



# Beyond the workflow: archivists' aspirations for digital curation practices

Colin Post<sup>1</sup> · Alexandra Chassanoff<sup>2</sup>

Accepted: 31 May 2021 / Published online: 14 June 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

## Abstract

The documentation of archival workflows plays an important role in digital curation practice. Capturing the various steps, tools, people, and software involved at different stages, workflow documentation visually represents complex activities, and at times, invisible labor. In this article, we reflect on findings from the OSSArcFlow project, a three-year, grant-funded initiative to investigate and document workflow activities of 12 cultural heritage institutions using three open-source software systems. Building on previously published research on documentation of current digital curation practices, this article reflects on the challenges the project team encountered in modeling archivists' aspirations for their workflows. While current practices could be accurately represented in linear process models, archivists' aspirations for how they might advance digital curation practices extended beyond adding or changing discrete workflow steps and often involved sociotechnical factors that could not be easily mapped. This article presents a taxonomy of archivists' aspirations for their born-digital archives, grouping these goals together around major themes that emerged throughout the research. Our findings show that workflow documentation is an essential artifact in helping archivists to understand gaps and challenges in their current workflows and to imagine the further development of digital curation tools, systems, and practices. Project participants especially benefited from engaging in this reflection on workflow documentation as part of a community of practitioners, with the opportunity to compare across institutional contexts.

**Keywords** Digital curation · Born-digital archives · Workflows · Sociotechnical factors · Communities of practice

---

✉ Colin Post  
ccpost@uncg.edu

Alexandra Chassanoff  
achass@ncsu.edu

<sup>1</sup> Department of Library and Information Science, University of North Carolina at Greensboro, Greensboro, NC, USA

<sup>2</sup> School of Library and Information Sciences, North Carolina Central University, Durham, NC, USA

## Introduction

Workflow models can be essential tools for archivists tasked with managing the long-term care of born-digital materials in institutional settings. In the current processing environment, workflows may involve the coordination of a wide range of tools and staff, potentially across departments. Models provide a means to visualize current practices and offer space for conceptualizing how best to improve gaps in the process (Anderson 2014). Advancing methods for developing digital curation workflow models was the primary motivation of OSSArcFlow (2017–2020), an Institute of Museum and Library Services (IMLS) funded project carried out by the Educopia Institute in collaboration with faculty and graduate students at the University of North Carolina—Chapel Hill School of Information and Library Science.

This article endeavors to analyze and reflect upon the aspirations of the OSSArcFlow partners for their digital curation workflows with the benefit of applying the lessons learned from the project. We consider how archival practitioners envision the ongoing development and maturation of their digital curation workflows in relation to their current practices, their aspirational goals, and a host of sociotechnical factors, including broad organizational aims, priorities, and mandates. First, we describe the challenges of modeling aspirational workflows and how these difficulties altered the intended course of the project. Next, we analyze the goals that partners discussed for their born-digital archiving workflows. We suggest that archivists think about the advancement of their workflows as a horizon of possibilities, dynamically shaped by the tools and resources available, organizational contexts, staff skills and competencies, and discourses that cut across professional communities. A better understanding of archivists' aspirations for their workflows, along with the many factors that recursively shape maturing digital curation practices, provides the necessary foundation for ongoing discussions around the development of open-source tools and best practices for born-digital archiving.

The project team worked with 12 partner institutions (see Table 1), each implementing at least one of three open-source software (OSS) systems for born-digital archiving: ArchivesSpace (<https://archivesspace.org/>), Archivematica (<https://www.archivematica.org/en/>), and the BitCurator Environment (<https://bitcurator.net/>). The project team and partners first modeled and analyzed 'as-is' workflows, with the intent of using these as a foundation for modeling 'aspirational' workflows with goals project partners hoped to advance toward within the life of the project. While the project team successfully collaborated with the partners to publish visual workflow diagrams of these institutions' current practices (<https://educopia.org/ossarcflow/>) and used these diagrams as the basis for systematic comparison and analysis (Post et al. 2019), modeling and advancing toward aspirational workflows stalled as two key assumptions undergirding the project proved far thornier than anticipated.

First, the project assumed that the partner institutions would have relatively mature digital curation workflows already in place; in fact, a major finding of

**Table 1** Overview of partner institutions

Partner	Description	OSS Environments
Atlanta University Center, Robert W. Woodruff Library (AUC)	AUC is an independent academic library providing information services to a consortium of historically black colleges and universities	Bitcurator, Archivesspace
District of Columbia Public Library (DCPL)	DCPL is a public library system serving Washington, D.C. The Special Collections employ two digital curation librarians	Bitcurator, Archivesspace
Duke University Libraries	Duke is a private research university in Durham, NC. Duke employs staff responsible for digital curation across a number of library units	Bitcurator, Archivesspace
Emory University, Stuart A. Rose Manuscript, Archives, and Rare Book Library	The Rose Library at Emory, a private research university in Atlanta, GA, collects a range of born-digital manuscript collections	Bitcurator
Kansas Historical Society (KHS)	KHS collects materials documenting Kansas history, and serves as the official repository of government records	Bitcurator
Massachusetts Institute of Technology Institute Archives and Special Collections (MIT)	The Institute Archives and Special Collections serve as a repository for institutional records of MIT, a major research university in Cambridge, MA	BitCurator, Archivesspace, Archivematica
Mount Holyoke College (MHC)	A member of the Seven Sisters and the Five College Consortium, MHC is a small liberal arts college in South Hadley, MA	Archivesspace
New York Public Library (NYPL)	A public library system serving New York City, NYPL has three research libraries that collect archival material, as well as a department for Special Collections and Preservation Services	Bitcurator, Archivesspace
New York University (NYU)	A private research university in New York city, NYU acquires a variety of archival materials, including those housed in the Fales Collection	Bitcurator, Archivesspace, Archivematica
Odum Institute	Part of the University of North Carolina at Chapel Hill, Odum manages social science data throughout the research lifecycle	Bitcurator

