is complete and current. If there is no sheet with the book, see the Digital Production Librarian.

Open Excel. Open the following file:

```
'Ulsdr100'(P):\Histpitt\Hpbooksheets\template\HP_spreadtemp.xlt
```

and immediately save the file using Save As. Save it to directory hpbooksheets as an **.xls** file, not as the default .xlt (Template).

The name of the file will be the ID number with two leading digits (00, 01, 02, etc., where numbers greater than 00 indicate multi-volume sets), followed by an "m" (for monographs) or "s" (for serials and periodicals).

Example:

```
00abc1234m, where abc1234 is the actual ID
```

The ID number is found near the upper left corner of the Digital Products Tracking Sheet. It may be in one of several forms: the old NOTIS number (00abc1234m), a new VOYAGER number (00a123456m), or the Historical Society of Western Pennsylvania's (HSWP) accession number (00hc12345m). Refer to APPENDIX III., Directory and Document Structure for additional details.

Insert the Header and Footer data. Go to FILE --> PAGE SETUP --> and click the HEADER/FOOTER tab. Type the filename in the center column of the HEADER (i.e., 00abc1234m), and type your name (e.g., R. Fletcher) in the center column of the FOOTER. Save the file again.

Be sure that the Autocorrect feature is disabled. Go to TOOLS --> AUTOCORRECT and uncheck all boxes.

Do not use the centering option for any column in the spreadsheet. The defaults are set in the template file. Numbers are right-justified, text is left-justified. The font is Times New Roman, 9-point (preset).

IMPORTANT:

During the process of filling out a spreadsheet, save frequently. Do not depend on the Excel AUTOSAVE feature, which may or may not be enabled.

3. Spreadsheet Columns

Note: This section is arranged by column content. Information specific to each column will be found under the appropriate heading and/or in the Appendices. The first two rows of the spreadsheet are not part of the document or text. They are "dummy pages" required

for production of the acid-free facsimile edition, and must be included on every spreadsheet as shown below:

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0001	000A		DOCPRN	[Complete title, plus single author's first and last name, separated by (\)]	00abc1234m	Omit covers, pastedown, flyleaf, and ("x" number of) blank pages
0002	000B		BLK001		00abc1234m	Verso of DOCPRN

Image sequencing and pagination are contained in the first two columns of the spreadsheet. (Together, these two columns are combined to make up the filename of the page image during the scanning process.) Note that the AccNo column always starts with 0001, and NatPg with 000A. Use uppercase letters in the Accession Number (AccNo), Native Pagination (NatPg), Additional Image (AddIMG), and Type (Type) columns.

3.1 AccNo Column

This sequence always begins with 0001 and increments continuously by one until the last page. There should never be any missing numbers in the numerical sequence in the Accession Number column.

The first column (AccNo) is the actual image file number. The first four digits of the file name indicate the accession number -- i. e., the sequential image number. The number will contain up to three leading zeroes depending on the number required to make a four-digit number (e.g., 0001, 0012, 0103, and so forth all the way up to 9999).

The recto (the front or right-hand page) will always be an odd Accession Number, and the verso (the back or left-hand page) will always be an even Accession Number.

3.2 NatPg Column

The second column (NatPg) is a record of the native pagination found in the book, with certain modifications to allow for automated scanning. The second part of each image file name indicates the native pagination, or lack thereof, plus any variations.

The front matter often contains one or more Roman numeral series. If this occurs, substitute an "R" for the leading digit of the four-digit numbers (i.e. R001, R002, etc.) in the series. Drop a leading zero from the four-digit number to accommodate the "R."

The native pagination may not necessarily start with page one (0001). Pagination begins with the first page that contains any type of information (a frontispiece, dedication, etc.), whether a page number is actually printed on this page or not.

(For example, if the first page that is part of the main text body (DOCBOD) starts at page 9, count pages backward from this page, not including Roman numerals, to find the actual first page (page 1.).

If there are additional (non-blank) pages before Page 1, and Roman numerals are not used, label these with letter sequences (starting with 000C, since 000A and 000B are always used for the first two rows). All pages in the text not specifically omitted in the Notes column, row 1 must be accounted for. See APPENDIX IV. Image File Naming Conventions for more details.

