Building Capacity for Born-Digital Archives Through Accessioning Workflows at the Friends Historical Library

Emily Higgs Kopin, Jessica Hutchison, and James Truitt

ABSTRACT

Large archival institutions have contended with the challenges presented by born-digital materials for much longer than smaller institutions, with the latter only recently beginning their own attempts to manage these items. This case study details the process of starting a born-digital archival accessioning program and iteratively designing workflows scalable for a small institution. We review our approach to acquiring storage space and assessing our existing born-digital holdings and needs, lay out our initial accessioning workflow, discuss our processes of documentation and iteration, and reflect on tensions with archival education and best practices. By sharing this case study, we aim to outline a practical approach to initiating a born-digital archiving program using robust and extensible accessioning practices. Further, we emphasize that a focus on iteration and a willingness to weigh best practices against contextual limitations can provide a crucial path forward for under-resourced institutions struggling to wrangle born-digital records.

KEY WORDS

Accessioning, Best practices, Born digital, Iteration, Liberal arts college

While large organizations have been wrestling with the difficulties of born-digital archival material for decades, many smaller institutions are only now beginning to engage with these materials. In this case study, three coauthors share the process of starting a born-digital archival accessioning program and iteratively designing workflows scalable for a small institution. Although our accessioning workflow has not fully resolved questions about arrangement, description, and access, stabilizing our born-digital holdings has given us more breathing room to grapple with them. In addition, clearly documenting the workflow has allowed other staff to share the work of accessioning and processing born-digital materials. By sharing this case study, we hope to outline a practical approach to initiating a born-digital archiving program using robust and extensible accessioning practices. Further, we hope to show that a focus on iteration and a willingness to weigh best practices against contextual limitations can provide a crucial path forward for under-resourced institutions struggling to wrangle born-digital records. The framework presented in the recently released *Archival Accessioning Best Practices,* which has both validated our workflows and shown paths for improvement, offers to make this process more achievable for other small institutions.[1]

Background: The Friends Historical Library of Swarthmore College

Located within a small liberal arts college in Swarthmore, Pennsylvania, the Friends Historical Library (FHL) of Swarthmore College collects, preserves, and makes available archival, manuscript, printed, and visual records concerning the Religious Society of Friends (Quakers) from their origins in the mid-seventeenth century to the present.[2] Its archival holdings include the records of Quaker congregations as well as the records and papers of organizations, families, and individuals associated with the Society of Friends. At Swarthmore College, FHL is one of four collections under the umbrella of the Special Collections and Archives department, which also includes the College Archives, Rare Book Room Collection, and Peace Collection. At the time of writing, there are eleven full-time professional staff in Special Collections, four of whom are specific to FHL.

Historically, FHL has accepted regular deposits[3] of paper records from Quaker meetings at many organizational levels[4] for preservation, access, and (at times) digitization. Like other living religious communities, the Religious Society of Friends has recently adopted computing technologies that have transformed their centuries-old traditions of recordkeeping and publishing. To begin to address this shift, archivists at FHL began to accept digital records from Quaker meetings in 2011 only in the form of PDF/A files, which staff understood to be preferable as a preservation format to proprietary word processing files. These files were ingested by the archivist into DSpace and given minimal descriptive metadata; on-campus users

and members of depositing meetings could access the records digitally. No other preservation actions were taken in addition to this process. Beyond these digital records, digital storage media (floppy disks, flash drives, CDs, DVDs, etc.) were accepted as part of multimedia accruals and deposits. No processing was undertaken on these media.

To continue building capacity for born-digital archival records, Emily Higgs Kopin was hired in 2019 as the first digital archivist for Swarthmore's Special Collections department. Shortly after Higgs Kopin started in this position, the role was reorganized specifically to support FHL collections.

Establishing Born-Digital Archival Accessioning

In her new role, Higgs Kopin chose to build out accessioning workflows as the first step toward establishing a more comprehensive and functional born-digital archives, particularly in order to avoid increasing the backlog of totally undescribed digital material. Prioritizing a healthy and sustainable accessioning workflow provides a minimum viable process to alleviate backlog and provides paths to researcher access for incoming records. By mitigating immediate preservation concerns, accessioning also buys staff time to devise workflows for description and access. Further, a clear, well-documented accessioning workflow for born-digital records can be performed by archival workers who are *not* digital records specialists, allowing the digital archivist to focus their limited work hours on more complex digital collections that require specialist intervention. However, in order to realize this vision, Higgs Kopin had to address the issue of where this content could be stored, accessioned, processed, and preserved in collaboration with other FHL staff.

