Starting with the Digital Doesn't Make it Easier: Developing Transparent Born Digital Acquisition Policies for Archives

Amanda Boczar
University of South Florida

Abstract

As organizations continue to overwhelmingly abandon all forms of paper-based record keeping, libraries are still adapting to increased offers of born digital archival donations. Simple misunderstandings or disconnects between the units facilitating donations and maintaining born-digital collections creates pain-points for donor relations and can result in a lack of transparency over how their records may be processed. To facilitate better donor transparency and cross-area collaboration over born digital records, Special Collections and archives need comprehensive policies and shifts in training and collaboration paradigms. This paper analyses the intersections of born digital archiving, collection development polices, donor relations, human-supported AI tools, and digital records education within American academic libraries to propose a functional toolkit for born digital acquisitions. Unrealistic expectations of collection processing, retention, growth, and publication onto openly accessible platforms can quickly overwhelm a libraries' digital collections' team due to size, need for digital forensics work, copyright limitations, or other capacity-related issues. Intertwined within this discussion is an additional discourse over the need to carefully curate our digital spaces not only for practical cost reasons, but due to the environmental costs of massive data storage solutions. Through an analysis of the elements stated above, the paper will reflect on the need to integrate born digital materials into archival acquisition procedures and provide practical solutions to meet this need.

Introduction

As organizations continue to overwhelmingly abandon all forms of paper-based record keeping, libraries are still adapting to increased offers of born digital archival donations. Scholars have published about the importance of maintaining born digital archives throughout the first two and a half decades of the twenty-first century, but many archives and special collections departments are still struggling to integrate these policies smoothly. Within American academic libraires, born digital archives are frequently lumped into special collections donations policies and paperwork by name only, without concrete accession and preservation practices. When institutions are fortunate enough to have staff focused on born digital archives, they are not always integrated in the donation process, creating knowledge gaps and confusion for both donors and those on the library side. Librarians and curators managing archives rarely have training in born digital archives management, and materials may be brought to digital collections, digital scholarship, or systems units with limited input before the paperwork is completed. In one recent encounter at my home institution, a well-meaning colleague who does not work with digital records suggested that a donor strategy might be to accept born digital records first in order to get ahead of the slog of processing physical donations and winning donor favor since the born digital items would be easier to process. Unfortunately, the collection included several intellectual property issues halting its acceptance and creating confusion for the donor. Simple misunderstanding or disconnect between the units facilitating donations and maintaining born-digital collections creates pain-points for donor relations and can result in a lack of transparency or fractured trust over how a donor's records may be processed.

American academic library archives are not always equipped with faculty or staff trained in digital curation practices, and courses on these subjects only became common within the past few years meaning library leadership will have rarely had any formal theoretical education on the subject. Archives and records management courses within American Library Association (ALA) accredited Master's programs overwhelmingly focus on paper-based or otherwise physical artifacts. Numerous professional development options exist, but can be cost or time prohibitive. Born digital donation acquisitions often require specialized knowledge of at least one of the following areas: obsolete and legacy media reformatting, digital storage, digital object organization, data management, or digital forensics. In addition, processes to make the collections usable for researchers can be time prohibitive without the ability to write code for large-scale data processing using shell scripts, XML, Python, or other programmatic solutions. These skillsets are rarely sought and hired within archives, however, and tend to be found in separated digital scholarship or systems units who face their own scaling problems. Facilitating work between units is possible but without initial evaluation of born digital assets, donations can be mired in disconnects over the rationale for digital curation strategies. Born digital archiving represents a time consuming, albeit necessary and significant, aspect of academic librarianship for the foreseeable future. This work, while similar, is distinct from that done by digitization units and requires its own workflow. When archives units are unable to address the curation needs of a donor's digital or mixed physical and digital format donations in clear, confident, and transparent ways, libraries risk appearing unprepared or inept at managing their resources.

