

Guidelines on File Naming Conventions for Digital Collections

The following guidelines were implemented on March 4, 2008 to facilitate the selection of file naming conventions. Digital Projects Advisory Group does not advise retrospectively changing file names for pre-existing collections.

A file naming convention is a set of agreed-upon rules used to assign identifiers to digital objects in a collection. A good file naming convention ensures consistency and uniqueness and should be established before scanning begins.

It is important to know the scope and details of the collection, prior to establishing a file naming convention. Take the size of the collection into account; the naming convention should allow for unique file names for each item. If the file naming convention will be based on some characteristic of the collection, identify outliers that don't fit the pattern and develop rules to deal with them. For example, if file names will be based on items' accession numbers, look for items with non-unique accession numbers. Determine in advance how these items will be distinguished, such as adding "_1", "_2", etc. to the end of the non-unique accession numbers.

File names can be meaningful or non-descriptive. Meaningful file names contain words, numbers, or abbreviations that bear some relation to the item being digitized; they may be composed of an abbreviated title, the accession number of the physical item, or some other descriptive identifier. Meaningful file names work best for medium to small collections and make it easier to identify and manage the digital files. Meaningful file names don't work as well for large collections because it is more difficult to create unique names for all items. It is also possible that the name's meaning will be lost or change connotation over time. Non-descriptive file names express no relationship to the item and are usually sequential numbers. Non-descriptive file names work well for medium to large collections and are easy to assign. Non-descriptive file names provide no identifying information. Thus, the files are harder to manage outside of the database that contains the associated metadata. The decision to use meaningful or non-descriptive file names should be based on the collection's characteristics and project specifications.

Use the following guidelines to develop a file naming convention:

1. Use 8 or fewer characters in the file name, whenever possible.
Eight character file names are backwards compatible with older operating systems. Shorter file names are easier to read and reduce the potential for error when typing the file name.
2. Use a 3 character file extension (i.e. ".tif", not ".tiff").
3. Use only alpha-numeric characters, except for dashes, - , and underscores, _ .
Dashes and underscores may have special uses in other operating systems or software, such as UNIX and MySQL. DPAG reviewed the contexts in which these characters are reserved and determined that dashes and underscores can be used in file names without a technical conflict.

4. Do not use special characters, such as, . \ / : * ? " < > |, except for dashes or underscores.
These characters are often reserved for use by the operating system.
5. All letters should be lower case.
Some operating systems are case sensitive. Using lower case consistently prevents problems, if the files are migrated to a case-sensitive operating system.
6. Do not use spaces in the file name.
Browsers and some older operating systems do not handle spaces well.
7. Use leading zeros.
If the file name includes numbers use zeros as placeholders. For example, a collection with 999 items should be numbered: mac001.tif, mac002.tif ... mac011.tif, mac012.tif, etc. (NOT mac1.tif, mac2.tif ...). This practice facilitates sorting and file management.

New file naming conventions should be recorded to document decisions and assist file management. Send the file naming convention documentation to the Digital Projects Advisory Group, so that it can be saved and posted on the [Resources for Building Digital Collections at UCB University Libraries](#) web page.

All documentation should include:

- A. Collection
Record the name and location of the digital collection(s) that use the convention, as well as the physical collection from which the digital collection was derived.
- B. Date and author(s)
Record the date that the convention was created. List the people who participated in creating the convention.
- C. Rules used to create file names
List all rules used in the convention. Don't forget rules that apply to outliers in the collection.

Example:

Rules

1. File name is 8-9 characters long with a three digit file extension.
2. File name consists of numbers and lower case letters only.
3. The base file name is derived from the city, date, and page number printed on the map.
4. The city is a three letter code created by the Map Library specifically for this digital collection.
5. The date is the last two digits of the year in which the map was drawn. The date is printed on the map.
6. The page number is printed in the upper right corner of the map sheet. The page number in the file name is a three digit number. "001", "002" ... "021", "022" ...
7. Some maps include one or more unnumbered sheets. These files are denoted with "un1", "un2", etc. in place of a three digit page number.
8. Revised map sheets have the same city, date, and page number as the original. To maintain unique file names, an "r" for revised is placed between the date and page number on all revised map sheets.

D. Structure of file names

Describe the structure of the file name and provide examples

Example:

Structure:

cccy[r]ppp.ext, where:

ccc = City code designated by the Map Library

yy = Last two digits of the year printed on the map

r = The letter "r" designates revised maps. [Optional]

ppp = Three digit page number printed in the upper right corner of the sheet or
"un" + sequence number for unnumbered sheets

Examples:

asp98003.tif

Map of Aspen, 1898, numbered sheet 3

cos07un3.tif

Map of Colorado City, 1907, third unnumbered sheet

sal86r001.tif

Revised map of Salida, 1886, numbered sheet 1

Documentation for meaningful file names should include:

E. Attribute(s) used to create file names

Describe which attribute(s) were used to name the files and where to find the attributes

Example:

Attributes:

The base file name is composed of the city code, date, and page number. Some maps also have a "revised" attribute.

City code: The city can be found on the title page, and index printed on each map. The corresponding code for the city is listed in this document.

Date: The date can be found on the title page, and index printed on each map.

Page number: If a sheet is numbered, it will appear in the upper right corner.

Otherwise, the sheet is designated "un" for unnumbered, followed by a single digit number that is assigned sequentially.

Revised: If a map is revised, it will be noted on the map, "Corrected Jan 1888"

F. Codes (if used)

If codes or abbreviations are used, provide a complete list of the codes/abbreviations and their meanings.

Example:			
<u>City Codes:</u>		<u>Other Codes:</u>	
ala	Alamosa	r	Revised sheet
ant	Antonito	un	Unnumbered sheet
arv	Arvada		
...			
wla	West Las Animas		
yam	Yampa		
yum	Yuma		

Documentation for non-descriptive file names should include:

E. First and last file name

If file names are in a consecutive order, record the first and last file name used.

File naming conventions may have already been determined for a collection. For example, grant funded projects may place stipulations on file names if the digital content is contributed to a collaborative initiative. File naming conventions may already exist for on-going scanning projects. Check with the granting agency and review the existing file naming conventions documented on the [Resources for Building Digital Collections at UCB University Libraries](#) web page before creating a new convention.

Related Resources

[Millennium Media Guidelines](#)

File naming guidelines have already been established for digital files that will be displayed in the catalog using Innovative's Millennium Media tool. See Part 1: Technical Considerations Section III.