3.3 Figures

Full-page figures within a book are often not numbered, especially when the reverse is blank. In this case, use the preceding page number and add a letter suffix as shown below. (Drop a leading zero from the four-digit number to accommodate the added letter.)

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0045	0049		UNSPEC		00abc1234m	
0046	049A		FIG001	Hillman Library	00abc1234m	
0047	049B		BLK001		00abc1234m	verso
0048	0050		UNSPEC		00abc1234m	

For multiple full-page figures, an example with two facing figures is shown below:

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0045	0049		UNSPEC		00abc1234m	
0046	049A		BLK001		00abc1234m	recto
0047	049B		FIG001	Hillman Library	00abc1234m	
0048	049C		FIG002	Hillman Library Annex	00abc1234m	
0049	049D		BLK001		00abc1234m	verso
0050	0050		UNSPEC		00abc1234m	

3.4 AddIMG Column

This column is used only if a color or grayscale image is required. For normal B/W images no entry is required. Consult Appendix IV., Image File Naming Conventions for details.

3.5 Type Column

The third column (Type) contains all the codes for automatic encoding. See APPENDIX II. Labels for Parts of a Document Structure for a complete listing.

3.6 Blank Pages

Blank pages occurring within the book are included as part of the native -- and logical -- pagination. They are "scanned" using a "dummy page," indicated in the Type column of the spreadsheet by using BLK001. If consecutive blank pages occur, then indicate this by using BLK001, BLK002, BLK003, etc. This enables all automated processes when the book is scanned. It is important to include blank pages. These are required for duplex printing of images when creating a facsimile reprint of the book.

3.7 Multiple Document Structures on the Same Page

In any instance where there may be two or more document structure types on the same page, use an underscore to separate the labels in the Type column (e. g., DOCBOD_CHAPTR or LOI001_LOT001).

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0011	R009		LOI001_ LOT001	Illustrations \\ Tables	00abc1234m	

3.8 Covers

When a cover differs from the title page, or there is no title page, it must be scanned if possible. Use the example shown below (which assumes that the cover is Page 1):

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0001	000A		DOCPRN	[Title as it is in the MARC record]	00abc1234m	Omit covers, pastedown, flyleaf, and ("x" number of) blank pages
0002	000B		BLK001		00abc1234m	Verso of DOCPRN
0003	0001		DOCCOV	[If it is significantly different from DOCPRN entry]	00abc1234m	
0004	0002		UNSPEC		00abc1234m	verso

Confirm any additional bibliographic information or title variation on the cover by checking the MARC record. If the additional information or title variation is noted in the MARC record, do not include the cover in the document structure.

Note: do not include the cover if scanning it would obviously result in an image with poor contrast. This usually occurs when the background -- the color of the book cloth -- is too dark.

3.9 Appendices

If an appendix appears within another appendix, it should be labeled again and not be considered as part of the previous appendix. For example, a page titled "Appendices" occurs, followed by separate sections entitled Appendix I. Biographies, Appendix II. Committees, and so on. Each section should be labeled APPEND in the Type column with its title recorded in the Title column.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0210	0204		UNSPEC		00abc1234m	
0211	0205		APPEND	Appendices	00abc1234m	
0212	0206		APPEND	Appendix I. Biographies	00abc1234m	
0213	0207		APPEND	Appendix II. Committees	00abc1234m	
0214	0208		BLK001		00abc1234m	Last page

3.10 Numbered Document Types

For document types that have the last three characters as numbers (i.e., TPG001, TOC001, IND001), any information in the Title column will only show up on the web document if the number is 001. Anything higher in the series (i.e., IND002) will be ignored by the script. The only exception to this rule is for Figures.

If, for example, there are embedded indices in a work, each must begin with 001, even if no special index tags apply, and the index appears to be continuous in the text;

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0210	0204		PNI001	Volume I Index	00abc1234m	
0211	0205		PNI 002		00abc1234m	
0212	0206		PNI 003		00abc1234m	
0213	0207		PNI 001	Volume II Index	00abc1234m	
0214	0208		PNI 002		00abc1234m	

3.11 Title Column

Enter any relevant chapter numbers and titles, and figure captions or descriptions. Certain SGML tags (in the Type column) always require titles or descriptions in the Title column:

DOCPRN
 TPG001 (only if different from DOCPRN)
 DOCCOV (only if different from DOCPRN)
 CHAPTR
 FIG(001, LDL, LDR)
 UNSFIG
 ADV001
 UNSADV
 TOC001

LOI001
 LOT001
 CRN001
 VOI001
 APPEND
 BIB001
 PNI001
 IND001
 SPI001
 SUI001

Additional information is found in APPENDIX II. Labels for Parts of a Document Structure.