Archives without a clear policy or dedicated professionals managing born digital acquisitions may find themselves either refusing collections of significance or taking collections they cannot appropriately manage. One common misconception that can result from the lack of a policy direction is the idea that donations should start with born digital content due to them being easier or less costly to process. Several factors challenge this misconception, as born digital donations carry similar challenges as reformatting physical collections. Items need to be assessed for copyright and intellectual property status to determine retention and methods of access, which often serve as the main barriers for making born-digital records accessible (Jaillant, 2022). Digital objects may contain private or confidential information, which can exist both in text or in embedded file metadata. Donations are often bloated with countless versions, duplications, or unrelated saved files that should be weeded prior to providing public access. Each of these issues

are similar to those faced by digitization units, however, born digital archives can be deceptively large as tens of thousands of records can be transported on a single hard drive or via cloud storage. For libraries without the capacity to curate and weed born digital archival donations, they quickly face problems with the cost and limits of digital storage, the need to evaluate and maintain digital records across a much shorter lifecycle, and the paradigm against deleting or destroying records based on the core American Library Association value of preservation (ALA, 2006; 2024). Updated in January 2024, the ALA Core Values indicate support for making educated choices with an eye toward sustainability and climate resiliency that will support justifications for born digital efforts in the future.

Due to these factors, it is necessary that born digital donations face more scrutiny than physical donations and that donors are made aware that items in their donations may be edited or deleted. It may seem easier to justify the weeding or deaccession of mail coupons or sensitive financial record from a physical archive than it is to delete private or redundant files from a donated hard drive since the objects are less visible to the donor and there may be concerns over the erasure of their archive, but early discussions can help to frame expectations. Donors may also not understand that having digital surrogates of materials on their donated hard drives does not equate to copyright ownership and drastically limits what can be done with their materials. Only in very rare cases, like the computers of author Salman Rushdie at Emory, would a library be able and interested in maintaining an exact replica of a donated digital environment (Rockmore, 2014; Waugh et al, 2016). In reality, donations will likely be reformatted, curated, have metadata applied, and uploaded to a digital collections repository or local access platform. Often, archivists weed collections without consultation of the donor by using the approvals granted on most Deed of Gift agreements. When archivists or subject curators do not address issues related to weeding and retention honestly with donors, in the way that an in-person review of physical archives would occur, donors may not understand that their records may be reviewed, rejected, or deleted upon gifting to an archive. Further, if the archivists or subject curators do not realize this weeding may need to occur for digital items, they may make unrealistic promises to donors about the creation of digital collections. Collections accepted on a loan or proof-of-concept model present even greater challenges to the work of curators. When born digital donations are accepted without careful evaluation or a clear distinction as a gift, the library takes on considerable risk related to donor satisfaction over the curation that will certainly need to occur with those records to avoid burying the library in a time-consuming project. Involving digital archivists or curators early in the process will help to prevent misunderstandings over donor expectations related to born digital and have the added benefit of helping to prevent over promising on digitization agreements as well.

Due to the lack of a dedicated digital curator at the University of South Florida (USF) until 2022, for many years born digital archival acquisitions were handled by archivists or librarians without a background in digital libraries, digital archiving, or digital preservation. When possible, the digital scholarship or digitization units provided advice, but were not the drivers of new projects. A review of posted policies and interviews with selected libraries reflected similar approaches across American Research Libraries (ARL) members, and universities in the American Association of Universities (AAU). To serve the donor community, USF had been creative in accepting donations over the years making use of internal departmental storage, collaborating with campus IT, or engaging with other organizations to support their growing digital archives. In other cases, archivists could only inventory donations on storage media as archival objects within the finding aids by recording any information on the labels, but the media was only checked if requested by a researcher. Due to a lack of equipment, funding, or expertise to reformat obsolete media, some donations remained in their original formats for decades. Physical decay of legacy media and obsolete software have greatly increased the urgency and difficulty of recovering the archives. Following a reorganization that placed Digital Collections into Special Collections and an investment in specialized subject curator positions to compliment general archivists at USF, a shift has occurred allowing for a more careful evaluation of materials. The new hires include the author's position as the Curator for Digital Collections, as well as specialists in two of five strategic collecting areas. In the past year, over 10 TB of born digital records have been donated to the library as part of one institutional records

transfer, drawing attention to the need to holistically address born digital acquisitions across subject areas. To facilitate the processing of new donations, the author is utilizing human-supported AI models that can be trained to recognize people, places, handwriting, and voices. The long-term repercussions of that work are unknown, but will need to be integrated into the discourse with donors due to privacy concerns over training AI algorithms on materials as part of the curation process.

