TRENDS IN
ARCHIVES
PRACTICE

MODULE 2

PROCESSING DIGITAL RECORDS AND MANUSCRIPTS

J. Gordon Daines III

SAA
SOCIETY OF
American
Archivists
CHICAGO

Appendix A: Case Studies

Duke University: David M. Rubenstein Rare Book & Manuscript Library
by Seth Shaw

The David M. Rubenstein Rare Book & Manuscript Library uses two methods to arrange and describe born-digital records. Using the first method, we note the records' logical arrangement, without changing the folder structure. Using the second method, we copy the files and folders into a new structure mirroring the collection's series structure, while retaining a copy of the files and folders as accessioned.

We employed the first method when a collector made a purely digital addition to the Common Sense Foundation records. This accrual represented a network file store snapshot, with known non-permanent records omitted. After securing the materials in our dedicated file store, we placed a working copy of the files on a computer that is dedicated to processing born-digital material to allow for arrangement and description. We used a variety of tools to review file contents. Quick View Plus (a commercial file viewer application supporting thousands of file formats) is a useful general-purpose tool. To examine specific formats more carefully, we made selective use of plain-text editors, the Microsoft Suite, Adobe Acrobat, Irfan Viewer (images), VLC media player (audio/video), and hex-viewers. Each top-level folder was quickly evaluated to identify whether it could be mapped to an existing series or subseries. If a match could not be identified, the second-level folders were analyzed in similar fashion. If second-level review failed, the processing archivist created new series and subseries as appropriate to fit the record content.

In practice, we have found that the processing archivist rarely needs to provide arrangement lower than the second level. Since logical arrangement is used, a folder or subfolder may be listed in multiple series as appropriate. For example, a new "Staff Files" series was created. Sub-folders within two "Staff Files" subseries merited inclusion with the "Board of Directors" series, and so were described there though no files were moved, allowing multiple access points from the finding aid.

When each folder's appropriate arrangement has been identified, it is added to the existing EAD series component element with its own component and container element. The container element's type attribute is "efolder" and label attribute is "Electronic Folder." The container element's text value is the path to the parent folder (if not the top level). The unit title is simply the folder's name. Extent lists folder count, file count, and megabytes (identified on Windows machines by right-clicking and selecting "Properties"). Earliest and latest date modified values are used for date ranges (identified using an in-house script). The scope and content note describe the record content in the same way that analog materials would be described, with additional notes describing common file formats as appropriate.

In some cases, we use an alternate method to arrange and describe the records. The Rubenstein Library's second method allows an archivist to create an electronic series structure (using filesystem directories). Once the series structure has been established, the archivist moves folders and files from their original locations into the appropriate series/directory.

This method was used for the Mark Bowden papers. Note that folders and files are moved only to a single location (and not copied), but may then be cross-referenced in multiple series as described in the previous method. This option is more appropriate when there are several small additions or content spread across multiple media with no artifactual value and is comparable to consolidating papers from several small boxes into a larger one. Associations within the media are not irrevocably lost in this method, because they are preserved in the accessioned structure, though not immediately available to the researcher via the "processed" arrangement.