**Table 1** (continued)

Partner	Description	OSS Environments
Rice University, Woodson Research Center	Woodson Research Center is the Special Collections and University Archives for Rice, a private research university in Houston, TX	Bitcurator, Archivesspace
Stanford University	Stanford is home to 23 libraries, all 19 of those under the direction of the University Librarian collect digital resources	Bitcurator, Archivesspace

the project was that these workflows remained largely ad hoc and still in the process of development. Second, the project assumed that archivists could pursue a fairly straightforward pathway in moving from as-is to aspirational workflows. At the outset, the project team imagined that partners would be able to identify specific actions that would improve their as-is workflows, and that these goals could be represented and mapped onto corresponding aspirational workflow diagrams. Although the as-is practices were amenable to visual diagramming, the discussions surrounding aspirational workflows quickly moved beyond specifying a set of linear sequences or steps that partners hoped to implement. Rather, it became apparent that a host of sociotechnical factors impacted workflow practices in both specific and generalized ways. These challenges of moving from as-is to aspirational workflows made clear the central importance of workflow models as instruments for reflective practice, helping archivists to recognize the broader socio-technical factors impacting their digital curation workflows.

## Related work

Archival scholars and practitioners have recognized the importance of modeling workflows to better understand and advance archival practices, especially around born-digital materials. Although the current born-digital archiving practices of the partner institutions were less established and fully formalized than the project team anticipated, the as-is workflow models effectively depicted what was happening on the ground as a snapshot of an ongoing, dynamic process. As Daines (2011) suggests, process mapping techniques are valuable approaches for archives to holistically understand their current practices and, with this overarching perspective, to identify gaps, challenges, and areas for improvement. “They facilitate an understanding of the big picture and help to define the boundaries of a business process” (p 133). However, Daines also notes that process models are built as linear sequences of clearly definable steps with predictable inputs and outputs (p 129). In other words, workflow models are well suited for documenting current practices but not necessarily apt tools for visually depicting possible new curation steps. As we will discuss, these as-is models of current practices are important tools for articulating workflow aspirations, even if these aspirations are difficult to diagram.

As the present research makes clear, linear process models cannot fully capture the dynamic and varied factors that impact the ongoing development of digital curation practices at institutions. The robust OSS tools in digital curation work—both those focused on in the project as well as others used by the partners and across the profession—make many workflows possible, but actualizing these workflows involves far more than stringing together a linear sequence of potential tasks. As Owens (2018) states, “ensuring long-term access to digital information is not a problem for a singular tool to solve. Rather, it is a complex field with a significant set of ethical dimensions. It’s a vocation” (2). Partners’ reflections on the as-is workflow models provided insight into this complex field, revealing both the sociotechnical and organizational factors shaping the horizon of digital curation work they saw as possible.

Rather than thinking about the use of discrete tools and technologies, we approached digital curation practices and workflows as part of broader information technology infrastructures. These infrastructures encompass not only physical structures and technological systems, but also individuals, organizations, and other social groups like teams and networks. Lee et al. (2006) argue that this “human infrastructure is integral to the painstaking process of creating a data sharing infrastructure,” though it is “both dependent and constrained by conventions of practice, existing standards and the organizations that created them” (2006, p 490). Lee et al. describe this human infrastructure in the context of distributed scientific research, but the development of born-digital archiving workflows is similarly shaped at this confluence of social, technological, organizational, and professional factors.

A key lesson from infrastructure studies scholarship is that shared systems, technologies, and standards are often applied in divergent ways at the local level (Star and Ruhleder 1996; Bowker et al. 2009, p 102). Both the repertoires of digital curation practices and the tools supporting this work are dynamic and recursive, open to change over time as a result of the negotiations required to implement these tools and practices—but the character and consequences of these negotiations differ across institutions. This is not a problem to be resolved by designing better software or defining standards with more precision. Rather, it is a generative tension resulting from the many ways that practitioners on the ground flexibly adapt the available tools to meet the immediate needs (and inevitable constraints) of the work at hand. As Bowker and Star state, “no such tool can be defined once and for all. They are always the products of continuing negotiation and change” (2009, p 158). In the case of born-digital archiving, then, the issue is not designing ArchivesSpace—to take an example from the present project—from the outset so that it meets all archivists’ needs at all times but recognizing that particular archivists at specific institutions will necessarily use the same tool in quite different ways.

Workflows also change in response to the ongoing development of standards and best practices for acquiring, processing, and providing access to born-digital collections. In recent years, successful working groups have contributed to the vital discourse of best practices for processing born-digital archival materials (Annand et al. 2018) and providing access to born-digital collections (Arroyo-Ramírez et al. 2020), and these recommendations for archival practice necessarily coevolve with the widely-used tools and systems needed to carry out this work. At the same time, there is growing recognition of factors impacting the execution of born-digital archiving that extend well beyond technology. Preservation needs are often not perceived uniformly across institutions, marked by significant gaps between those in leadership roles and the practitioners responsible for carrying out preservation tasks. As Rieger (2018) notes, institutions often lack a shared preservation mandate across departments and the coordination of tasks required for undertaking preservation work is challenging. Empowering practitioners to propose and/or implement policy is one recommended strategy to address the difficulties arising from the siloed nature of distributed digital preservation work (Blumenthal et al. 2020).

The lessons learned from the OSSArcFlow project can inform the developmental trajectories of digital curation tools and workflow activities by shedding light on the impact of these sociotechnical factors. Community-supported and

collaboratively-developed tools and systems are crucial components of archival infrastructure, but a failure to plan for and fully grapple with the tensions among diverging interests of disparate stakeholders, varying local contexts of use, and the coordinating pieces that need to coalesce in the form of workflows, threatens to undercut these efforts at the outset, as examples from the history of digital libraries and digital humanities projects demonstrate (Hedstrom 1998; Dombrowski 2014). Through better understanding archivists' aspirations for their digital curation workflows, and advancing methods for documenting and sharing these aspirations within professional communities of practice, the ongoing development of digital curation tools and systems can be responsive to the generative tensions intrinsic to large-scale information infrastructure.