Preparation: Acquiring Storage

Prior to hiring a digital archivist, FHL was following some baseline practices for digital storage with room for improvement. The repository had relied partially on external storage media (external hard drives and CDs) for storing digital archival content. Staff were following guidelines like the "3-2-1" rule[5] but did not have plans or procedures to monitor the health of external storage devices. Swarthmore Information Technology Services (ITS) staff had provided a CIFS share on Amazon EC2 in 2017 for Special Collections use, and that storage was nearly full in 2019 when a digital archivist was hired. At the time, this storage was a great start to provide a digital "workspace" for staff, but the share did not adhere to best practices for long-term digital preservation; for example, all staff members had read and write permissions, and no procedures were in place for checking the health of storage or the fixity of files.

As Swarthmore College already had the infrastructure and staff knowledge to manage storage with Amazon Web Services (AWS), the digital archivist worked with ITS to set up improved long-term storage for digital preservation, while existing networked storage was reallocated specifically for digital collections undergoing processing. ITS staff were able to set up a new configuration with AWS for long-term digital preservation storage: a dedicated S3 bucket duplicated in another geographically distinct Glacier instance for long-term storage of digital assets.

Because this type of storage is relatively costly compared to external hard drives and local networked storage, the digital archivist drafted a tiered approach to the use of dedicated digital preservation storage based on the National Digital Stewardship Alliance's Levels of Preservation.[6] The new S3 bucket was reserved for digital assets that fit the following conditions:

1. Swarthmore is the only custodian of the content.
2. The content is part of an accessioned collection with a deed of gift.
3. The digital files are the only copy of the content, or the analog content is fragile or inaccessible (especially in the case of at-risk audiovisual materials).

Digital surrogates of analog materials not considered "at-risk" (i.e., archival records that could be redigitized in the case of disaster) are stored in Islandora, FHL's digital asset management system, which is backed up both locally in a consortium-managed data center and in vendor-managed cloud storage.

Preparation: Inventory and Environmental Scan

With improved digital storage infrastructure, the first task for developing new accessioning workflows was a survey and inventory of existing born-digital materials: both the PDF/A Quaker meeting records in DSpace and the many unprocessed storage media in the backlog. The results of this process indicated that the majority of unprocessed born-digital materials were in common and identifiable word processing file formats from the 1990s to the present; many storage media were readable on contemporary computing equipment, with obsolete storage media mostly confined to 3.5" floppy disks (with some 5.25" floppy disks); and that materials were often duplicative of published materials owned in analog format. These findings were crucial in starting to think about standardized workflows that could handle the *majority* of born-digital materials in the collection. While every collection has some difficult and complex digital objects, knowing that the majority of holdings were text-based and in readable formats kept staff from focusing too many resources on planning for hypothetical cases of multimedia artwork and software that other repositories may encounter more routinely.

In addition to surveying holdings, the digital archivist also spoke with depositors and donors in informal meetings to ascertain what unknown materials *could* be

given to FHL with assistance and what roadblocks existed for donors who may wish to do so. These conversations overwhelmingly indicated that although the limitation of acceptable formats to PDF/A could be a useful boundary when no digital records specialists were on staff, this was ultimately preventing Quaker recordkeepers from depositing their materials at all. Many of the donors and depositors to FHL are not fluent in digital technologies that lie outside of very specific use cases. Although the digital archivist had been trained in previous professional archival contexts to request precustodial interventions from the donor such as deduplication, file renaming, and file format migration/normalization, it was not practical to request this from donors struggling with simply transferring files from one computer to another. Similarly, transfer methods developed to ensure fixity (a best practice for digital preservation) that required donor-side use of specialized software were not feasible.[7]

Although the digital archivist may have envisioned implementing a well-formed accessioning workflow based on established practices at other institutions, these conversations made clear that planning extensible accessioning workflows would be a balancing act between aspiring to best practices and providing actionable pathways to facilitating digital acquisitions without overloading donor or staff capacity in the particular context of FHL. To address this balance, FHL staff thought deeply about the "why" behind aspirational practices at other repositories, considering the possibility that our own institutional goals and needs may not be completely compatible with popular practices in other contexts. Based on this process of examining goals and justifications behind best practices through the lens of FHL-specific priorities, Higgs Kopin completed an initial accessioning workflow draft in early 2022. This workflow is described in the following section, with some comparisons to the National Best Practices for Archival Accessioning Working Group's *Archival Accessioning Best Practices*; while this resource was released several years after FHL's 2022 workflow draft, the *Best Practices* have since proven to be useful benchmarks for assessing and improving the workflow.[8]