To facilitate better donor transparency and cross-area collaboration over born digital records, Special Collections and archives will benefit from comprehensive policies and shifts in training paradigms. This paper analyzes the intersections of born digital archiving, collection development polices, donor relations, human-supported AI tools, and digital records education within American academic libraries to outline necessary shifts in the field to broadly enact best practices for born digital acquisitions. Unrealistic expectations of collection processing, retention, growth, and publication onto openly accessible platforms can quickly overwhelm a libraries' digital collections' team due to size, need for digital forensics work, copyright limitations, or other capacity-related issues. Intertwined within this discussion is an additional discourse over the need to carefully curate our digital spaces not only for practical cost reasons, but due to the environmental costs of massive data storage solutions. Through an analysis of the elements stated above, this paper will showcase a clear need to more explicitly integrate the unique challenges and benefits of born digital archives into acquisition procedures and provide practical solutions to meet this need. Through an assessment of industry best practices and an evaluation of how peer and aspirant libraries are responding to those best practices, this paper will seek to provide a practical framework for libraires to consider how they can start to address the growing demand for born-digital library support in a way that puts transparency with donors at the forefront. Addressing what donors should know about their born digital donations in clear and concrete ways should lead to better interdepartmental communication and improved services with an eye on ethical weeding and preservation practices.

Literature Review

Existing literature on the topic of born digital archiving, acquisitions, and collection management focus on issues related to the scope, access strategies, and the difficulties in maintaining effective born-digital preservation strategies. Articles frequently point to an interest across the field in ensuring that the cultural memory of born digital spaces are preserved in some way. Most of these works were published in the twenty-first century but are preceded by a robust field of scholarship dedicated to similar concerns over digitized collections. Few works address the complex intersections of these fields with issues of donor relations, archival education, and the climate impacts of exponential growth. In addition to the literature prepared by scholars, existing born digital acquisition policies from organizations both within (Yale's Beinecke Library)[1] and out of scope (New York Public Library)[2] support the findings of the study.

As described in the introduction, Emory University was an early adopter of comprehensive born digital archival procedures. An article from the team at that time titled, "A Comprehensive Approach to Born-Digital Archives," describes how the Rushdie donation served as a critical impetus for taking on a more comprehensive approach to born digital archiving (Carroll et al., 2011). Emory's impressive undertaking involved a careful consideration of donor requests along with the desire to make the materials as accessible as possible. The notoriety of Rushdie placed even greater pressure on the library to produce a high-quality archival experience, as coverage from *The New Yorker* indicates (Rockmore, 2014). Building on the work of Carroll et al. at Emory, Jan Zastrow's 2016 conference report "Digital Acquisitions and Donor Relations: Assets, Apprehensions, and Anxieties" focuses on the importance of keeping donor expectations brief and anxiety free. Part of her approach includes remaining format agnostic in the acquisitions

---

[1] Yale Beinecke Library, https://beinecke.library.yale.edu/article/born-digital-archival-acquisition
[2] New York Public Library, https://nypl.github.io/digarch/sitevisits/acquiring

process. While this is helpful in completing the donations, the additional benefits of collaborating between the physical and digital archivists have the potential to further improve services. Sloyan (2016) pays careful attention to the process of digital appraisal and sensitivity review in their article. Although the focus for their work takes place in a health library, the issues are applicable across library systems in different ways. The balance over how to address speed and thorough review are critical elements that continue to drive decision making.

More recently, Velte and Wikle's 2020 study "Scalable Born Digital Ingest Workflows for Limited Resources: A Case Study for First Steps in Digital Preservation" focuses on taking small and management steps to offset the impact of born digital donations at the University of Idaho. Of particular interest in their article is an outstanding literature review and detailed workflows covering best practices for born digital archives and digital preservation practices serve as an excellent resource for any academic library seeking to improve their services. Their focus is not on donor relations and transparency, however. Acknowledging that best practices are well established for small and large libraries for processing born digital donations, this work now turns to an analysis of the acquisitions policies of ARL and AAU libraries and how those policies impact transparency.