## Methods

In the initial plan for the project, staff from each of the 12 partner institutions were to participate in two series of semi-structured interviews, each series including a main interview session lasting approximately 60 min, as well as at least one follow-up interview to discuss additional points of interest or clarification specific to the partner. The first series of interviews centered around the current digital curation practices at the institution, feeding directly into the as-is workflow diagram, which was constructed iteratively with many opportunities for partners to provide feedback. The second series of semi-structured interviews focused on partners' aspirational workflows; succeeding this diagramming exercise, a second round of interviews gave partners an opportunity to reflect on the gaps and challenges identified in the first round of interviews and to imagine how their workflows might continue to develop as they address these difficulties. Dowdy and Raeford (2014) describe applying this method in a library-wide, systematic audit of workflows for managing e-resources at Duke University, noting how the audit team created 'to be' workflows that sought to correct the specific issues identified in the audit.

As discussed in the previous section, the method of visualizing linear process models worked well for documenting and sharing current digital curation practices but could not capture the complexities of partners' aspirations for their workflows. Drucker (2014, p 79) observes that abstraction from direct observation and experience is a feature of maps, and we can think of these workflow models as maps of current practices. Maps were a useful kind of document for many of the project goals, as the workflow models abstracted away many concrete contingencies in how steps were carried out and thus facilitated comparison across different institutions. Although these process maps sparked critical reflection on current workflows and provided a platform for recognizing challenges and envisioning a horizon of possibilities for advancing digital curation practices, they were not able to capture the richness and details of the local context in a way that could be directly integrated into a visual diagram. Some aspirations did fall into the category of specific steps that partners wanted to change or gaps between steps that could be readily plugged, but the partners' reflections also raised issues that spanned across many steps, difficulties with tools that intersected with factors like staff training or lack of IT

support, desires for hypothetical tools or not yet supported functionality for existing tools, and open questions being discussed across the profession.

Once it became clear that modeling partners' aspirational workflows would only capture a slim margin of their actual aspirations and related issues, the project team changed course for the second half of the project. The project team proposed facilitating a series of five breakout group discussions that brought the partners together around issues that cut across the individual aspirational interviews (Chassanoff and Post 2020). The project team suggested a few breakout group topics based on our observations of these shared issues and concerns, with topics including describing born-digital collections, scaling up workflows, and staff training and skill building. From our preliminary analysis of the as-is workflow interviews (Post et al. 2019), we already understood that partners' workflows were not uniform in the steps, tools, and order of processes, and so we anticipated that there would be many questions in these areas for partners to fruitfully discuss. For each session, one or two partners volunteered to lead the discussion, first presenting a case study of their experiences with the particular issue at their own institution, and then fielding questions that opened up the conversation to other partners making comparisons to their experiences at different institutions.

Along with the breakout groups, this mid-project change of course also impacted our plans for research and analysis. Instead of generating visual diagrams, we have conceptually modeled partners' aspirations for their workflows in a taxonomy (see Fig. 1). We developed the taxonomy using an axial coding method (Corbin and Strauss 1990, p 13), organizing codes emerging from the qualitative data gathered by the project team into a hierarchy of categories and sub-categories. The qualitative data sources primarily included the aspirational and as-is workflow interviews, as well as meeting notes and exercises used to check in on partners' progress toward goals discussed in the interviews. Using

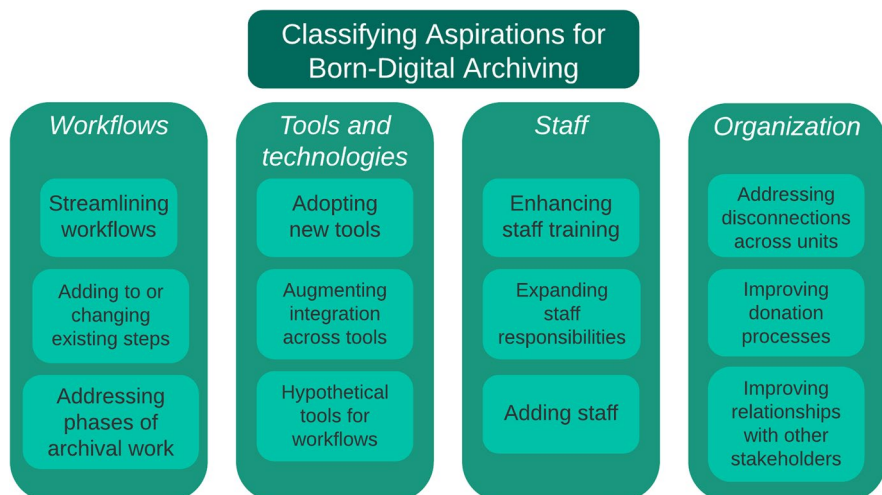


Fig. 1 Taxonomy of aspirations



the qualitative analysis software NVivo, the coding was first performed by one project team member, who developed a set of codes through an initial pass over the data and then iteratively grouped these codes into increasingly more generalized categories through several additional passes. Another project member then tested out this set of categories to conduct their own coding of the data, and using a review feature in NVivo, the two team members compared how they applied the codes to the data and discussed discrepancies in how specific codes and more general categories were applied.

While this was not designed as a grounded theory study, we have followed a constructivist grounded theory approach to this analysis, recognizing the interpretive role played by the researcher (Charmaz 2014, p 13). The separate coding passes by the two team members were not intended as an assessment of inter-coder reliability but rather served to further discussion and deepen interpretation by bringing in multiple perspectives. We have previously presented results from this analysis pertaining to partners' as-is workflows (Post et al. 2019), but this earlier paper did not get into partners' workflow aspirations. For the present paper, the authors have returned to this analysis, specifically focusing on the codes relating to partners' aspirational goals. One of the authors performed an additional round of selective coding, or a process of unifying categories around a single core category (in this case, aspirations for workflows) (Corbin and Strauss 1990, p 14), to assemble the taxonomy described below.