Accessioning Workflow (2022 Version)

FHL defined the goals of its born-digital accessioning workflow as (1) stabilizing digital records upon receipt, (2) establishing baseline intellectual control through documentation, and (3) establishing a path to access. Based on previous experience using the BitCurator suite of open tools for accessioning and processing functions,[9] FHL defined five core steps for born-digital accessioning:

1. Receipt of electronic materials
2. Creation of accession records in ArchivesSpace
3. Virus scan
4. Bagging and creation of checksum
5. Transfer to appropriate storage location

Contrary to some recommendations and based on the aforementioned discussions,[10] FHL has not asked donors to run checksums before transferring material; while ensuring fixity is a worthy goal, this practice severely impacts FHL's ability to receive any digital materials from our donor community. Instead, to smooth the process with donors and repeat donations, FHL receives most digital material through institutional Gmail and Google Drive accounts, a method that presents a much lower barrier to donors.

Because these services run their own virus scans, materials received this way are downloaded directly to processing storage. Files sent via cloud providers that do not automatically scan for viruses (e.g., Dropbox) are downloaded instead to a laptop running BitCurator. Then, to quarantine this device, its ethernet cable is disconnected until virus scanning is complete. Likewise, physical carriers are mounted via write-blocker to the BitCurator laptop with the ethernet cable unplugged.[11] In keeping with recent professional guidelines that rethink disk imaging as a catch-all best practice,[12] staff only create disk images of carriers that are older, contain unusual formats, or appear blank to the BitCurator laptop's filesystem. While this practice is aligned with recently developed best practices, it also works particularly well as an approach for FHL due to limited staff time for born-digital accessioning and processing.

Once materials have been received, the first priority is to establish intellectual control by creating an accession record in ArchivesSpace. In addition to information usually found in an accession record (provenance, dates of material, etc.), documentation for born-digital objects includes the size of the digital content in bytes and is linked to a digital object record that documents the location of the files.[13] The accession record is also used to document virus scans, bag files, and transfer files to preservation storage.

After an accession record has been created, a virus check is performed on material that did not receive one during the transfer process using ClamAV software packaged with BitCurator, either through the command line or the ClamTK GUI. The results of the scan are recorded and set aside to be added to the bag as a tag file and to the accession record in ArchivesSpace through a virus-check event. At this point, the BitCurator laptop can be reconnected to the network via ethernet cable.

Bagging completes the tasks of stabilization and documentation begun in previous steps. FHL staff use the Bagger program[14] to create checksums for the accessioned files and compile metadata that will be stored with the files themselves. The checksum manifest (a list of the bagged files and their checksums) means that the materials' fixity can be confirmed in the future.[15] The bag-info.txt file provides basic metadata that can be used to identify and describe the files independently of the ArchivesSpace record. In addition, the standardized directory structure of the bag will facilitate performing bulk operations, if necessary, and aligns with Best

Practice 10.4 in *Archival Accessioning Best Practices*: "Use consistent and standardized packaging practices and/or directory structures for digital files."

The last step of accessioning is to transfer the bagged files to an appropriate storage location.[16] Transfers are made using a tool with robust checksum verification[17] to ensure that fixity is maintained. At the moment, most born-digital archival records receive only accessioning as baseline processing. Files meeting these criteria are wrapped in a TAR file and transferred to the preservation storage bucket. Additional processing with BitCurator's suite of tools is reserved for accessions that are particularly old, likely to be sensitive, or include uncommon or heterogeneous file formats. Finally, these materials are moved to the processing storage space if they were not initially downloaded there.

Initial Testing, Feedback, and Documentation

After initial development of this workflow in 2022, FHL began work with Jessica Hutchison, a graduate student at Drexel University, on a sponsored capstone project to test, improve, and document the draft workflows presented in the previous section. The capstone project "provides a closer connection to information work to ensure that library and information science students have practical work experience by the time they graduate and are prepared to enter the workforce at a professional level" and is a requirement for Drexel's MSLIS program that grants students course credit toward their degrees.[18] Hutchison's primary ten-week capstone objective was to test and improve FHL's born-digital accessioning workflows and to write documentation to train future undergraduate student workers.