Methods

To better understand how USF's repositioning on born digital archives fits within the state of the field, the author reviewed the published references to born digital collections and policies for the libraries at the 107 ARL or AAU member universities and conducted interviews with faculty and staff at those institutions. These limitations are placed on the survey to ensure that the libraries share similar missions by serving major American academic universities. The author's home institution is a member of one of these organizations, but not the other, making the survey population reflective of the organization's aspirations.

In addressing policies, procedures, and published information related to born digital in ARL and AAU libraries, the author made a point to specifically address born digital materials donated to special and digital collections units. This is a conscious choice to avoid pulling the thread on the definition of born digital assets that may also include everything within the institutional repository including electronic theses and dissertations, faculty publications, and open access journals. Some may also include in their definitions born digital media purchases made by circulating and e-resources departments that would quickly distract from the intended scope of this paper. One area that did receive consideration, but does occasionally fall outside Special Collections, are university archives and records management. University archives are often a fundamental element of special collections departments, including at the author's home institution, and work with intra-institutional donors present their own challenges that the solutions in this paper can likewise help to mitigate. While not the primary focus of this research, oral histories and web archiving projects are also considered part of the born digital archival strategy when self-identified as the focus by the institution.

Published Policies Review

To place USF within conversation with other academic libraries, the author conducted an observational assessment of the websites and published materials for the 107 ARL and AAU libraries. The goal of the review was to identify if those institutions had clearly published born digital acquisitions policies, workflows, or dedicated digital archivists or curators to manage such work. University libraries for each institution were searched broadly, and specific attention was not granted to any one campus or subject-specific library within an organization.

Interviews

To better understand the observational assessment, the author also conducted a preliminary set of qualitative interviews with eleven librarians and curators from seven selected ARL and AAU institutions sharing diverse perspectives based on their public/private status, size, endowment, and age. The small sample size that would benefit from further investigation, but efforts were made to seek out a representative sample. The interviews sought to provide context for the published details about born digital archiving practices and seek a stronger understanding of the published policies. The interviews allowed the author to directly engage with practitioners about how their libraries evaluate and accept born digital archives, who within the library managed born digital donations, how they engage with donors, how they manage exponentially increasing storage demands, and how they respond to issues related to climate change. Seven libraries responded to invitations to discuss their practices under an understanding of anonymity, with eleven faculty and staff sitting for interviews. The questions asked in the interviews are listed in Table 1.

Table 1. Interview questions.

1	Does your library have a formalized born-digital acquisition policy?
2	Who is responsible for evaluating and accepting born-digital donations? What is their educational/experience background? For example, digital humanities, digital libraries, archives, systems, or computer science.
3	Is the digital collections unit specifically funded to handle born-digital needs, such as server space and curation of digital materials?
4	Approximately how many donors do you work with each year who are donating born-digital materials? Roughly how many TB of content did your library accept last year in born-digital donations? Less than 1, 1-5, 5-10, more than 10.
5	When working with born-digital donors, when in the process do you address issues such as copyright, weeding collections, and retention of materials?
6	Was your born-digital acquisitions policy written to give consideration to issues related to climate change?
7	How are weeded materials disposed of? Are any materials returned to the donor as part of the born-digital acquisitions process?
8	If your institution received funding to enact one change to improve your overall born-digital acquisitions policy, what would you do?

Results

In reviewing the published online polices and descriptions for born digital archival donations, the author recognized a stark lack of consistency related to how libraries approached their born digital archives in relation to more established library units including the physical archives, cataloging, technical services, or even digital scholarship. Organizational structures and job titles vary heavily from institution-to-institution, and many libraries have posted searches for digital archivists in the last five years. A more targeted assessment of the job advertisements may provide additional relevant details, but not all calls are still open or easily accessible.

