MODULE 16

ACCESSIONING DIGITAL ARCHIVES

Erin Faulder

SOCIETY OF
American
Archivists

Case Study 3: Real-World Accessioning Is Messy

By Max Eckard, Assistant Archivist for Digital Curation, University of Michigan, Bentley Historical Library

While functional models such as the Open Archival Information System (OAIS) Reference Model and the Producer-Archive Interface Methodology Abstract Standard (PAIMAS) are useful for identifying, defining, and providing structure to the relationships and interactions between information producers and archives, their unblemished facades mask an ugly truth: real-world accessioning of digital archives and manuscripts is often messy. Despite the best efforts of archivists, the information objects to be archived are not always well defined, a formal agreement as to what will be delivered and when is not always made (and, even when it is, is not always followed), and the complementary processes of transferring objects and validating transferred objects are not always straightforward. Rather than detail a particular accession, this case study provides examples of the many ways in which the successful accessioning of digital archives is not as structured or formal as professional standards would lead one to believe.

Defining the Object of Preservation

While the producer and the archives may have a good idea at the onset of a project as to what records should be transferred, it can still be difficult to define the primary object of preservation. When a university photographer, for example, offered her current and future portfolio of materials to the archives at Grand Valley State University, a public liberal arts university in Allendale, Michigan, the archivists knew they had an exciting opportunity to preserve the history and memory of student life and social customs at their institution. However, deciding which digital objects would be preserved was more difficult. Archivists had to determine if original, lossless RAW-formatted images needed to be accessioned or if derivative JPEGs (with associated metadata) would suffice. Since the producer was willing to transfer all of her existing photos (around 24,000 images captured over approximately eight years), the answer to this question had a significant impact on the volume of data to be initially transmitted and stored: 396 GB versus 60 GB, respectively. Likewise, the answer would affect the level of preservation service that the archives could provide. Since the archives had

not established normalization pathways for RAW images, the archives could only provide bit-level preservation for them. Determining the precise version of content to be transferred therefore had major implications for the overall cost and feasibility of the project, especially since future accruals were expected.

After considering the perspective of the data's primary designated community—students, faculty, and staff of the university as well as outside researchers and the general public—the producer and archives decided to set up a workflow to transfer a representative sample of the derivative JPEGs from the "top ten" university events each year, with a spreadsheet of associated metadata. This brought the overall volume of the initial transfer to 1,600 images and 4 GB, and the estimated volume of annual accruals down to a manageable 200 images and 500 MB. Both parties acknowledged that because this relationship was ongoing and the volume and type of images as well as the ability of the archives to handle them would likely change over time, details of the transfer would need to be revisited periodically to ensure that the premises of the initial agreement still held true. Both parties were satisfied, and, in due course, accessions of these important materials began.

Protecting Confidential Information

While archivists actively promote open access to the records in their care, some accessions contain confidential or personally identifiable information that requires restrictions on access. Archivists at the Bentley Historical Library thus needed to proceed carefully to meet strict requirements for data security when they accessioned the email of a university administrator and realized, after conducting a review and scanning for personally identifiable information (PII) with a tool called Identity Finder, that the records contained sensitive information related to the Family Educational Rights and Privacy Act (FERPA) and the Health Insurance Portability and Accountability Act (HIPAA) as well as Social Security numbers. Archivists had originally intended to apply a standard twenty-year executive records restriction to the content since the producer had not mentioned this type of data during previous negotiations. The discovery of personally identifiable information complicated this initial decision. Furthermore, the presence of FERPA- and HIPAA-protected data meant that the accession was not eligible for deposit into the Bentley's preservation repository because

the underlying storage was not rated as secure enough for data classified as sensitive or critical to the operation of the university.

In seeking to be responsible custodians of this information, archivists needed to determine an alternate approach to ensure it could be stored in a manner that would protect its authenticity, integrity, and security. The Bentley already maintained a secure Networked Attached Storage (NAS) device that was used as a local storage for reference scans and internal projects. Archivists thus decided to create a partition on the NAS and provide access to a limited number of senior staff, thereby establishing a dark archive, with backups replicated to University of Michigan storage that was approved for sensitive data. Since then, archivists have worked with Library Information Technology to explore options for institutional repository storage and will soon move their repository to a storage protocol that provides a secure environment to store sensitive university data.

Transfer and Validation of Digital Objects

While the transfer of digital objects from the donor to the archives may seem straightforward, anomalies abound, as archivists at the Bentley recently discovered when accessioning a number of videos documenting campus events. A hard drive with these materials showed up relatively unannounced on the archivist's desk one day: the donor happened to be in the area and felt it was most convenient to drop off the drive, with an expectation of retrieving it at the end of the week.

When archivists attached the hard drive to the Bentley's removable media station as part of the transfer process, nothing showed up. Given the prevalence of the Macintosh operating system on campus, particularly for those who work with video editing software, archivists suspected that the drive was formatted using Hierarchical File System Plus (HFS+), a proprietary file system developed by Apple that cannot be read natively by the Microsoft Windows operating system. After consulting professional listservs, archivists downloaded HFSExplorer, a free and open source tool for reading HFS file systems on Windows. The application allowed the archivist to mount the hard drive, view its contents, and transfer authentic copies of the files to local quarantine storage.

Closing Thoughts

Accessioning digital materials will look different at every archives, and at any given institution it may take a long time and many small, incremental improvements to approximate PAIMAS or thoroughly address the functions of the OAIS ingest process. At some point, an archives will almost certainly be asked—or obliged—to accession important, mission-critical digital archives before it has established processes or even the basic infrastructure needed to handle them. Even for those archives that do have procedures in place, the digital world is complex and dynamic—it won't be long before some accession comes along that isn't covered. The good news is that by embracing these challenging situations, archives and archivists alike can transform them into opportunities for growth.