The engagement of an unpaid capstone student to design professional accessioning workflows presents several tensions with the *Best Practices for Archival Accessioning*, which were in development at the time of Hutchison's capstone. Namely, the "Successful Accessioning Labor Practices" section of the *Best Practices* recommends that archivists "treat accessioning positions as skilled, professional labor" (Best Practice 13.1) and "create permanent positions" (Best Practice 14.1).[19] Higgs Kopin and Hutchison designed the capstone to function as an educational training opportunity in born-digital accessioning as well as an opportunity for Hutchison to further develop her expertise in translating professional archival processes into legible materials for a nonspecialist audience. In this way, Hutchison operated as a consultant in a scoped project and was not solely responsible for the development of a new program in a new workplace, which is often an untenable task for a student project and is unrealistic for any professional on a ten-week timeline. As is illustrated by the rest of the case study, the launch of FHL's born-digital accessioning program relied on multiyear work by multiple permanent archival professionals, largely due to the iterative nature of developing workflows and policies.

The authors strongly encourage the adoption of Best Practices 13.1 and 14.1 across archival institutions.

Throughout the project, Hutchison learned workflows for born-digital accessioning and processing. Higgs Kopin recommended literature to supplement Hutchison's learning and to further develop workflows, such as Rachel Searcy's article "Beyond Control," which outlines archivists' efforts to create accessioning workflows at New York University. Searcy's case study was particularly helpful in considering ways of approaching an accessioning workflow that is aimed toward an end goal (in Searcy's case, an access-driven workflow; in FHL's case, a student-archivist-driven workflow). Other documentation consulted during this phase is included in the Appendix. In expanding on drafts created by Higgs Kopin, the goals of this project were to ensure that the final documentation was approachable by a nonspecialist audience, usable by undergraduate students lacking prior experience working with born-digital materials in an archival setting, and empowering for student workers to understand not only how to approach this work, but also *why* they are doing each step. Aspirationally, students using this documentation would be able to approach born-digital materials with confidence and autonomy.

This documentation has created greater capacity for digital accessioning at FHL, though not exactly in the way expected. Higgs Kopin and Hutchison imagined student workers using the documentation to handle routine accessions, freeing the digital archivist to devote time to more complicated cases. However, FHL has not yet assigned students to born-digital accessioning work. Due to a lack of documentation for the collections of legacy removable media in FHL's backlog, staff have found that these items' complicated provenance requires an archivist's expertise too often for it to be a good student task. Nevertheless, aiming the documentation toward undergraduates means that it was written at a level that facilitates training new and nonspecialized professional staff. In 2022, FHL filled a pandemic-era vacancy by hiring an archivist whose primary responsibility is Quaker meeting records. The documentation streamlined her onboarding into the born-digital accessioning workflow and has assisted her in working largely independently on routine born-digital meeting records. This represents a major victory in the process of designing and documenting a born-digital accessioning workflow, as training "non-digital" archivists on born-digital accessioning has allowed a more sustainable division of accessioning labor in the department and a standardization of that labor across professional roles.

In 2023, Higgs Kopin moved into a new position at Swarthmore College Libraries, and FHL archives technician James Truitt was promoted to digital archivist. Because both Truitt and Higgs Kopin were employed by FHL prior to the transition and both remained at Swarthmore College Libraries in their new positions, they were able to plan for extensive cross-training and continue meeting throughout the transition process. Hutchison's work on creating and compiling documentation

eased Truitt's transition into the new role, significantly cutting down the amount of technical training required as he was able to work through an existing accessioning manual. Consequently, Higgs Kopin and Truitt were able to allocate more time to discussing decision-making processes, understanding the context of past policy decisions, and identifying areas of difficulty that could use a novel approach. Higgs Kopin and Truitt were able to treat the documentation as a first draft that Truitt was invited to change and improve. Version control on that documentation has allowed for transparent changes and iterations to the workflow by Truitt, which are described in the next section. The overall success of this transition was aided by the depth of documentation provided, the availability of both colleagues pre- and post-transition for support and training, and the intention set by Higgs Kopin and Truitt that healthy workflows are highly collaborative and iterative endeavors.

Iteration

FHL staff designed the accessioning workflow with the intention of iterating upon it, and this framework has allowed us to continually improve our processes as we have gained experience using them. One of the first improvements made to the baseline accessioning workflow was an addendum to the deed of gift and deposit agreements. This addendum, a "digital preservation agreement," allows FHL to acquire donors' informed consent regarding the level of preservation to which FHL commits and gives donors the opportunity to choose from a small menu of options for conditions on access (e.g., unrestricted online access, on-site access only, etc.).[20] By building donors' understanding of what FHL will do with their materials and giving them some control over access conditions, the agreement helps avoid miscommunications with individual donors, demonstrates FHL's commitment to its donors at large, and gives repository staff a sound foundation for choices regarding access to born-digital material.

