

FILE NAMING STRATEGIES TO AID IN SUCCESSION PLANNING

By Leigh Grinstead, Catalyst Fund Program Lead and Senior Digital Services Consultant,
Lyrrasis

During the Listening Sessions and Survey for the Cultural Stewardship Succession Planning Initiative, project leaders heard many questions about methods for naming digital files so that they would not get lost when a staff member left, and would be easier to find from project to project. The information below can be utilized to assist your organization in developing file naming conventions.

IDENTIFIERS AND FILE NAMING

A file-naming scheme should be established prior to digital capture. The development of a file naming system should take into account whether the identifier requires machine- or human-indexing (or both – in which case, the digital file may have multiple identifiers). File names can either be meaningful (such as the adoption of an existing identification scheme which correlates the digital file with the source material), or non-descriptive (such as a sequential numerical string). Meaningful file names contain metadata that is self-referencing; non-descriptive file names are associated with metadata stored elsewhere that serves to identify the file. In general, smaller-scale projects may design descriptive file names that facilitate browsing and retrieval; large-scale projects may use machine-generated names and rely on the database for sophisticated searching and retrieval of associated metadata.

RECOMMENDED CHARACTERISTICS OF FILE NAMES

- Are unique - no other digital resource should duplicate or share the same identifier as another resource. In a meaningful file-naming scheme, names of related resources may be similar, but will often have different characters, prefixes, or suffixes appended to delineate certain characteristics of the file. An attempt to streamline multiple versions and/or copies should be made.
- Are consistently structured - file names should follow a consistent pattern and contain consistent information to aid in identification of the file as well as management of all digital resources in a similar manner. All files created in digitization projects should contain this same information in the same defined sequence.
- Are well-defined - a well-defined rationale for how/why files are named assists with standardization and consistency in naming and will ease in identification of files during the digitization process and long afterwards. An approach to file naming should be formalized for digitization projects and integrated into systems that manage digital resources.
- Are persistent – files should be named in a manner that has relevance over time and is not tied to any one process or system. Information represented in a file name should not refer to anything that might change over time. The concept of persistent identifiers is often linked to file names in an online environment that remain persistent and relevant across location changes or changes in protocols to access the file.
- Observant of any technical restrictions – file names should be compliant with any character restrictions (such as the use of special characters, spaces, or periods in the name, except in front of the file extension), as well as with any limitations on character length. Ideally, file names should not contain too many characters. Most current operating systems can handle long file names, although some applications will truncate

file names in order to open the file, and certain types of networking protocols and file directory systems will shorten file names during transfer. Best practice is to limit character length to no more than 32 characters per file name.

GENERAL GUIDELINES FOR CREATING FILE NAMES

- It is recommended that an institution use a period followed by a three-character file extension at the end of all file names for identification of data format (for example, .tif, .jpg, .gif, .pdf, .wav, .mpg, etc.) A file format extension must always be present.
- Take into account the maximum number of items to be reformatted and reflect that in the number of digits used (if following a numerical scheme).
- Use leading 0's to facilitate sorting in numerical order (if following a numerical scheme).
- Do not use an overly complex or lengthy naming scheme that is susceptible to human error during manual input.
- Use lowercase characters and file extensions.
- Record metadata embedded in file names (such as scan date, page number, etc.) in another location in addition to the file name. This provides a safety net for moving files across systems in the future, in the event that they must be renamed.
- In particular, sequencing information and major structural divisions of multi-part objects should be explicitly recorded in the structural metadata and not only embedded in filenames.

NAMING DERIVATIVE FILES

The file naming system should also take into account the creation of derivative files made from the preservation files. In general, derivative file names are inherited from the preservation file, usually with a qualifier added on to distinguish the role of the derivative from other files (i.e., "pr" for printing version, "t" for thumbnail, etc.) Derived files usually imply a change in dimensions, for example, image resolution, and/or file format from the preservation copy. Derivative file names do not have to be descriptive, as long as they can be linked back to the preservation file. For derivative files intended primarily for Web display, one consideration for naming is that images may need to be cited by users in order to retrieve other higher-quality versions. If so, the derivative file name should contain enough descriptive or numerical meaning to allow for easy retrieval of the original or other digital versions.

FILE NAMING CONVENTIONS

Institutions may consider using the following conventions to assure consistency and ease of use. If followed, these file naming conventions can be utilized across various projects and by all staff.

- No spaces in filenames.
- Use underscores _ to separate discrete information.
 - For example, a historical society might use, repository HTUWIL_ vs collection ID MS0422_ vs container b0002_ vs folder f004_ vs sequential identifier 0001)
- Use hyphens - when the next numbers are part of the previous whole.
 - For example, HTUWIL_MS0422_b001_f9876_CHS16010-01
 - CHS16010-02.
 - Except in the case of a collection that was only cataloged to the box, and folder level, the Norton Ranch Collection, b001 f0001 indicates box, folder.
- After an image or scan number, A, B refers to multiple images that are exact duplicates or so similar that the cataloger lazily used one identifier, while a, b refers to the front and verso of an image or page.

- All scans of a single item, no matter how many scans it takes to capture, have the same file name, except for the a, b, c, added at the end.
 - For example, all pages from a single pamphlet will have the same file name, with a-z added at the end.
- Use three leading zeroes for box, folder, and sequence, so that each section is consistently 4 digits.
- For material whose digital file name references a location.

For example, b, f, the final sequence reflects the order in which material was digitized, not the order within the folder unless the entire folder was digitized in one fell swoop (the latter is ideal but unrealistic).

FILE NAMING PROTOCOLS AND PROJECT MANAGEMENT

As a long-time project manager, I can say with authority, having overseen projects with multiple institutions and many partners collaborating across organizations from different states and regions, one thing that is often overlooked during the project planning stage is the need for the development of consistent file naming protocols.

The need for such standard operating procedures becomes more and more critical as the number of collaborating partners expands. Over time, as digital files begin to be submitted, a project manager is tasked with sorting, organizing, and ultimately creating a preservation repository documenting the work of a partnership, collaborative or group. If a project has been grant-funded, those documents are often required to be made public and/or may need to be submitted to a federal or state agency as part of the interim or final deliverables. Therefore, establishing criteria up front is vital and the larger the project, the more important this can be.

Some examples of how to craft a specific strategy for project file naming that may be used across institutions and organizations for grant and collaborative projects include.

- Use of underscores _ to separate discrete information is very important to your documentation.
 - Let's say this was a grant project. One might have every submission start with the grant number.
 - PE-295821_
 - This is an example of the number series assigned to NEH Preservation and Access Education and Training grants.
 - 2023_16_c\$f_wmpatersonu_idea.pdf
 - This is an example of a Lyrasis Catalyst Fund Grant application.
 - If there was not a grant or grant number, the project might have a project name that is used as the primary identifier.
 - With large multi-party projects creating a list of three or four-digit alpha or numeric unique identifiers for each project partner can be useful.
 - The word 'report' could be changed and used in another form. In this case it is envisioned as a monthly, interim, or final report.
 - Soundmodel_report_4digitinstitutionID_YYYY_MMDD.docx
 - Soundmodel_report_4digitinstitutionID_YYYY_MMDD.pdf

For ways to plan for file naming, detailed questions to consider, and more explicit definitions, see chapter nine of [FADGI Technical Guidelines for Digitizing Cultural Heritage Materials Creation of Raster Image Files](#). This work is available for worldwide use and reuse under CC0 1.0 Universal.