On the first row of the spreadsheet -- designated 0001000A and labeled DOCPRN (production note) -- enter the title of the work as it appears in the MARC record in the Title column. If there is a single author, include the first and last name. Separate it from the title by a space, double back-slash, space, as shown below.

Note: Double back-slashes (\) and double forward-slashes (//) are not interpreted in the same way by the computer. Back-slashes tell the computer to place what follows on the next line down. Forward-slashes are read as punctuation.

Use the standard capitalization convention for titles. Capitalize the first letter of the first word and the first letter of each word thereafter, except prepositions, articles, and coordinate conjunctions. Do not use a period after the title of a document structure (even if it is used in the actual title).

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0001	000A		DOCPRN	Historic Pittsburgh \\ Miller	00abc1234m	

For the remainder of the book, record the actual titles of document structures as they appear in the book. Always include chapter designations when they appear--that is, "Chapter One," "Chapter 1," or "Chapter I." Be sure that the titles match the labels in the Type column.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0403	0401		CHAPTR	Chapter XXVI. Robert	00abc1234m	

There is no label type for divisions of a book larger than a chapter (e.g., "Book Two," "Section One," or "Part Three"). Although there is no need to designate such larger divisions in the document structure, they should be recorded in the Title column for clarity.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0403	0401		CHAPTR	Book Two \\ 	00abc1234m	

				Chapter XXVI. Robert		
--	--	--	--	----------------------	--	--

Any information created or supplied by the author of the spreadsheet should be enclosed in brackets.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0045	0041		FIG001	[Robert]	00abc1234m	

If a page contains more than one document structure, such as a chapter title and a figure, separate each title or caption with a space, two back-slashes, and another space.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0011	R009		CHAPTR_FIG001	Chapter I. \\ Bob	00abc1234m	

3.12 Figure Captions and Titles

A caption is a title or headline used for a figure, plate, etc. A legend is a more complete description of a figure, plate, table, etc., often in the form of a sentence. In the Title column of the spreadsheet, use captions whenever possible (in terms of style, treat a caption the same as a title). It is acceptable to use a caption from a list of illustrations in the front matter, or to supply a caption summarizing the more complete description of a legend, but the latter should be enclosed in brackets.

If a page contains a figure with multiple images with separate captions, separate each title or caption with a space, two back-slashes, and a space.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0011	R009		FIG001	Doris \\ Robert \\ Liz	00abc1234m	

What is included, or not, in figure legends or captions depends on the format of the book. Lengthy titles should be shortened or paraphrased in the spreadsheet author's own words, but must include any searchable terms, such as place-names, or proper names. Any text added by the spreadsheet author that is not part of the book text should be enclosed in brackets ([...]).

- *Figures with captions or legends.*
Record as described above. These will be normally captured during the OCR process.
- *Figures without captions or legends.*

If the subject is not obvious, try to find a description in the text, and paraphrase it. Include any searchable keywords or phrases (for example, [Members of the Fletcher family]). Enclose the text in brackets.

- *Multiple figures with captions within the figure.*
Paraphrase long descriptions, but be sure to include all place-names, proper names, and any other relevant information. Normally OCR will not capture this information since it is part of the figure, and not the text.
- *Multiple figures with captions with all information in the caption or legend.*
Record as described above. These will be normally captured during the OCR process.

3.13 NOTISID Column

This column contains only one piece of information, which is the ID number of the book. The ID number is the same number as the filename (i.e., 00abc1234m, 00a123456m, or 00hc1234m). It is not necessarily a NOTIS number.