Although the taxonomy itself is, in a narrow sense, the product of the authors, we have approached this research from a community of practice perspective in two important ways. Methodologically, what we have learned has been gained through direct collaboration with professional archivists. Echoing the findings of Clemens et al. (2020), we observe that archivists are able to articulate, and in some cases, actualize how to advance digital curation workflows through interacting, participating in, reflecting upon, and sharing resources within communities of practice. Lave and Wenger (1991, p 99) characterize communities of practice as existing through a dynamic balance between inculcating new members in the established values and practices and integrating new approaches and perspectives as the community is reproduced through this new membership. As archival practitioners continue to respond to the emerging challenges and subsequent changes involved in working with born-digital collections, these are especially crucial reproduction cycles for archival communities of practice. Lave and Wenger (1991) describe how communities of practice generate 'learning curricula' through situated opportunities for the development of new practices and approaches; the OSSArcFlow project provided such a learning curriculum to project partners. Among the results of this present research, we demonstrate how archival communities of practice can serve as forums for archivists to negotiate differences in how they apply community-supported and collaboratively developed tools and attendant best practices for using those tools. Foregrounding these tensions between local contexts and the global domain of professional practice is essential to the long-term sustainability of this shared archival infrastructure.

## Taxonomy of aspirations

Across the aspirational interviews and the breakout group sessions, OSSArcFlow partners shared a wide range of aspirations for their digital curation workflows, from concrete and immediately-actionable goals to more amorphous desires for yet to be developed tools. As discussed above, this diversity (especially in terms of time-scale and scope of the aspirations) inhibited our attempts to fully model the partners ‘to be’ workflows as linear diagrams. Instead, we have applied qualitative coding techniques to group aspirations into major categories, each of which exhibits the range and diversity characteristic of the overall set of aspirational goals. The first major category is **aspirations** for workflows, which encompasses goals for particular workflow steps, broader phases of archival work, and the workflow at large. The other major categories group aspirations according to sociotechnical factors impacting digital curation work identified in our earlier research (Post et al. 2019), including goals related to tools and technologies, staff, and organizational or administrative factors.

While we refer to this as a taxonomy, the discussion below should quickly make clear that there is a **great deal of overlap and interaction** among the various categories. We delve into one pressing example of this interaction at the end of the section, touching on the widespread concern for scaling up workflows to address increasing amounts of incoming and backlogged digital materials. The categories are therefore not mutually exclusive, and nor are they exhaustive, as the categories we propose here are based on a relatively limited sample of project partners from 12 institutions. We put this forward, then, as a provisional and pragmatic taxonomy, a framework that can further—and be furthered by—discussions within archival communities of practice regarding the maturation of digital curation workflows. Practitioners can use this taxonomy as a sounding board for their own aspirations, making sense of difficulties they may be encountering in reference to those experienced by practitioners at disparate institutions; further research and continued discussions can, in turn, add to or alter this taxonomy. This is a working knowledge artifact that we hope will be taken up by archival communities of practice, the importance of which for maturing digital curation workflows we discuss in the following section.

## Goals for workflows

As partners reflected on their as-is workflow documentation and diagrams, they identified many changes they would like to make as they advanced their digital curation practices. Among these goals, several partners suggested specific steps they would like to add to their existing workflows: Emory wanted to assign **unique identifiers to incoming materials**; NYPL and Rice suggested **integrating steps for normalizing certain file types**; KHS and AUC hoped to **add steps for verifying the fixity of digital files**. These aspirations, which could be readily added to linear workflow diagrams, were the kind anticipated by the project team. Discussions about what steps should be included in digital curation workflows were indeed prevalent in both the

aspirational interviews and the breakout group sessions, but these discussions also brought up concerns related to adopting new tools, expanding staff training or roles, and grappling with organizational mandates or confronting conflicting institutional priorities. The aspiration to add or change a workflow step tells only a part of the story more fully divulged in the other taxonomy categories.

Along with these goals for specific steps, partners articulated aspirations for their workflows in general, imagining how workflows could be better managed or how methods for broader phases of archival work could be improved. Eight partners described a goal of streamlining their existing workflows. Both Duke and MIT articulated a desire for a “single pipeline” for materials to move through from accessioning to access, and DCPL wanted a “single entry point” for collection metadata. In our earlier analysis (Post et al. 2019), we suggested that the largely ad hoc nature of partners’ as-is workflows, which might be customized on the fly to handle unique cases or altered to accommodate disparate kinds of materials, was a primary driver of this desire to streamline. In a review of workflows for removable media at the Bentley Historical Library, Eckard and Hagen (2018) echo the sentiments of the partners as they similarly strove to identify a middle ground between “developing custom workflows for each unique collection and trying to shoehorn everything into a single workflow” (2018, p 4). However, this further analysis has drawn out many sociotechnical factors impeding this desired simplification. AUC mentioned redundancies across various digital curation tools and the need to consolidate and clarify what tool is to be used for each workflow activity. KHS described a similar need for coordination but across the various staff and departments responsible for different aspects of digital curation work. As a repository for social science data contributed by researchers, Odum enumerated the many opportunities for errors and omissions in the submission process that contribute to confounding workflows later down the line.

While these complicating factors will never be entirely eliminated from digital curation work—making the “single pipeline” an understandable but unachievable goal—many partners (10) expressed a corollary aspiration of further formalizing existing workflows. If all the complications cannot be eliminated, then formal workflow documentation can direct staff to consistent courses of action when those complications are encountered. A major takeaway from the OSSArcFlow project, though, is that formalization is not a straightforward process of documenting what is currently being done on the ground, but rather an iterative process of reflecting on current practices, responding to gaps and challenges, trying out new approaches, and repeating the cycle. Rice described their goals for revising documentation in this way: tweaking workflows depending on outcomes of experimentation while also documenting their rationale behind these decisions, ensuring that documentation will not just be a static snapshot of how steps were arranged or how tools were used but part of an ongoing process of development. As we discuss below, this reflective and iterative process benefits from being carried out through communities of practice that extend beyond individual institutions.

