

12-11-2018

The Digital Librarian: The Liaison between Digital Collections and Digital Preservation

Marina Georgieva

University of Nevada, Las Vegas, marina.georgieva@unlv.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/lib_articles



Part of the [Library and Information Science Commons](#)

Repository Citation

Georgieva, M. (2018). The Digital Librarian: The Liaison between Digital Collections and Digital Preservation. 1-5.

https://digitalscholarship.unlv.edu/lib_articles/589

This Article is protected by copyright and/or related rights. It has been brought to you by Digital Scholarship@UNLV with permission from the rights-holder(s). You are free to use this Article in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Article has been accepted for inclusion in Library Faculty Publications by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

Author: Marina Georgieva, Visiting Digital Collections Librarian, University of Nevada - Las Vegas

The digital librarian: the liaison between digital collections and digital preservation

Overview

At UNLV Libraries, the role of the Digital Collections Librarian goes beyond the traditional routine tasks of digitization, metadata management, project management, workflow development and team management. Digital Collections Librarians serve as links between digitization and digital preservation and do everything in between to draft sustainable digital preservation workflows alongside their colleagues in the Special Collections Technical Services Department. Technical Services Librarians are responsible for the preservation of born-digital archival materials, whereas the Digital Collections Librarians' roles entail being information architects directly engaged in the process of preparing master files of in-house and outsourced reformatted materials for digital preservation.

In recent years, the UNLV Libraries Digital Collections Department has completed numerous large-scale digitization projects that yielded hundreds of thousands new archival digital objects that require long-term preservation. Currently all these archival files are stored on a server, referred to as 'The Digital Vault'.

One of the invisible, often overlooked, yet very important roles of the Digital Librarian is to verify that all images from completed digitization projects are properly organized in meaningful easy-to-navigate directories and that all files are in the appropriate file format. It is common practice for folder directories (created and organized during the actual process of digitization) to remain intact and be moved to the Digital Vault for long-term storage in their original order. There they get merged in the collection-appropriate existing folders or, if necessary, a new folder is created.

Additionally, UNLV Digital Collections has thousands of images from legacy collections stored in the Digital Vault. All of these digital objects live on the Digital Collections website, but some of the archival master folders have redundant data; others are saved in inappropriate file formats, and still others have non-normalized file naming. In the recent years, there has been an effort to clean up and restructure these legacy folders in order to make the archival files easily discoverable and to optimize the storage space before the content of the Digital Vault gets

migrated to a new more robust system (UNLV Special Collections and Archives is currently building an instance of Islandora CLAW that will back up files in Amazon Glacier).

The role

The role of the UNLV Libraries Digital Librarian that relates directly to the digital preservation is outlined in the poster presented at 2018 NDSA DigiPres Forum ([click here for access](#)). Here we will just briefly touch upon few of the major responsibilities:

File naming conventions

For current digitization projects, file naming has been normalized and it happens in a structured and logical way depending on the type of collection being digitized. During the process of preparing collections for digitization, the librarian analyzes the content, makes decisions regarding the grouping of the digital objects and assigns collection-level and item-level digital identifiers. To achieve consistency and logical arrangement, the digital librarian maintains and updates spreadsheets with assigned and available digital identifiers.

For example, if the collection consists of archival photographic materials, the assigned digital collection alias will be 'PHO' with the sequential numeric identifiers. These identifiers will logically follow the structure and numbering of all other previously digitized photo collections.

As mentioned earlier, most of the newly digitized collections remain in the original directory structure that was developed during the scanning process. The digital librarian ensures that the file naming on directory level and on file level is accurate and the data set is ready to be moved to the Digital Vault.

It is important to mention that often digital librarians need to deal with and manage more identifiers beyond those that identify archival structure (collection, folder) and those that identify the intellectual unit (item) so that they can accurately reflect the structure of materials. So they also need to create a third type which may involve multiple image files that comprise a single digital object; for example, back and front of a printed item or multiple items on a page in a scrapbook.

Legacy collections bring more challenge and sometimes need some clean up as their file naming may be inconsistent. Depending on the project, the digital librarian may decide to keep the file structure intact or to rearrange the folders in more normalized way that follows the current preservation practices.