3.14 Notes Column

This column contains all pertinent information that needs to be included in the spreadsheet that does not fit elsewhere. Certain non-text pages at the beginning and end of the book require special treatment. The pastedown is the outer leaf of an endpaper that is pasted down to the inside of the front or back cover of a book. The inner leaf of an endpaper (the flyleaf) should not be counted in the number of blank pages. Sequences of blank pages are not scanned, especially at the beginning and end of a book. Indicate the omission of blank pages (and the cover) in the Notes column of the production note (DOCPRN) on the spreadsheet. The standard convention for this should read as follows:

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0001	000A		DOCPRN	Historic Pittsburgh	00abc1234m	omit cover, pastedown, flyleaf, and [x number of] blank pages

Certain parts of a document, such as a preface, introduction, or dedication, are not considered essential elements of the structure and are, therefore, not indicated in the Type column. But, such non-essential parts may be indicated in the Notes column of the spreadsheet to reduce structural ambiguity.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0011	R009		UNSPEC		00abc1234m	Preface

3.15 Two-Page Figures

A figure or plate that covers two pages, verso to recto, should be clearly indicated in the Notes column of the spreadsheet. The double page image will be treated and eventually reprinted as a one-page foldout. We will have to slightly manipulate the native pagination in order to keep the pages in their correct verso or recto position in the digital and reprint versions. Also, flag these double page images with a post-it note or strip of paper so that our vendor (NMT) will be aware of the figure prior to disbinding the book.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0045	0047		UNSPEC		00abc1234m	
0046	047A		BLK001		00abc1234m	added blank verso to maintain pagination
0047	0048		FIG001	Hillman Library	00abc1234m	figure covers two pages, verso to recto. Scan as one recto image.
0048	048A		BLK001		00abc1234m	verso of image
0049	0049		BLK002		00abc1234m	added blank recto to maintain pagination
0050	0050		UNSPEC		00abc1234m	

If a two-page figure is a plate or map that is preceded and followed by a blank page, use the following method (Foldout Figures). If it is tipped-in using a hinge, in such a way as to allow for its removal without damage, then unhinge or remove the plate and treat it as a foldout. Use only one Native Pagination value (page number) for both halves -- i. e., the front -- of the figure and one value for the blank side -- usually the verso -- of the figure.

Note: For binding purposes, the left edge should be tipped-in and the page should be folded to fit the text block.

3.16 Foldout Figures

A foldout map or figure should also be indicated in the Notes column, including the overall dimensions in inches (width-by-height). It is important to note that a foldout map should be scanned as one image, and always spell foldout as one word. Flag the foldout figure in the original book with a post-it note or strip of paper so that our vendor(NMT) will be aware of the figure prior to disbinding the book.

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0046	0048		UNSPEC		00abc1234m	
0047	048A		FIG001	Map of ULS	00abc1234m	foldout map

						(12x10 inches), scan as one image
0048	048B		BLK001		00abc1234m	verso
0049	0049		UNSPEC		00abc1234m	

3.17 Documenting Decisions

When you make any choices, document your decision in the Notes column of the spreadsheet so that others will know what you did and why. Convey any information you believe is important.

4. Last Pages

At the end of a book, the last page must be even-numbered. If the last page is odd-numbered, a dummy page must be added after the page and documented in the Notes column, as shown in the first example below:

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0299	0275		UNSPEC		00abc1234m	last page
0300	0276		BLK001		00abc1234m	verso of last page

If the last page is even-numbered, follow this example:

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0299	0275		UNSPEC		00abc1234m	
0300	0276		BLK001		00abc1234m	last page

The rear covers of pamphlets or books are sometimes part of the document structure and therefore need to be scanned. A DOCCOV entry near the beginning of the spreadsheet may indicate this. If the back cover is odd-numbered, follow this example:

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0299	0275		FIG001	[Title]	00abc1234m	recto of back cover
0300	0276		FIG002	[Title]	00abc1234m	back cover

If the back cover is even-numbered, follow this example:

AccNo	NatPg	AddIMG	Type	Title	NOTISID	Notes
0299	0275		FIG001	[Title]	00abc1234m	
0300	0276		FIG002	[Title]	00abc1234m	back cover

Verify that the last cell of the spreadsheet is actually the one that contains one of the structures above. When finished with the spreadsheet, press CTRL + END. This activates

the last cell of the spreadsheet. If the last cell does not contain one of the above notations, extra cells have somehow been incorporated into the spreadsheet. Delete any groups of cells that occur after the "last page," or "verso of last page," notation.

