

MODULE 16

ACCESSIONING DIGITAL ARCHIVES

ERIN FAULDER



SOCIETY OF
**American
Archivists**

Appendix C: Case Studies

Case Study 1: Leveling Up Procedures: Results of Accessioning Records from Sudan Mediations

In the summer of 2013, Digital Collections and Archives at Tufts University (DCA) accessioned four born-digital collections totaling approximately 6,000 unique files. The DCA acquired the collections as part of the World Peace Foundation (WPF) grant titled “Documentation, Research, and Writing on the African Union High-Level Implementation Panel for Sudan.” The WPF requested the DCA’s assistance to preserve personal records that documented the mediation efforts in Sudan, South Sudan, and Darfur. This case study reviews some of the specific challenges of accessioning these collections and how the process increased the capacity of the archives to accession digital records.

The archives first wrote its accessioning procedures for digital records in 2007. The procedures met local accessioning and stabilizing needs for recovering files from obsolete media uncovered in primarily analog collections. The processing archivist identified the media (often a 3.5-inch floppy disk), write-protected the media when possible, and ran a virus scan over the files on disk after inserting the media into the appropriate drive. When the virus scan software confirmed that there were no threats, she copied the files from the disk to the medium-term processing storage area with a read-me text file in a file directory corresponding to the collection. The read-me file identified the following key metadata elements: who conducted the stabilization and when and how the stabilization off media happened, the accession number, a list of files and their checksum values created by Advanced Checksum Verifier, and any notes that identified anomalies and what the archivist had done to address them. The archives already had experience transferring digital records from university offices, often following an appraisal conversation that occurred as part of the records survey. However, during the six years with these procedures, the archives had never taken a large quantity of digital records from nonuniversity creators from actively used systems. There were no documented procedures in place to work with donors to support an accession of this scale.

When the WPF approached the archives as a partner in the preservation of the records, the archives agreed to provide technical expertise and long-term preservation services for the collections. The WPF hired and managed graduate students from the Fletcher School of Law and Diplomacy to catalog the records at the item level with guidance from an archivist. The WPF staff also coordinated all of the negotiations with the donors, who were also the creators. The added layer of communication between the archives and the creators made asking and answering questions about the transfer and accessioning process difficult. The creators were all involved in the Sudan, South Sudan, and Darfur mediation efforts. Their records presented legal, ethical, and cultural sensitivity concerns that demanded a high level of intellectual and physical control to manage them properly. The archives had to develop strategies to address the complexities of the collections, the time sensitivity imposed by the grant schedule, and the large number of parties involved in the negotiations.

The archivists had a number of meetings with the WPF staff to gain an understanding of the grant deliverables, the purpose driving the grant project itself, and the complex political situation of the mediation efforts that led to the records' creation. Given this background, the archives recommended that the archivist assigned to the project give the WPF graduate students, who would be cataloging the materials and liaising with the donors, lessons in archival principles. The archivist's coaching would help the WPF staff and graduate students guide conversations with donors. The archivist would also provide strict cataloging guidelines so the metadata created would align with regular archival standards and systems.

For the first month, the archivist helped the WPF staff revise cataloging they had already completed on the collection that the first donor had transferred to the WPF's control prior to the archives joining the project. The opportunity to simultaneously review and correct the already-completed cataloging, while closely appraising the records, allowed both archivists and graduate students to understand the complexity of the collection and the potential legal, ethical, and security concerns that the next three collections would pose.

In the second month, the DCA worked with the WPF staff to coordinate the transfer of the three other collections. Because many

of the creators lived and worked outside the United States, there was only one opportunity to simultaneously transfer records and gather as much contextual information about the records as possible in a single face-to-face conversation. Given the context gleaned from the deep appraisal and processing of the first collection, the DCA developed a one-page, ten-question donor survey that asked specific questions about known issues with these records. The archivists carefully reviewed the form with WPF staff to ensure they understood why the questions were being asked, what kind of information would be useful to get from the creators, and what kinds of answers would provide red flags that required additional probing. These red-flag answers focused specifically on intellectual property and restriction concerns. Answers to the questions would help document the creators' record-creating practices. These answers would then help identify gaps in the records transferred or questions about the records' authenticity. While guiding the conversation through a third party was not the most efficient means of gathering information from the creators, the structured questions and the month of intensive training about archival principles allowed the WPF staff to have productive conversations with the creators. The donor survey, and conversations with some of the donors during transfer, revealed that key records were not part of the transfer. These records were primarily emails, produced as a result of the way this team worked, and photographs, which documented mediation meetings. This gap came about because the donors self-selected the records they thought were pertinent without prior archival appraisal. The donor surveys and all information from these conversations became part of the collections' accessioning documentation.