Some tweaks to the existing workflow steps have also been made. For example, the exact procedures for virus scanning, packaging, and transferring to preservation storage have been slightly changed to account for quirks of the command line utilities. Other larger changes are still being contemplated; for example, one complication of the current workflow is that the bagging process overwrites the original timestamps for file creation and modification, because (by default) Bagger creates copies of the files being bagged. Since the first draft of the workflow, one workaround that has been found for this has been to use the "Bag in Place" function, which bags the original files rather than creating copies and facilitates the addition of tag files. However, this function tends to scramble the order of the metadata fields in Bagger's graphical user interface (GUI). One solution to this issue might be moving to a different GUI application like DART for bagging.[21] FHL staff have also experimented with generating digital forensics XML (DFXML) files with the walk-

to-dfxml utility.[22] Another solution under consideration is running the software Brunnhilde more regularly.[23] FHL usually runs Brunnhilde with BulkExtractor to conduct personally identifiable information (PII) scans, but the tool's core capabilities, including documenting file modification times, may make it worth using on a broader basis.

Reflections

Archival Education

Archival education is often focused on best practices, which are necessary standards for the archival profession. However, these best practices and ideal situations are often not what digital archivists find once they are employed in the field, and they can be limiting to staff making an initial attempt to steward born-digital archives. In their 2020 case study "Preservation Not Paralysis: Reflections on Launching a Born-Digital Preservation Program," authors Kyna Herzinger, Caroline Daniels, and Heather Fox point out that "available resources tend toward one of two extremes, consisting of either broad conceptual models that seem disconnected from the concrete actions and specific tools necessary to steward born-digital content, or one-off projects that are so specialized as to seem impractical when applied to different circumstances. . . . The beginner's steps seemed hidden in a complex end goal."[24]

Truitt, who succeeded Higgs Kopin as full-time FHL digital archivist in 2023, is currently a part-time MLIS student at San José State University and has found significant discrepancies between the iSchool digital preservation curriculum and the accessioning workflows developed by colleagues. iSchool classes tended to heavily emphasize disk imaging, the OAIS model, and Trusted Digital Repositories. These are all worthy topics, but they primarily reflect the capabilities and concerns of large, well-resourced institutions. Across-the-board disk imaging has been falling out of favor among digital archivists as part of a broader turn away from a forensic approach,[25] and the many functional entities of the OAIS model and the high standards of Trusted Digital Repository certification are out of reach for the many small and medium institutions that can dedicate only one full-time employee at most to their born-digital records.[26] The focus on these topics may be due in part to the backgrounds of the faculty teaching these classes, who had done digital archives work at large, robustly-staffed universities and state libraries.

It is worth reflecting on how much of the literature on born-digital records comes from these larger institutions that are richer in staff and resources than the vast majority of cultural heritage organizations, and on how their perspective might shape archival education and the discourse in the field. For example, during a discussion of the ethical considerations of disk imaging, one of Truitt's professors related that his

institution had such a large backlog of disk images that he expected any concerns over privacy or confidentiality to be negated by the amount of time items would have to wait to be processed. This scale of storage is out of reach of most archives. As A*CENSUS II found, about 60 percent of archives have three or fewer full-time equivalent staff and an annual nonpersonnel budget of less than $50,000.[27] These institutions likely cannot afford to store terabytes of bulky, unprocessed disk images for decades on end. Likewise, much early and foundational work in digital archives was produced by large well-resourced institutions because they were the only ones that had the technical capacity, staff bandwidth, and financial means to grapple with born-digital records at the time. When standards are developed by such institutions, who is left out, or left without practicable guidelines? As more and more cultural heritage organizations begin grappling with born-digital records, will they be well served by existing standards developed by large, well-funded repositories?[28]

Best Practices and "Good Enough"

Even after iterations since the original 2022 draft, the FHL born-digital accessioning workflow breaks with a number of best practices. To return to a previously cited example, FHL usually forgoes pretransfer forensics and dedicated transfer tools in favor of ease of use for donors, as the anxiety of donors is a larger issue in receiving and preserving records than the preservation of precustodial metadata. Some of these breaks stem from a misalignment with FHL's institutional profile as a repository for private organizations not subject to the same regulations as government organizations. Much of the born-digital archives literature comes out of government repository use cases and their mandate to comply with records retention laws, and FHL shares some but not all of the priorities of these repositories. By skipping certain steps, FHL is able to minimize its backlog while also providing *reasonable* assurance of the longevity of born-digital materials and the preservation of information crucial to FHL's particular community of donors, depositors, and researchers.