Decisions on archival file formats

UNLV Libraries Digital Collections have chosen TIFF file format for long-term preservation of archival master files. TIFF is the preferred format for in-house digitized reflective materials and transparencies.

The file format for digitized periodicals may vary depending on the project. In-house digitized periodicals and newspaper clippings are preserved in TIFF just as photographs and films, while periodicals digitized as part of the National Digital Newspaper Program are stored in the original Library of Congress approved data sets. These data sets include newspaper pages in JP2, PDF and TIFF formats along with the accompanying metadata encoded in XML METS/Alto schema.

Legacy collections may contain files in JPG format. This usually applies to collections accessioned as already digitized materials. The reason why they usually they remain in this format is that UNLV Libraries Special Collections do not have holdings of the original materials and therefore, it is impossible to re-digitize the items in the proper archival format.

Building directories in the Digital Vault

Current digitization and digital preservation efforts follow well-established practices regarding how files are nested in directories so that they have logical structure and are easy to navigate.

For example, the archival master files of a digitized photographic collection get migrated to the general folder that holds all archival files of all photo collections. This directory contains a blend of additional sub-folders that represent compound objects and files that represent the single objects. It is nested in a higher level Photo Collections folder.

To illustrate the scenario above, please examine the following example.

[...] Digital Vault\PHO Photo Collections

> PHO Archival Images

> pho003089

- pho003089_001.tif

- pho003089_002.tif

- pho003089_003.tif

- pho003089_004.tif

> pho003090

- pho003090_001.tif

- pho003090_002.tif

- pho003090_003.tif
- pho003090_004.tif
- pho003090_005.tif
- > pho003091
 - pho003091_001.tif
 - pho003091_002.tif
- [...]
- pho003092.tif
- pho003093.tif
- pho003094.tif
- [...]

Legacy collections usually get reorganized, especially if the folder structure is not logical or there are redundant files and folders.

Outsourced periodicals are kept in the original directory structure as created by the vendor. Data sets arrive separately and each of them is considered a batch and is stored separately in a parent-level folder hosting only outsourced periodicals.

[...] Digital Vault\NDNP Local Backup

- > Batch Aurora
 - *Original data set as received by vendor*
- > Batch Beatty
 - *Original data set as received by vendor*
- > Batch Caliente
 - *Original data set as received by vendor*
- [...]

Communication with the IT Department

The Digital Vault is a directory with limited access – librarians get “view only” mode and they need to communicate all needs for data migration, remediation requests and new decisions regarding folder structure to the IT department who maintains the Digital Vault.

The role of the digital librarian in this communication is critical, because of the limited access to the server. Usually this communication is informal writing – simply emailing the requests or updates along with instructions what needs to be accomplished. Recently, for some larger and more complicated clean-up projects, the UNLV digital librarians adopted Google Sheets. The advantages of Google Sheets is that more than one person can access and edit the document simultaneously to communicate changes and project updates.

The archival files prepared for migration are stored in a temporary location and once the move is complete, the digital librarian checks if all files and directories were moved successfully, and if there are any corrupted files or other discrepancies that need attention. Upon verification, the files in the temporary location are deleted permanently.

In remediation scenarios when legacy stuff needs to be cleaned up (deleted, moved elsewhere or restructured) usually the communication includes initial written instructions followed up by one-on-one meeting. The one-on-one meeting talks through all the necessary changes and serves as a final overview of the project as the nature of these actions is irreversible.

Conclusion

The day-to-day job of digital librarians has a slightly different focus than digital preservation, and yet digital librarians play a valuable role in building structure, organizing, and cleaning up data. Digital librarians are very user-focused - not just on internal end-users of the archival masters, but on library users who may need delivery of master images via a library system (online) or directly (via image reproduction). Their outstanding organizational skills and attention to detail not only make the data easily discoverable and ready for migration, but also optimize the storage space and lay the groundwork for a smooth migration to a new, more robust system. In institutions just starting to make baby steps in digital preservation, the digital librarian plays a key role of advancing a step closer to a robust and efficient long-term digital preservation strategy.

Link to poster presented at 2018 NDSA Digital Preservation Forum

https://drive.google.com/file/d/1fOHjb0g_2lawsymEihQJXa81TWS_MREY/view?usp=sharing