Key issue

Traditional and digital preservation: appraisal, accessioning and arrangement

AMY KIRCHHOFF
Archive Services Product
Manager
Portico

Digital preservation is the act of ensuring that electronic content of enduring value remains usable and authentic over the long term. While there are unique aspects to digital preservation, it is built on long-standing practices developed by traditional archivists of physical collections. The appraisal, accession and arrangement processes that are an early part of traditional archival procedures are also critical early steps in digital preservation.

Key terms

- *Appraisal:* 'the process of determining whether records and other materials have permanent (archival) value.'[1] The appraisal process determines what content will be preserved and for how long and what content may be otherwise disposed.
- *Accessioning*: the process of establishing 'legal, physical, and intellectual control over the collection'.[2]
- *Arrangement*: 'the process of organizing records and papers to reveal their contents and significance'[3]

In digital preservation, each of these traditional activities is critical and is undertaken with a new perspective.

A story

A botanist requires preservation services and ships her computer's hard drive to a digital preservation archive with a brief note describing the drive's directory structure. Most files would probably be organized according to the provided documentation, but there would surely be anomalies. Perhaps images would appear in 'My Documents', rather than in 'My Pictures'. Perhaps most of those images in 'My Documents' are commercial clip art and thus have a different archival value than images of plant specimens in her 'My Pictures' folder. Perhaps among the clip art images in 'My Documents' are misfiled but valuable plant specimen images. When considering only a single hard drive, one could imagine reviewing each file manually to appraise its archival value, but if we multiply this single hard drive across the careers of many faculty, it is clear that an automated approach is required.

The story above highlights one of the challenges raised by the large-scale preservation of electronic scholarly publications. For each supplied file, the preservation archive must understand the file's purpose and authenticity and properly organize it into a unit of intellectual content – these units of intellectual content can then be preserved for discovery and use by future scholars. Portico is a digital preservation archive supporting the scholarly community, with a capacity to ingest approximately 1 million e-journal articles or 10 million individual files a month. Portico takes a highly automated approach to traditional appraisal, accessioning and arrangement activities and offers a case-study of how traditional archival activities inform large-scale digital preservation practices. (For further information on Portico's preservation approach, see 'Archiving Electronic Journals: An Overview of Portico's Approach'.[4])

Appraisal, accessioning and arrangement in action: the Portico case

The appraisal guidelines used by traditional archives in assessing the archival value of content take into consideration many factors, such as the rarity of material, circumstances of creation, archivist's assessment of future research value, and effort required to organize. In a similar manner, digital preservation services have archival policies guiding appraisal. Large-scale digital preservation services typically appraise content at both the collection level and the file level. For example, based upon the needs expressed by the scholarly community, Portico selected electronic journals as the first collection of content to be preserved. Within that 'collection', Portico does not make appraisal decisions on a title-by-title basis, although Portico does make file-by-file appraisal decisions.

With digital preservation, accessioning occurs very early in the archival process. Much as traditional archivists secure legal agreements with depositors, Portico secures perpetual, archival rights to preserve identified scholarly content. Once a clear legal framework for Portico's preservation activities is established, Portico manages content from receipt through ingest into the archive. In traditional archives, this same inventory process often involves assigning accession numbers and organizing boxes, whereas Portico has an automated work flow to manage the accession and arrangement processes.

With digital preservation the appraisal, accession and arrangement of content at the individual file level are highly intertwined. In general, files that can be arranged into intellectual units of content are appraised as having archival value. Files which cannot be so arranged have more dubious value as they cannot be made discoverable to scholars of the future. For example, Portico must be able to identify whether any given TIFF image is an equation, a figure, or a page image and to which article and title it belongs – without that context, the TIFF image is of very little enduring value.

To arrange publisher files into reasonable units of intellectual content, Portico defines rules that describe the individual publisher's content package and maps this content packaging to an archival content model. (The Portico model is based on the Digital Item Declaration Language (DIDL) of MPEG-21[5] and implemented in a Portico version of METS[6]). Any file that follows the specified rules can be properly arranged and is appraised as having archival value. As with personal hard drives, however, there are files in the content supplied to Portico that do not follow the rules. In these cases, following Portico's preservation methodology, normal automatic processing stops and the problematic files are addressed. If the quantity of problematic content is small, staff manually analyze the files and, through visual inspection of the files and the publisher's website, identify the function of the files and the articles to which they belong so that normal arrangement and appraisal can continue. When there is a significant amount of problematic content, special rules and tools are developed to automatically associate files with an article or issue, even when we cannot positively identify the function of the files. Portico archival procedures guide us to be cautious in disposing of content even though this can result in the preservation of files of lesser archival value (e.g., US postal service forms used in the mailing of particular issues).

So although digital archives such as Portico rely upon different infrastructure and tools from traditional archives, many digital preservation practices – including appraisal, accessioning and arrangement – are informed by traditional archival practice. In both traditional and digital archives, clear policies for exceptional cases are needed.

References

1. Pearce-Moses, R, A Glossary of Archival and Records Terminology. (2005). http://www.archivists.org/glossary/term_details.asp?DefinitionKey=3 (Accessed 22 Apr 2008).

2. Hunter, G S, *Developing and Maintaining Practical Archives*, 2nd ed. Vol. 122, *How-To-Do-It Manuals for Librarians*. New York, NY: Neal-Schuman Publishers, Inc., 2003., p. 87.

3. Ibid. p. 113.

4. Kirchhoff, A, and Fenton, E, *Archiving Electronic Journals: An Overview of Portico's Approach. Papers from Portico*. 2006. http://www.portico.org/news/PapersFromPortico.1.Overview.pdf (Accessed 12 May 2008).

5. MPEG-21 Part 2: Digital Item Declaration Language (DIDL). In *Technology Reports*, [Website] Cover Pages (2004).

http://xml.coverpages.org/mpeg21-didl.html
(Accessed 12 May 2008).

6. METS: Metadata Encoding & Transmission
    Standard. [Website] Library of Congress.
    http://www.loc.gov/standards/mets/ (Accessed
    12 May 2008).

Article © Amy Kirchhoff

E-mail: amy.kirchhoff@portico.org

To view the original copy of this key issue, published in *Serials*, the journal of the UKSG, click here:

http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=21&Issue=2&spage=146

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=21&issue=2