Readers accustomed to reviewing born-digital workflows may notice other actions that are not included in the accessioning workflow presented here, particularly regarding file formats, PII, and deduplication. While FHL staff may characterize file formats during full processing using Brunnhilde, characterization and validation have not been necessary for most accessions, beyond a human-generated description of the file types (e.g., "16 PDFs"). Staff also do not normalize files during the accessioning process. As Trevor Owens has pointed out, far more digital content is lost to failures of bitstream preservation than to obsolescence of file formats, and FHL staff are reluctant to focus on the latter until we have a good grip on the former.[29] PII scans are bypassed for many materials because of the predictable nature of the records (Quaker meeting minutes) and precustodial assurances.

FHL does not currently deduplicate as part of accessioning. The repository simply does not have enough duplicated materials in most of its born-digital accessions to warrant this step. However, this may change as workflows evolve and staff continue working through the backlog of removable media. FHL may also apply such a policy selectively; for example, much of its removable media comes from a single organization, and the contents are thus more likely to be duplicative.[30]

As illustrated in the many iterations of these workflows, none of these decisions are final. As priorities and capacities shift, the best practices currently omitted from the accessioning program may become goals that FHL staff can reach in the future. However, by freeing staff from the need to have the final answer worked out today, this iterative approach has allowed FHL to take steps toward preserving born-digital records for the long term.

Conclusion

FHL staff have found that a robust accessioning workflow can address baseline digital preservation concerns for born-digital materials upon their receipt through stabilization and basic description, thereby laying the groundwork for staff without digital expertise to process much born-digital material. Further, this case study highlights the necessity of iteration and starting with reasonable, achievable workflows that are not necessarily immediately compliant with best practices. In particular, FHL was able to develop and iterate on a baseline born-digital accessioning workflow through prioritizing the following considerations:

- The abilities and needs of our donor community may take priority over best practices that are necessary for other institutional contexts, especially government records repositories.
- Best practices are contextual and will vary from institution to institution, both in terms of viability and usefulness.
- Workflows are never complete or correct, and they improve through iteration and collaboration with professional peer groups, official standards, and within staff teams.

Ultimately, while best practices provide guiding principles and ideals to strive for, practical applications in born-digital archiving can differ quite a bit from professional recommendations and dominant modes of archival education. For smaller institutions in particular, "perfect" can be the enemy of "good"; archives workers should not let an imagined ideal prevent them from building workflows to provide a sustainable way forward for born-digital collecting. It is the authors' intention to share our case study so that other archives workers in similar institutions might replicate this process of considering best practices and moving forward iteratively by weighing their own institutional priorities against the recommendations of the professional literature.

The newly published *Best Practices for Archival Accessioning*, and particularly the combination of best practice statements with additional recommendations for "going further," acknowledges this tension between what is "best" and what is crucial as a baseline for a successful accessioning program. *Best Practices* defines many of the same goals that FHL concurrently sought to identify and fine-tune throughout this multiyear process, and it also provides steps for building on archival standards through iteration, based on available resources and institutional goals. At FHL, we can use the *Best Practices* to review our existing policies, identify gaps and opportunities for improvement, and implement new workflows based on what is best for the needs of our donors, our institutional goals, our available resources, and our priorities and timelines. We encourage archival workers at other institutions—particularly those intending to grow their capacity for born-digital archival stewardship—to prioritize building and documenting a strong accessioning program based on the *Best Practices* in conversation with what is "good enough" in their own contexts and with iteration as the mechanism for success.

Appendix: Resources Consulted

Annand, Susanne, Sally DeBauche, Erin Faulder, Martin Gengenbach, Karla Irwin, Julie Musson, Shira Peltzman, Kate Tasker, Laura Uglean Jackson, and Dorothy Waugh. "Digital Processing Framework." eCommons: Cornell's Digital Repository, August 2018. https://ecommons.cornell.edu/server/api/core/bitstreams/c96991df-7c49-4110-8b03-1828b017d8e5/content.

BitCurator. "BitCurator." Accessed September 30, 2024. https://bitcurator.net/bitcurator/.

Dundon, Kate, Laurel McPhee, Elvia Arroyo-Ramirez, Jolene Beiser, Courtney Dean, Audra Eagle Yun, Jasmine Jones, et al. "Guidelines for Efficient Archival Processing in the University of California Libraries (Version 4)." UCLA Library, May 1, 2020. https://escholarship.org/uc/item/4b81g01z.

Erway, Ricky. "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media." Dublin, OH: OCLC Research, 2012. http://www.oclc.org/research/publications/library/2012/2012-06.pdf.