Historic institutions with large endowments tend to have the most clearly identifiable born digital archives policies, which excellent online documentation from field leaders like Yale University and Princeton University. [3, 4] To serve their fifteen branch libraries, Yale provides a full Born Digital Accessioning Service for Yale Special Collections, which includes publicly shared steps for born digital archival accessioning and their processing worksheet (Kuhl, 2018). Ivy League universities are not the only libraries with clearly documented practices, with a

---

[3] Yale University, https://guides.library.yale.edu/borndigital
[4] Princeton University, https://library.princeton.edu/special-collections/workflows/born-digital

notable example from the University of California system, including UCLA (University of California Systemwide Libraries, 2017). One of the most recent approaches to drafting a strategic approach to born digital acquisitions practices comes from the University of Kentucky, who presented their efforts at the 2023 Digital Library Federation Forum (Mummey et al., 2023). Kentucky's work is notable for their transparency over the cost of born digital growth and their efforts to weed born digital acquisitions. While these programs show clear strides in outlining or evolving their born acquisition polices, they are in the minority of the observed libraries when it comes to discussing born digital acquisition policies directly.

Through the interviews with eleven faculty and staff across seven ARL and AAU libraries, no two libraries reported the same approach to their born digital acquisitions. Three of the seven reported that they did not have a born digital acquisitions plan at all. Four of the universities stated that they had at least a general policy, as covered in their Special Collections Deed of Gift. Born digital is rarely distinctly addressed in archival policies, but rather listed as a donation type or covered in general Deed of Gift language, and thus covered by weeding policies drafted with physical archives in mind.

One area that interviewees all addressed was how their organizational structures supported or challenged their efforts. Professionals responsible for managing born digital acquisitions had roles in many different units including special collections, digital collections, digital preservation, or digital scholarship. Organizational divisions between digital and physical archives can at times create confusion over donor relations responsibilities and a resistance to overstepping across unit responsibilities. Only one interviewee reported that their staff responsible for processing digital archives was always included in donation conversations. When asked how many born digital acquisitions were brought in during the past year, two of the seven interviewees were completely unsure how many the special collections unit had brought in. Only one of the interviewees could confidently report their born digital acquisitions. Some universities relied primarily on their special collections teams to make decisions, while others occasionally made decisions using a cross-departmental committee structure. With the often split services across departments for those who work on born digital, budgets prove to be another area where there is little consistency. Outside of institutional repository server space or hosted digital preservation solutions, few libraries have dedicated budget lines for acquiring, weeding, and preserving born digital archives. Each of the institutions interviewed also shared that they hoped to integrate climate impacts and solutions into their future plans and revisions as they worked to meet the increasing demands of born digital archival donations.

Lessons Learned

Viewed together, the observational policy review and interviews do not provide a cohesive transparent approach to born digital donor relations for libraires of various sizes to adapt, but rather point to several areas where libraries could streamline their workflows. Years of discussion and publishing on this topic have failed to result in a shared approach among libraries. For example, one of the areas of concerns at the core of USF's efforts is related to addressing ethics within the donation process. Only one other institution mentioned ethics during their interview as a concern with their donor relations. Each university faces unique challenges related to specific projects, collecting areas, or tools that drive their decision making. Keeping donors more aware, earlier, about potential ethics, intellectual property issues, weeding, and online accessibility of collections are all focuses of revised efforts at USF. In addition, the use of Otter.ai, Transkribus, and Amazon Web Services to train transcription models and review donations for content all involve the loading and analysis of donations by algorithms. To ensure that donors approve of this use of their materials, we have started to address this work in our donation conversations.

Through this study, four key areas emerged as possible avenues to improving transparency in donor relations for born digital acquisitions. First, and perhaps the most expected answer, being a need for greater financial support. Dedicated born digital funding or increased personnel working with the acquisitions will free up capacity in already overworked departments

who are usually engaging in decades of backlog work from legacy media donations. Funding, however, can often be one of the most difficult areas to control from within library units.