Along with aspirations for workflows generally, partners also described goals for improving processes for broader phases of archival work, including acquisition, appraisal, accessioning, description, and arrangement. Beyond a single workflow

step, these aspirations involved larger transformations in how digital curation is done at each stage of the archival enterprise. In some cases, these aspirations were articulated as actionable goals, even if they stretched across many discrete activities. For instance, five partners wanted to advance a self-deposit approach to acquiring materials that would be available for individuals or entities making frequent donations to the institution. NYU already had an application in place that would support this, while AUC and DCPL had ideas for potential ways to accomplish this but had yet to take concrete steps toward implementation. In other cases, though, partners' aspirations for phases of archival work involved more complex interactions between the affordances of software environments and existing archival practices. Nine partners identified providing greater access to born-digital collections as a priority but all were in the process of exploring a range of options for facilitating both reading-room and online access. Partners used the aspirational interviews and breakout group sessions to explore possible changes to current descriptive practices in order to enhance user access. For example, Duke led a breakout session describing the development of their current approach to representing born-digital objects in ArchivesSpace. Initially, Duke configured the software for reading-room access only. With the advancement of their online repository, they were able to modify in-house descriptive practices in ArchivesSpace to facilitate direct access to certain prioritized objects through finding aids.

Importantly, making these changes to accommodate born-digital objects does not necessarily require the adoption of an entirely new archival paradigm. As Pearce-Moses and Kaczmarek (2005) point out, it's not the "what" of digital archives that's different but the "how" (2005, p 17). Aspirations for broader phases of archival workflows follow neither linear and predictable developmental pathways nor are they always radically altered to accommodate digital materials. Instead, digital archiving workflows develop through iterative processes in which practitioners reflect on the accumulation of existing practices, professional standards and norms, and new approaches afforded by tools like ArchivesSpace. For Duke, these intersecting practices can be summarized as: 1) existing localized practices for archival description and collections using ArchivesSpace; 2) descriptive archival standards that reflect those adopted by the archival profession writ large; 3) the capacity and ability of practitioners to adopt existing norms and standards for describing born-digital materials in ArchivesSpace. In other words, integrating ArchivesSpace successfully into existing digital curation workflows practices is not as simple as adding a single step to an existing workflow. Instead, digital curation workflows are developed through an ongoing process of negotiation among local and global contexts with rippling effects that affect other aspects of stewardship.

## Goals for tools and technologies

Partners' aspirations for tools and technologies hewed close to goals for workflows, as the use of these tools are not ends in themselves but rather means to achieve digital curation outcomes. However, as partners use a wide range of tools—spanning homegrown scripts, open-source systems, and commercial offerings—foregrounding

goals for tools draws attention to how the concrete details of the implementation of a particular tool within a given context shapes the horizon of possibilities for digital curation work. Ironically, this influence can perhaps best be seen through partners' discussions of hypothetical tools that do not yet exist, clearly illuminating gaps between the kind of work that partners can carry out with the tools already at their disposal and the work they wish they could do but for which they lack the necessary utilities. These hypotheticals ranged from the highly specific, such as NYPL's desire for an open-source email converter tool to be built into the BitCurator Environment, to the broad, like KHS's ideal of a single system that would accommodate universal access to all of their digital collections. In both cases, the desired tool illustrates in relief an aspiration for their digital curation workflows: to preclude the need to go outside BitCurator to normalize email formats; to simplify complications in maintaining multiple access platforms. In some cases, partners' aspirations to adopt new tools or to expand the use of already implemented tools likewise corresponded to some goal for or perceived gap in existing digital curation workflows. MHC mentioned potentially adopting Bagger as part of a remote acquisition approach more sophisticated and efficient than mailing out hard disks or copying files onto flash drives.

Aspirational goals for adopting new tools or expanding the use of existing tools rarely aligned directly with some discrete workflow step. For instance, institutional support for implementing new tools impacts the scope and character of these goals. AUC had committed to implementing Archivematica as part of their participation in OSSArcFlow, but at the midpoint in the project AUC still counted this as an aspirational goal. Their existing hardware was insufficient to run the system, and while their involvement in the grant helped to convince their IT department to invest in a more powerful machine, they were still deep into the protracted process of acquiring this equipment. In addition to these interactions with myriad sociotechnical factors, many of the partners' goals for technologies were complicated by how they encompass multiple workflow steps and disparate tools. Among the most ubiquitous aspirations surfaced in the research, partners sought to augment integration across workflow steps by facilitating handoffs between systems and automating digital curation activities where possible, echoing the above-discussed desire to streamline workflows.

This particular aspiration was anticipated by the project team and was a primary motivation for the OSSArcFlow project, which originated from the recognition that there are many open-source tools for digital curation work but limited means for efficiently stringing these tools together in workflows. Initially, the project scope included the development of scripts to address metadata handoffs between OSS systems. Because many partners were still relatively early in the process of implementing these systems, the project team decided to focus on four institutions (Duke, MIT, Stanford, and NYPL) who already had robust multi-system workflows in place. However, details of how these systems were configured at each institution made the development of shared scripts quite difficult. For example, both MIT and Duke expressed a similar aspiration to automate the manual process of transferring metadata generated from BitCurator reports into various ArchivesSpace description fields (see Table 2).

**Table 2** Comparison of two aspirations

MIT aspiration	Duke aspiration
“I want to update descriptive information and extent in archivesspace based on processing (disk imaging, file extraction, etc.) using bitcurator without having to manually do this.”	“Automate reuse of metadata from acquisitions and analysis activities into descriptive work”

After drilling down into how each institution had customized BitCurator and ArchivesSpace, the project team realized that each institution wanted to extract slightly different information from BitCurator reports and wanted to integrate this metadata into slightly different descriptive fields in ArchiveSpace. Despite a similar goal, a script developed to realize this aspiration at Duke would not function at MIT and vice versa.

Aspirations shared across institutions are distinctly shaped by the details of local context. This dimension of local context has significant implications for ongoing efforts to develop integrations across OSS for digital curation, and for the sharing of digital curation resources more generally. While more automation and fewer manual processes were common motifs, details of local practices and technical implementations will always preclude automation at some level. Shared tools supported by a community of practitioners are essential to fulfilling digital curation functions at cultural heritage organizations (Gengenbach et al. 2016), but the implementation of these tools involves a balance between local autonomy and needs of the archival profession as a whole. Bowker et al. (2009) characterize information infrastructure as distributed across a global and local axis and a technical and social axis, necessarily balancing factors at the ends of both axes. Efforts to automate workflow activities by integrating handoffs across various OSS need to strike such a balance.