Note: In some cases (for reasons known only to now-deceased Microsoft programmers) these extra cells cannot be deleted by normal means. If this occurs, copy (minus the extra cells, of course) and paste the spreadsheet to a new template. Save as the same filename (overwrite the old file in \hpbooksheets). Also verify that both AccNo and NatPg are even numbers in the last cell. Generally, there should be no odd numbers in this last line.

An odd number is a good indicator that versos and rectos are off at some point in your spreadsheet.

On completion of the spreadsheet, proofread the document again. Run the spell checker found in the Tools menu. This may be cumbersome at first, but ADD odd terms to the dictionary (such as those found in the Type column, plus proper names and place names). Eventually the process will become faster as the spell checker's dictionary accumulates these standardized terms. Sign off on the Digital Products Tracking Sheet with your initials and the date. Notify the Technical Projects Manager.

APPENDIX I. Conditions and Problems with Native Pagination

Production note	Solution: These pages are added to every scanned title. The DOCPRN is the first file of every work, and the "blank" DOCPRN verso (BLK001) is always the second file. Note: See production note example attached to Guidelines for Historic Pittsburgh.
Unnumbered front matter that clearly is not a part of any logical sequence of the native pagination	Solution: For a half-title page or dedication page that is the first page of a book, and does not display a page number, use 000C. Find the first numbered page. If it is not Arabic numeral 1, Roman numeral I or i, logically assign page numbers backward until you get to page one. Pages not included should be designated in sequence with letters of the alphabet, starting with 000C.
Roman numeral pagination	Solution: For page ix use R009, for page III use R003, and so forth. If a plate was inserted between page iii and page iv, use R03A and R03B. Start with R, followed by Arabic numerals. If unnumbered pages occur within a roman numeral sequence, append a letter to the preceding numbered page and remove a zero.
Unnumbered pages and plates clearly not part of the logical sequence of the native pagination, on either side	Solution: For an unnumbered plate between pages 32 and 33, use 0032A and 0032B (for the recto and the verso of the plate). Use the preceding page number and append the letter A, then proceed alphabetically with succeeding pages. Eliminate one zero for each letter used. See under "Roman numeral pagination" if the unnumbered page is in a roman numeral sequence.
Incorrect page number displayed versus the correct	Solution: Use logical pagination, not what is on the

and logical pagination	piece. For example, if there are two pages between pages 6 and 9 that display the number 8, the page following page 6 would be called page 7 by inference.
A.6 Page numbering sequences that are duplicated in a single volume	Solution: Number the longest sequence as usual, and assign a letter prefix for each subsequent sequence. Use 0001, 0002, 0003, and then A001, A002, B001, B002, etc. Only the longest sequence will be without an alphabetic prefix letter. Exclude the letters I, L, O, and R.
Pagination exceeds four characters (e.g., 1000A and 1000B, or A100A and A100B)	Solution: Use the letter Z as a substitute for 1[000] and drop the appended letter, or just drop the appended letter Z000 and Z000, or A100 and A100. Even though the second part of the file name for several images may have the same four characters, the accession number -- or the first part of the file name -- will still be different
Pagination with characters not permitted in ISO9660	Solution: Use underscoring in place of any illegal characters. For page 22.6 use 22_6. Ignore adornment characters such as asterisks.

APPENDIX II. Labels for Parts of a Document Structure (Arranged Alphabetically)

A zero-padded three-digit number representing a related sequence of document structures follows some of the labels listed in the table. Please observe that some document structures, such as a table of contents or a list of illustrations, may not always be titled as such (e. g., a list of illustrations may simply be titled "Figures").

When in doubt about the document structure label, use the default value UNSPEC. SGML tags are applied only to those parts of a document's structure that are considered important to helping the user navigate through the work as it is displayed on the World Wide Web. For example, after the table of contents, a preface may be found in the front matter. But while the table of contents would be helpful to the user, it is doubtful he or she would find the preface to be of much use. Therefore, we specifically note the table of contents (TOC001) and label the preface "UNSPEC."