The second area of consideration is related to organizational structures. Working within or adapting organizations structures to give voice to digital archivists and curators as part of the donations process is crucial to ensuring that donors are aware of the complex processes engaged by libraries to evaluate content, check for viruses, weed, dispose of, preserve, and perform regular checksums and maintenance on their digital files. Whether digital library professionals are employed in shared units with special collections and archives or take part in cross-departmental meetings and trainings, opening communication within libraries will only further improve transparency and reduce confusion for donors as they work through the process. This includes transparency at the university level over records management and university archives, which is very well established at some organizations and less so at others.

Third, in the years to come it will be critical to keep an eye toward climate change. Climate resiliency remains a major challenge for transparent donor relations, as donors may or may not be familiar with the impacts of server use and data storage. Integrating climate into donor relations conversations, however, creates several internal and external hurdles that may extend beyond what some libraries can practically handle in the near term. One interviewee suggested that compounding discussions over climate with possible cost savings may be a way to build a bridge with library administration on the topic. As more research is done on the topic, it is likely that climate transparency will provide the next wave in adaptations to born digital archiving and donation procedures.

Finally, to improve donor transparency on born digital, it will become necessary to accept the need to formalize born digital acquisition policies. Placing limits and earmarking budgets to support stable born digital archival growth in similar ways that other library units are expected to work will elevate born digital acquisitions to a core academic library service.

Conclusions

The inconsistent adoption of born digital practices in varied library departments illustrates that ARL and AAU libraries are still adapting to the growing born digital demands of their donors. Through qualitative assessment using observational surveys and interviews, this paper provides a preliminary assessment of how libraries can improve their holistic approach to born digital acquisitions, with an emphasis on its impacts on transparency and donor relations. Few libraries are taking the same approach to born digital currently. Through internal assessment and consideration of the areas outlined above –ethics, budgets, organization structures, climate resiliency, and formalization of policies– American academic libraries can streamline services, improve their donor relations, and better meet their preservation goals for digital collections services.

References

American Library Association. (2006; 2024) *Core Values of Librarianship*, Retrieved January 25, 2024, from http://www.ala.org/advocacy/advocacy/intfreedom/corevalues

Carroll, L., Farr, E., Hornsby, P., & Ranker, B. (2011). A Comprehensive Approach to Born-Digital Archives. *Archivaria*, *72*, 61–92. https://archivaria.ca/index.php/archivaria/article/view/13360

Jaillant, L. (2022). How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. *Archival Science*, *22*(3), 417–436. https://doi.org/10.1007/s10502-022-09390-7

Kuhl, N. (2018). *Born Digital Archival Acquisition*. Beinecke Rare Book & Manuscript Library. https://beinecke.library.yale.edu/article/born-digital-archival-acquisition

Mummey, M., Dorpinghaus, S., Bryan, R., Collier, E. (2023). Conference Presentation. *Will our future selves thank us? An examination of born-digital curation practices at UKL*. Digital Library Federation Forum. St. Louis, MO.

Rockmore, D. (2014). The Digital Life of Salman Rushdie. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/digital-life-salman-rushdie

Sloyan, V. (2016). Born-digital archives at the Wellcome Library: Appraisal and sensitivity review of two hard drives. *Archives and Records*, *37*(1), 20–36. https://doi.org/10.1080/23257962.2016.1144504

University of California Systemwide Libraries. (2017). UC Guidelines for Born-Digital Archival Description. *UC Office of the President: University of California Systemwide Libraries*. Retrieved from https://escholarship.org/uc/item/9cg222jc

Velte, A. & Wikle, O. (2020). Scalable Born Digital Ingest Workflows for Limited Resources: A Case Study for First Steps in Digital Preservation. *Preservation, Digital Technology & Culture*, *49*(1), 2-13. https://doi.org/10.1515/pdtc-2020-0004

Waugh, D., Russey Roke, E., & Farr, E. (2016). Flexible processing and diverse collections: A tiered approach to delivering born digital archives. *Archives and Records*, *37*(1), 3–19. https://doi.org/10.1080/23257962.2016.1139493

Zastrow, J. (2016). Digital Acquisitions and Donor Relations: Assets, Apprehensions, and Anxieties. *Computers in Libraries*, 36(5), 16-18. https://www.proquest.com/docview/1826446270?sourcetype=Trade%20Journals