### Goals for staff

Compared to other taxonomy categories, partners communicated a more limited range of goals related to staff, namely adding new staff members, expanding digital curation responsibilities to existing staff members, and enhancing staff training and familiarity with digital curation activities and tools. These aspirations are all interrelated—for instance, the inability to hire new staff dedicated to digital curation work necessitates spreading these responsibilities around among existing staff—and are all also deeply affected by broader socioeconomic conditions of the larger organization. Archival institutions are now grappling with the sweeping effects wrought by COVID-19, which has had immediate impacts on staffing in the way of layoffs and hiring freezes, the full extent of which will not be understood for some time.

Several partners (AUC, Duke, MIT, MHC) expressed aspirations to add more staff members, whether for specific workflow areas, such as hiring more processing archivists, or adding staff to specifically focus on digital curation work. Nine partners, including the four just mentioned, described plans for expanding digital curation responsibilities to existing staff. At the time of the project, Rice lacked a dedicated



digital archivist and rotated digital archiving responsibilities among four full-time archivists. Whether bringing in new staff or expanding digital archiving responsibilities to existing staff, partners recognized that digital curation involves a growing and dynamic skill set, and thus requires ongoing staff training and professional development. Seven partners discussed goals to augment staff training and familiarity with digital curation tools and techniques. KHS cited integrating new tools into digital curation workflows, specifically the BitCurator Environment, as a key motivating factor for this training. In many cases, staff have not had previous training in digital curation activities, and so systematic skill-building efforts are fundamental to successfully achieving aspirations for exploring new tools or techniques.

### Goals for organizations

In the preceding categories, we have already made mention of several ways organizational factors shape aspirations pertinent to workflows, tools, and staffing. Workflows follow the contours of organizational structure, and decisions regarding implementing new technologies or changing staff responsibilities often need to pass through formal channels and require institutional buy-in. Partners mentioned many adjacent departments, committees, and agencies across their organizations that had some stake in digital curation. Eight partners discussed aspirations for generally improving relationships with these stakeholders as a means to advance digital curation workflows in their organizations.

Related to this, partners discussed goals specific to how digital curation activities cut across organizational departments and units. Acquisition of materials is one area where workflows cut across organizational units or outside the organization entirely: curators may be responsible for acquiring collections that archivists then process and make accessible, or archivists may work directly with a variety of donors in and outside the institution. In either case, important decisions about how, when, or what materials are collected are made at this juncture that in turn have ramifications further down the workflow. 10 partners mentioned improving communication with curators or donors at this initial stage in the workflow. Odum described the manual process of going back and forth with research data depositors as a sticking point. Duke and Emory both outlined plans for better equipping curators to play an established role in the appraisal and accessioning of born-digital materials that then get handed off to archivists. MHC wanted to increase awareness of the born-digital records program among offices across campus and formalize how university records are collected. As Dowdy and Raeford (2014) report in their assessment of e-resource workflows at Duke University, email and human memory proved to be “the weakest links in our decentralized, fragmented, error-prone system” (2014, p 178). Addressing those points where workflows depend on the involvement of entities outside the archives unit was a focus of partners’ goals related to organizational factors.

## Scaling up

As we have already indicated throughout this section, the categories of the taxonomy are porous, with myriad interactions between aspirations. Many of these interactions would benefit from further research as they represent critical juncture points where digital curation practices and tools are shaped by sociotechnical factors. One such interaction point warrants highlighting here: a general recognition among the partners that current workflows will need to scale up to address the increase in born-digital acquisitions and growing backlogs of unprocessed digital collections. This concern permeates all the above categories as partners anticipate that the need to manage ever-larger digital collections will require more efficient workflows, improved tools, and new staff skills, along with attendant organizational adaptations.

The impact of scaling up can be expressed in many ways, from the staff time involved in digital curation activities to the anticipated financial investment required to care for burgeoning digital collections. Rosenthal et al. (2012) demonstrate that digital preservation requires ongoing investment, the costs of which are substantial and increasing, including recurring expenses for digital storage as well as the labor needed to curate materials over time (2012, pp 513–514). Developing models to more readily prepare for these costs have long been called out as an essential component of preservation planning (Cloonan and Sannett 2002, p 87–90). As Cochrane, von Suchodoletz, and Rechert (2014) discuss for costing emulation, just one particular digital curation activity, many unpredictable factors make approximating costs difficult, challenges that are only amplified when considering holistic costing of digital curation. Uncertainties about the future size of digital acquisitions and how the scale of collections will transform digital curation workflows undermine any reasonable attempt to derive costing models.

Another response to this situation, though, is to critically interrogate the conditions contributing to scaling up and to reassess curation practices that encourage unabated acquisition and promise immediate, unlimited access. Christen and Anderson (2019) call for a slowing down of archival practice to create space necessary to question the complicity of current practices and processes in colonial regimes of power. In particular, this space for interrogation can help us to understand how these anxieties of scaling up workflow practices relate to broader questions of sustainability raised by Tansey (2015) and Pendergrass et al. (2019). In response to worsening climate chaos, these authors urge archivists to evaluate how the practices of cultural heritage stewards contribute to this crisis. The recommendations offered have implications across digital curation workflows not only addressing the concern of ever-growing digital acquisitions, but also challenging the assumption that all collections need to be permanently preserved at the bit level (Pendergrass et al. 2019, p 186). As the present research illuminates, sociotechnical factors and aspects of local context mediate how any generalized best practices are carried out on the ground, and in turn, shape the pressing questions motivating ongoing discussions relevant to maturing workflows. Regardless of how archivists address scaling up (and other open issues), the critical space described by Christen and Anderson will be essential to articulate, assess, and realize aspirations for digital curation workflows. Despite



differences across local contexts, communities of archival practitioners afford this necessary space for reflection.