Given the location of many donors, and the information's sensitivity, the archives asked the WPF to request that the donors transfer records they considered part of the collection to a flash drive and hand deliver the flash drive to the WPF staff. The WPF staff copied the files from the flash drive to the WPF's secure network drive. The archives staff then copied the files from the WPF network drive to a medium-term storage server. The WPF copy of the records became the working set, used for appraisal, arrangement, and description. The copy in the archives was the stable version of the SIP that sat on the secure server until the AIPs were ingested into the institutional repository. Ideally,

given the nature of the records, the archives would have preferred the creators to run checksums and/or use a tool to capture file system metadata before copying the files to the flash drive. However, given the nature of the situation, this was impossible. Instead, the gap in control became part of the accessioning documentation.

In the third month, the WPF graduate students spent their time cataloging the new collections with the assistance of the archives. While much of this work could be considered processing, and outside the accessioning efforts, the line between the accessioning and processing was blurry. Through visual inspection (opening every file of the working copy of the SIP), they identified and recorded the title, creator, and date as well as noting any records that had confidential or sensitive information in a standard data entry spreadsheet that the archives uses. When there were specific questions about content or context, the WPF staff occasionally reached out to the creators with specific clarification questions. This was moderately successful soon after the transfer, but, as time passed, the donors were less responsive. Often the date used to identify a record came from the creation date of the digital file. In one collection, due to the nature of the transfer process, all files had the same file creation date. In this case, the last modified date became the item's date. This spreadsheet was imported into the collection management system used by the archives and used to generate the EAD finding aids and the metadata about publicly available objects.

The archivists, meanwhile, filled out the submission agreement form on behalf of the producers (the WPF) using signed donor agreements, the donor surveys, and other information provided by the WPF. The archivists also followed other standard accessioning and processing procedures. As preservation and technical experts, the archivists ran checksums for all objects using Advanced Checksum Verifier, created the intellectual arrangement scheme, and normalized the file formats using a third set of copies of the SIP in the processing storage space. In doing so, the processes that were designed primarily for transferring and stabilizing small quantities of records off obsolete media revealed their inability to scale reliably. The standard read-me text files were not structured data, and thus the values could not be programmatically extracted for import into the collection management system. The

data received from the checksum tool had the same problem. The standard file normalization procedures were labor-intensive and difficult to scale beyond about 150 objects. Not requiring the donors to transfer their records with checksum values meant a lack of information about objects prior to their arrival at the archives; this meant transfer validation activities beyond a virus scan were meaningless. The multitude of tools used for discrete tasks was acceptable at a small scale, but, at the larger scale, they hindered the archivist's ability to efficiently manage data about the archives' electronic records.

Ultimately, accessioning these collections was successful. But, as a result of these experiences, the archives made several changes to its regular procedures to ensure that the next time a similarly complex set of collections arrives a better infrastructure would be in place to manage it with greater control and efficiency. First, the archives instituted Bagger as a packaging tool for all SIPs. All records that are transferred to the archives are bagged, whether the transfer comes from obsolete media or over a network. Bagger runs the checksums and allows archivists to capture most of the same data previously included in the read-me text file. One reason Bagger became the SIP packaging tool is because it uses a customizable GUI form for filling out local accessioning data (i.e., accession procedure, accession number, name of person accessioning, collection number, media type, original media annotations, virus scan software tool, other tools that may have been used, and transfer notes). Moving to this tool combined two steps (creating checksums and a read-me text file) into one while also using a standard packaging format. Bagger also produces a file directory printout. In an institution that uses both Macs and PCs, Bagger can be installed on any of the computers. With training, the simple GUI form can be filled out by any of the archives staff or student workers, which increases accessioning capacity and puts the data in a regularly structured form for easier reuse. Second, the archives has spent time diagramming the potential and likely ways transfers can come in to map out workflows and create stable procedures for those transfers. Third, the lessons of working with an intermediary in negotiating with creators has highlighted the need for a standard set of questions to ask producers—who may in turn ask them of creators—to get the kind of information that helps archivists gain intellectual control over a

collection. Finally, the archives began looking at ways to scale up file normalization processes so that they are less labor-intensive. While the old procedures allowed the archives to successfully accession content transferred through a particular set of procedures for many years, this experience demonstrated the value of regular evaluation and iteration of procedures. Assuming an iterative approach to developing workflows has increased skills and capacity in the archives to handle more complex transfer and accessioning situations.