Invest In Open Infrastructure. "Infra Finder." Accessed September 30, 2024. https://infrafinder.investinopen.org/solutions.

Owens, Trevor. *The Theory and Craft of Digital Preservation*. Baltimore, MD: Johns Hopkins University Press, 2018. https://doi.org/10.31229/osf.io/5cpjt.

Schaefer, Sibyl, Cynthia Wu, Eld Zierau, Nancy McGovern, and Andrea Goethals. "Digital Preservation Storage Criteria." Center for Open Science, 2018. https://doi.org/10.17605/OSF.IO/SJC6U.

Searcy, Rachel. "Beyond Control: Accessioning Practices for Extensible Archival Management." *Journal of Archival Organization* 14, nos. 3–4 (2017): 153–75. https://doi.org/10.1080/15332748.2018.1517292.

Searcy, Rachel, and Weatherly A. Stephan. "ArchivesSpace Manual for Local Usage at NYU (Version 2)." Google Doc. Accessed September 30, 2024. http://bit.ly/nyu-archivesspace.

Notes

[1] National Best Practices for Archival Accessioning Working Group, *Archival Accessioning Best Practices* (GitBook, 2024), https://accessioning.gitbook.io/archival-accessioning-best-practices.

[2] "Friends Historical Library," Swarthmore College, accessed September 30, 2024, https://www.swarthmore.edu/friends-historical-library, captured at https://perma.cc/Y4AM-DXT3.

[3] FHL has "deposit" agreements with Quaker meetings that allow for the meeting to remove records. Although these agreements differ in many ways from donations, similar archival processes are performed on both types of acquisitions.

[4] "Meeting" is the name for a congregation or organizational unit in the Society of Friends, ranging from preparative and monthly meetings (equivalent to a church, worship group, or parish) to quarterly and yearly meetings (larger administrative units). For more information, see "A Friendly Glossary," Friends Historical Library, https://www.swarthmore.edu/friends-historical-library/a-friendly-glossary, captured at https://perma.cc/68EP-WAZF.

[5] This rule advises keeping three copies of valuable digital content on at least two types of storage media, at least one of which should be geographically separated from the others. See, for example, Paul Ruggiero and Matthew A. Heckathorn, *Data Backup Options* (Washington, DC: United States Computer Emergency Readiness Team, 2012), https://www.cisa.gov/sites/default/files/publications/data_backup_options.pdf.

[6] "Levels of Digital Preservation," National Digital Stewardship Alliance, April 25, 2019, https://doi.org/10.17605/OSF.IO/QGZ98.

[7] For example, Mathew Kirschenbaum, Richard Ovenden, and Gabriela Redwine, *Digital Forensics and Born-Digital Content in Cultural Heritage Collections* (Washington, DC: Council on Library and Information Resources, 2010), p. 38.

[8] Coauthor Emily Higgs Kopin also participated in the National Best Practices for Archival Accessioning Working Group and contributed to the *Archival Accessioning Best Practices*.

[9] The BitCurator Environment is a "stack of free and open source digital forensics tools and associated software libraries, modified and packaged for increased accessibility and functionality for collecting institutions." "BitCurator," BitCurator, accessed September 30, 2024, https://bitcurator.net/bitcurator/, captured at https://perma.cc/4W6D-RQK3.

[10] For example, Kirschenbaum, Ovenden, and Redwine, *Digital Forensics and Born-Digital Content*, p. 38; Best Practice 8.2: "Generate a file directory and checksum before transfer."

[11] These practices align with Best Practices 8.5 ("Assess born-digital materials within a quarantined environment"), 8.6 ("Run a virus scan on born-digital materials"), and 8.7 ("Use a write blocker when interacting with new collections").

[12] "Disk Imaging Decision Factors," DANNNG!, January 10, 2022, https://dannng.github.io/disk-imaging-decision-factors.html, captured at https://perma.cc/MK7M-L2HM.

[13] Compare Best Practices 11.1 ("Create an accession record for each new collection and accretion as a standard part of the accessioning workflow"), 11.2 ("Create accession records promptly after receipt of materials"), and 11.6 ("Consistently document the presence of special formats in accession records").

[14] Bagger (https://github.com/LibraryOfCongress/bagger, captured at https://perma.cc/2JPU-BHGM).

[15] At the time of writing, FHL staff are developing more automated fixity checking.

[16] This is in keeping with Best Practice 10.5, "Deposit digital files into designated storage for digital archival materials."

[17] For example, the command-line utilities rsync and robocopy, or the applications Grsync and Teracopy. This aligns with Best Practice 8.4 ("Use a dedicated file transfer utility to move digital materials to digital collections storage").