## Realizing aspirations through communities of practice

The changes made halfway through the project were largely motivated by the notion that partners could better articulate and understand their own workflow aspirations by discussing their goals and challenges with other project partners. In effect, the context of the project generated the three dimensions of a community of practice described by Wenger (1998, p 73): mutual engagement in negotiating meaning, participating in a joint enterprise, and developing a shared repertoire. Through regular project meetings and breakout groups, the partners came together to discuss common issues and challenges, and in doing so, defined a particular kind of joint enterprise with a distinct shared repertoire. While the project partners did not constitute a true community of practice in and of themselves, having been brought together somewhat artificially through their shared involvement in the project, OSSArcFlow provided means to closely consider the role of born-digital workflows in archival communities of practice. Our reflections on the analysis of the partners' aspirations—and how these developed through a shared enunciation of issues and challenges—suggest ways to support existing communities of practice as spaces for archival practitioners to articulate and realize aspirations for born-digital workflows.

The partners all participated in communities of practice outside the bounds of OSSArcFlow, and many mentioned the importance of their participation in these communities for sharing resources, troubleshooting issues with tools, and discussing emergent digital curation activities. OSSArcFlow has highlighted the significance of workflows as artifacts in these existing communities, in particular, and has helped to develop practices for creating and sharing workflows as crucial elements in archival repertoires. Wenger argues that practice is negotiated at the intersection of reification and participation (1998, p 62), as technologies, policies, and other concretized artifacts are interpreted (and reinterpreted) in the process of getting put to use. Digital curation workflows are a perfect example of this juncture: these documents reify how things have been done in the past, communicating a set of steps to follow while also providing a material grounds for reflecting on what could be done differently. In communities of practice—both within particular institutions and more fluidly across the profession—a repertoire for generating, sharing, and advancing workflows can help archivists to recognize the stoppages and gaps in their work and prompt the articulation of aspirations for overcoming those challenges.

Developing and sharing workflows are an important discursive practice. Lave and Wenger (1991, p 109) frame discourse as a kind of practice, emphasizing discourse among practitioners as a significant site of learning. Newcomers to a community learn how practitioners talk and how to talk with other practitioners. For new and veteran archivists alike, workflows are important artifacts in these discourses, and the present project has demonstrated the value of learning how to talk about, through, and with these workflows. During the breakout group sessions, partners centered discussions around each other's workflows to relate challenges and

brainstorm possible solutions. As Orr (1996, p12) attests, “war stories” are common currency swapped among practitioners, though these stories do more than build camaraderie. Stories enter into a bricolage of other tools that help practitioners to formulate and solve problems.

Workflows are themselves statements that enter into practitioners’ discourses, a representation of how archival work is carried out as well as a point to which someone else can respond. Practices for sharing and commenting on workflows are just as important as practices for creating this documentation. Sharing workflow documents within communities of practice offers opportunities to compare and contrast with other practitioners’ experiences. In addition to finding common ground, this exchange can also illuminate differences in context and varying sociotechnical factors that may influence archival work at one institution but not another. The enunciation of both similarities and differences in digital curation work has implications for the development of open-source tools in particular, as development that stands to enhance the use of a tool at one institution may in fact inhibit use in another context. The significance of workflow documentation in archival discourse is increasingly being recognized, evident in the many resources accumulating under the banner of **Community-Owned Workflows** (COW) initiative to provide one example ([https://coptr.digipres.org/Workflow:Community\\_Owned\\_Workflows](https://coptr.digipres.org/Workflow:Community_Owned_Workflows)). Approaching the creation and use of workflows as crucial skills in archivists’ repertoires suggests further programs of research and development, such as how to train new professionals to effectively ‘talk with’ workflows or how discussions around workflows can feed back into the development of open-source digital curation tools.

## Conclusion

Workflows are generative artifacts, integral to the development and maintenance of digital curation practices—not just rote prescriptions for practice. Sociotechnical factors and aspects particular to local context impact not only how digital curation work gets done, but also shape how practitioners envision the advancement of this work. Putting digital curation workflows into practice, therefore, is not just a matter of carrying out precisely defined steps and processes; challenges that arise in carrying out this work cannot be resolved solely through refining these steps and processes at evermore precise levels. Instead, our findings suggest that the archival profession should encourage and reward research embedded within communities of practice, in particular to investigate how the implementations of digital curation tools and the arrangements of digital curation work configure the lived realities of archival practitioners. Workflow documents are crucial artifacts supporting archivists as they make sense of and negotiate these configurations—literally, these artifacts are at the crux of current and future practice.

The development of archival practice does not reside within one institution; rather, archivists negotiate the meaning of practice across the professional community. Digital curation workflows, in particular, can enter into this professional discourse as statements about the nature of current practices and as points around which to structure reflections and discussions about the continued development of

tools and techniques. As digital curation is fundamentally sociotechnical work, discussions around workflows in broader communities of practice enable practitioners to make comparisons across their various institutional contexts, to make sense of many factors that impact their work locally by understanding the factors that impact others' work elsewhere. These discussions also make clear that digital curation is far from a one-size-fits-all undertaking. Digital curation systems and techniques from one institution cannot simply be uprooted and implemented at another institution. This does not obviate the need for communities of practice but intensifies and amplifies their role.

**Acknowledgements** The authors would like to acknowledge the contributions of the many other project team members along with the sustained engagement and generous participation of the project partners that made this research possible. Additionally, the authors would like to thank Kassidy Hof-Mahoney for her work to copyedit, format, and provide general feedback on an earlier draft of the article.

**Funding** OSSArcFlow is funded by a grant from the Institute for Museum and Library Services (IMLS grant LG-71-17-0016-17).