DOCUMENT STRUCTURE	ADV001	YES	EXPLANATION
Advertisements			Use when there is a full-page display of advertising in a monograph. Note the person(s) or company being promoted in the Title column.
Advertisements combined with text, that also might require an additional image for color or grayscale	UNSADV	YES	Use for pages where an advertisement is embedded in the OCR text, and also meets the criteria for an additional gray scale or color image. This designation is a combination of UNSPEC and ADV001.

Appendix	APPEND	YES	Use for the first page of an appendix. List the title of the appendix. Second and subsequent pages of the same appendix are listed as UNSPEC or other labels as applicable.
Back matter Note: This includes any reference material that appears after the body of the text.	DOCBAC		Use to indicate the first image of any back matter. Combine DOCBAC with the label for the structure that begins the back matter - e. g., use DOCBAC_IND001 if it's an index, etc.
Bibliography	BIB001	YES	Apply to bibliography for the whole volume or issue; do not apply to bibliographies and footnotes at the end of chapters or articles. Number each page of bibliography in consecutive order (ie., BIB001, BIB002, BIB003, etc).
Blank pages	BLK001		When blank pages are to be scanned, use the label BLK001, regardless of where the blank page is located. Use BLK002, BLK003, etc. (when there are two or more consecutive blank pages).
Body Note: This is ABSOLUTEY REQUIRED!!!	DOCBOD		Label the first page following the front matter with DOCBOD. If it is the same as the first page of the first chapter, combine it with the label for chapter - i. e., DOCBOD_CHAPTR.
Chapter	CHAPTR	YES	Use this label for the first page of each chapter
Chronology	CRN001		For a chronological list of events, use CRN002, etc., if more than one page. Note: A chronology may also appear in the back matter.
Copyright page (or notice)	TPG002 or UNSPEC		Use TPG002 if the copyright page is the verso of the title page (this is often the case); otherwise, consider it front matter that does not require unique designation and use UNSPEC.
Dedication	UNSPEC		This is front matter and does not require a specific label.
Document Cover	DOCCOV	Only if title is significantly different than DOCPRN entry	Use if there is an illustration or image on cover. [if the image is too dark for scanning, don't bother]. Also use when the cover information is significantly different than the title page or if there is no title page.
Errata	ERRATA	YES	Use to indicate an "errata sheet" or "addenda" that is often tipped-in, and which corrects an error or omission that would cause an misunderstanding otherwise. This is not a usual part of the book.

Figures (illustrations, plates, charts, maps, etc.) on a whole page in landscape orientation (left)	FIGLDL	YES	Use for full-page illustrations that need to be rotated 90 degrees to the left for correct Web display.
Figures (illustrations, plates, charts, maps, etc.) on a whole page in landscape orientation (right)	FIGLDR	YES	Use for full-page illustrations that need to be rotated 90 degrees to the right for correct Web display (this is often the case).
Figures (illustrations, plates, charts, maps, etc.) on a whole page in portrait orientation - the orientation for Web display	FIG001	YES	Use for full-page or multi-page figures. Use FIG002, FIG003, etc. for a series of figures (or plates) that are not separated by blank pages. Do not use for figures embedded within the text of a page.
Figures combined with text, that also include a caption and might require an additional image for color or gray scale	UNSFIG	YES	Use for pages where an illustration with a the OCR text, and meets the criteria for an additional B/W, gray scale or color image. This designation is a combination of UNSPEC and FIG001.
Foreword	UNSPEC		This is front matter that does not require a specific label.
Glossary	UNSPEC		This is most often back matter that does not require a specific label.
Half-title page	UNSPEC		This is front matter that does not require a specific label.
Index, author or name	PNI001	YES	Any personal name index, use PNI001, PNI002, etc.
Index, general or comprehensive	IND001	YES	The default label for indexes not otherwise specified, use IND001, IND002, etc.
Index, special	SPI001	YES	Examples of a special index would include a place name index, a species index, and so forth. Note: Special indexes can sometimes appear in the body of a book.
Index, subject	SUI001	YES	Other titles include "Topical Index," use SUI001, SUI002, etc.
Index, volume	VOI001	YES	Use for an index at the end of each volume of multi-volume work.
Introduction	UNSPEC		This is front matter that does not require a specific label.
List of illustrations	LOI001	YES	Use LOI001, LOI002, etc., if more than one page.
List of tables	LOT001	YES	Use LOT001, LOT002, etc., if more than one page.
Notes	UNSPEC		This is back matter (or matter which may appear at the end of a chapter) that does not require a specific label.
Preface (or acknowledgments)	UNSPEC		This is front matter that does not require a specific label.