[18] "Internship and Capstone Opportunities," Drexel University College of Computing & Informatics, June 29, 2023, https://drexel.edu/cci/student-experience/career-resources/internship-and-capstone-submission-form/, captured at https://perma.cc/G2LT-24N2.

[19] "Successful Accessioning Labor Practices," in *Archival Accessioning Best Practices*.

[20] This addendum was implemented before the publication of *Archival Accessioning Best Practices* but aligns with Best Practice 4.9 ("Include born-digital materials in legal agreements").

[21] DART (https://aptrust.github.io/dart-docs/, captured at https://perma.cc/E7ER-H8PM) is a tool for packaging files and uploading them to remote repositories; it can be run through a GUI or the command line.

[22] DFXML Working Group, https://github.com/dfxml-working-group/dfxml_python, captured at https://perma.cc/A8CA-TNWC.

[23] Brunnhilde (https://github.com/tw4l/brunnhilde/, captured at https://perma.cc/RKC7-XSN4) is a command-line utility for generating aggregate reports on directory structure, file sizes and modification times, duplicate files, and file formats. It relies on the file-characterization tool Siegfried (https://www.itforarchivists.com/siegfried), and can optionally be run with the PII-scanning tool BulkExtractor (https://github.com/simsong/bulk_extractor, captured at https://perma.cc/63GK-VMVD) and virus-scanning tool ClamAV.

[24] Kyna Herzinger, Caroline Daniels, and Heather Fox, "Preservation Not Paralysis: Reflections on Launching a Born-Digital Preservation Program," *Collections* 17, no. 4 (2021): 348, 354, https://doi.org/10.1177/1550190620978221.

[25] "Disk Imaging Decision Factors," DANNNG!

[26] Makala Skinner, "A*CENSUS II: Archives Administrators Survey," *Ithaka S+R*, January 31, 2023, https://doi.org/10.18665/sr.318227.

[27] The overwhelming plurality of archives surveyed (44%) had an annual operating budget of less than $20,000 (excluding personnel). The next largest category, with budgets between $20,000 and $49,999, accounted for just 17% of archives, and the drop-off continues from there. Skinner, "A*CENSUS II: Archives Administrators Survey."

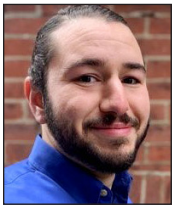
[28] After this article had gone through peer review, Truitt had the good fortune to take another digital preservation class at San José State. This course, taught by Dr. Katherine Skinner, more consistently highlighted the perspectives of small and underfunded institutions.

[29] Trevor Owens, *The Theory and Craft of Digital Preservation* (Baltimore, MD: Johns Hopkins University Press, 2018), https://doi.org/10.31229/osf.io/5cpjt.

[30] Deduplication can lead to significant storage savings. In one case, preservationists at the University of Kentucky managed to reduce the size of an accession by 90% by discarding duplicate files. Megan Mummey, Andrew McDonnell, Emily B. Collier, Sarah Dorpinghaus, Ruth E. Bryan, "Will Our Future Selves Thank Us? An Examination of Born-Digital Curation Practices at the University of Kentucky Libraries," Digital Library Foundation Forum, St. Louis, MO, 2023, https://uknowledge.uky.edu/libraries_present/263. More broadly, recent attempts to generate metrics for born-digital processing indicate that processing reduces storage size by an average of 80%, though some collections actually end up growing as, for example, zip files are unpacked. Carol Kussmann and Lara Friedman-Shedlov, "Insights and Exploration of Metrics for Born Digital Records Processing," Best Practices Exchange, Sacramento, CA, 2024, https://bpexchange.wordpress.com/wp-content/uploads/2024/08/2024-bpe-kussmann-friedman-shedlov-notes.pdf.

ABOUT THE AUTHORS

Emily Higgs Kopin is the head of digital collections strategy for the Swarthmore College Libraries, where she oversees the management and preservation of digital special collections and other library-owned assets. Previously, she was the digital archivist at the Friends Historical Library of Swarthmore College.



James Truitt is the digital archivist of the Friends Historical Library of Swarthmore College, where he previously served for several years as an archives technician. He is an advocate for the wider use of regular expressions by archivists and is currently an MLIS student in his final year at San José State University.

Jessica Hutchison is a project archivist at the American Philosophical Society (APS), where she processes History of Science collections. Prior to the APS, she received her MSLIS from Drexel University in 2022 and has held positions at several institutions in Philadelphia, Pennsylvania, including the Penn Museum Archives and the Library Company of Philadelphia.