**Declaration**

**Conflicts of interest** The authors declare that they do not have any conflict of interest.

## References

- Anderson EK (2014) Workflow Analysis Library Technology Rep 50(3):23–29
- Annard S, DeBauche S, Faulder E, Gengenbach M, Irwin K, Musson J, Peltzman S, Tasker K, Jackson LU, Waugh D (2018) Digital processing framework. <https://hdl.handle.net/1813/57659>. Accessed 21 May 2021
- Arroyo-Ramírez, E, Bolding, K, Butler, D, Coburn, A, Dietz, B, Farrell, J, Helms, A, Henke, K, Macquarie, C, Peltzman, S, Watson, CT, Taylor, A, Venlet, J, Walker, P (2020) Levels of born-digital access. Digital Library Federation. Accessed 21 May 2021 from <https://osf.io/r5f78/>
- Blumenthal K, Griesinger P, Kim JY, Peltzman S, Steeves V (2020) What's wrong with digital stewardship: evaluating the organization of digital preservation programs from practitioners' perspectives. *J of Contemporary Archival Stud* 7(13). Accessed 21 May 2021 from <https://elischolar.library.yale.edu/jcas/vol7/iss1/13>
- Bowker G, Star SL (1999) *Sorting things out: classification and its consequences*. MIT Press, Cambridge
- Bowker GC, Baker K, Millerand K, Ribes D (2009) Toward information infrastructure studies: ways of knowing in a networked environment. In: Kjastrup L, Allen M (eds) Hunsinger J. *International Handbook of Internet Research* Springer, Dordrecht, pp 97–117
- Charmaz K (2014) *Constructing grounded theory*, 2nd edn. Sage, Thousand Oaks
- Chassanoff, A & Post, C (2020) OSSArcFlow Guide to documenting born-digital archival workflows. K Skinner, J Farrell, B Locke, C Perry, K Smith, & H Wang (eds). Educopia Institute. Accessed 20 May 2021 from <https://educopia.org/ossarcflow-guide/>
- Christen K, Anderson J (2019) Toward slow archives. *Arch Sci* 19(2):87–116
- Clemens A, Hagenmaier W, Meyerson J, Appel R (2020) Participatory archival research and development: the born-digital access initiative. *Proven J Soc Ga Arch* 36(1):4–24
- Cloonan M, Sanett S (2002) Preservation strategies for electronic records: where we are now—obliquity and squint? *Am Arch* 65(1):70–106
- Cochrane E, von Suchodoletz D, Rechert K (2014) Developing costing models for emulation based access in scientific libraries. In: IPRES 2014: Proceedings of the 11th International Conference on Digital Curation, Melbourne, Australia

- Corbin J, Strauss A (1990) Grounded theory research: procedures, canons, and evaluative criteria. *Qual Sociol* 13(1):3–21
- Daines GJ III (2011) Re-engineering archives: Business Process Management (BPM) and the quest for archival efficiency. *Am Arch* 74(1):123–157
- Dombrowski Q (2014) What ever happened to Project Bamboo? *Lit Linguist Computing* 29(3):326–339
- Dowdy B, Raeford R (2014) Electronic resources workflow: design, analysis and technologies for an overdue solution. *Ser Rev* 40(3):175–187
- Drucker J (2014) *Graphesis: visual forms of knowledge production*. Harvard University Press, Cambridge
- Eckard M, Hagen A (2018) Revamping the ‘difficult (potentially)’ but ‘mostly good’ and ‘pretty smooth’ removable media workflow at the Bentley Historical Library. In: *IPRES 2018: Proceedings of the 15th International Conference on Digital Preservation*, Boston, MA. Accessed 21 May 2021 from <https://doi.org/10.17605/OSF.IO/UREY>
- Gengenbach M, Peltzman S, Meister S, Graham B, Waugh D, Moran J, Seifert J, Dowding H, Carleton J (2016) OSS4EVA: Using open-source tools to fulfill digital preservation requirements. *Code4Lib J* 34 Accessed 21 May 2021 from <https://journal.code4lib.org/articles/11940>
- Hedstrom M (1998) Digital preservation: a time bomb for digital libraries. *Comput Humanit* 31(189–202):1998
- Lave J, Wenger E (1991) *Situated learning: legitimate peripheral participation*. Cambridge University Press, Cambridge
- Lee CP, Dourish P, Mark G (2006) The human infrastructure of cyberinfrastructure. In: *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*. ACM, Banff pp 483–492
- Orr J (1996) *Talking about machines: an Ethnography of a modern job*. Cornell University Press, Ithaca
- Owens T (2018) *The theory and craft of digital preservation*. Johns Hopkins University Press, Baltimore
- Pearce-Moses R, Kaczmarek J (2005) An Arizona model for preservation and access of Web documents. *DTTP: Doc People* 33(1):17–24
- Pendergrass KL, Sampson W, Walsh T, Algana L (2019) Toward environmentally sustainable digital preservation. *Am Arch* 82(1):165–206
- Post C, Chassanoff A, Lee CA, Rabkin A, Zhang Y, Skinner K, Meister S (2019) Digital curation at work: modeling workflows for digital archival materials. In: *Proceedings of the 19th ACM/IEEE Joint Conference on Digital Libraries*. IEEE Computer Society Press, Urbana-Champaign, pp 39–48
- Rieger OY (2018) The state of digital preservation in 2018: a snapshot of challenges and gaps. Accessed 21 May 2021 from <https://doi.org/10.18665/sr.310626>
- Rosenthal D, Miller E, Adams I, Storer M, Zadok E (2012) The economics of long-term digital storage. In: *Memory of the world in the digital age*, Vancouver, British Columbia, Canada pp 513–528
- Star SL, Ruhleder K (1996) Steps toward an ecology of infrastructure: design and access for large information spaces. *Inf Syst Res* 7(1):111–134
- Tansey E (2015) Archival adaptation to climate change. *Sustain: Sci Pract Policy*. 11(2):45–56
- Wenger E (1998) *Communities of practice: learning, meaning, and identity*. Cambridge University Press, Cambridge

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Colin Post** is an Assistant Professor of Library and Information Science at the University of North Carolina at Greensboro. He is interested in new approaches to stewarding digital cultural materials both in and outside of cultural heritage institutions. His dissertation investigated how artists engaged with digital technologies are curating interactive immersive exhibitions in online galleries.

**Alexandra Chassanoff** is an Assistant Professor at the School of Library and Information Sciences at North Carolina Central University where she teaches and conducts research on born-digital cultural heritage. From 2017 to 2018, she was the OSSArcFlow Project Manager. Previously, she worked with the BitCurator and BitCurator Access projects while pursuing her doctorate in Information Science at UNC-Chapel Hill. She co-authored (with Colin Post and Katherine Skinner) the *Guide to Documenting Born-Digital Archival Workflows*.