Production note	DOCPRN	YES	This is a "dummy" page containing information about the electronic or facsimile edition (it will always be followed by a DOCPRN verso labeled BLK001). The Title column always includes the title from the spine, cover, or title page. It can be incorporated into the TEI header later.
Publisher's device	UNSPEC		Any embellishment (colophon, etc.) that may appear on the last page, it does not require a specific label.
Table of contents	TOC001	YES	Use TOC002, etc., if the table of contents is more than one page.
Tables	UNSPEC or FIGLDL/LDR		This is usually text that does not require a specific label. If the table is not vertical, treat it as a figure that needs to be rotated for correct Web display.
Title page		TPG001	YES NOTE: Only if title page contains essential information not on the cover or spine. Use TPG001 for the title page and TPG002 for the TPG verso, which will often be the copyright page, unless it is blank. Note: Half-title pages are labeled UNSPEC.
Unspecified pages	UNSPEC		Use this label for any pages not otherwise covered in this table.

APPENDIX III. Directory (Books and Serials) and Document Structure

Directory Structure

The directory structure is intended to allow for as much automated processing as possible. To that end, absolute consistency is expected to result in better quality work. The directory structure is suitable for single-volume and multi-volume works. Single-volume works have empty subdirectories for volumes. Each title in the Historic Pittsburgh Full-Text Collection bears a unique designation called IDNO in the header for the electronic edition (hereafter, called "header") and in all directories where data pertaining to the item are stored. The first two digits represent volume numbering. Two zeros mean an item is a single volume monograph, for example. The last letter is either an "m" (for monograph) or "s" (for serial or periodical). The root identifier directory -- i. e., the first file folder -- will ordinarily consist of the following three parts:

1. A prefix of 00 for single-volume monographs or 01, 02, 03, etc. for multi-volume monographs or serials;
2. A seven-digit identification number, which consists of a combination of numbers and letters. The NOTIS ID is from the old ULS NOTIS database. The new VOYAGER number is used for titles acquired after implementation of VOYAGER.

The HSWP accession number, or other unique designation, is created for works not owned by the ULS.

3. A suffix of m for monographic works or s for serials and periodicals.

There are several categories of ID numbers (IDNO) in the Historic Pittsburgh project

- Single volume monographs where the print copy was cataloged for ULS in NOTIS:
Example: 00akc9350m
We take the seven-digit NOTIS record number and put it in the middle.
- Multi-part set cataloged for ULS in NOTIS
Example: 04adc9345m
This is read as volume 4 of a multi part set, cataloged as a monograph. The TEI-like header would be the collective record for the set, but with volume numbering added to the title for display on the browse list. Other volumes would have the IDNO of the form:
01adc9345m for volume 1 02adc9345m for volume 2, etc. We don't record the extent for the volume, but leave the extent as:
4 v. ill. 24 cm., for example. Volume of a serial cataloged for ULS in NOTIS:
Example:45akf8435s
This is read as volume 45 of a serial or the 45th volume of a serial. If the serial has no apparent volume numbers, we infer the volume number from the catalog record and any holdings information we can locate.
Sometimes the implied volume number is arbitrary.
- Items from the HSWP collection:
 - o 00hc10333m for a single volume monograph
 - o 03hc00345s for a volume of a serialThe first two digits represent volume numbering. The hc stands for HSWP (the History Center). The five digits following hc are the accession number for the cataloging record derived from the HSWP catalog. This number does not show up in the public version of the HSWP Catalog on the Historic Pittsburgh Web site. It is accessible through a copy of the catalog on the development server, Nessie.
- Items cataloged for ULS after the Voyager implementation:
Because the bibliographic record number in Voyager is all numbers, the DRL has devised a way to separate the volume numbering digits from the rest of the IDNO. We are using the first nine alphabets to represent the first number of the Voyager bibliographic record number:

a=1
b=2
c=3

d=4
e=5
f=6
g=7
h=8
i=9

So, a single volume monograph acquired since the Voyager implementation would be called: 00c234567m (where the VOYAGER number would be 3234567).