OSSArcFlow

Guide to Documenting Born-Digital Archival Workflows

AUTHORS

Alexandra Chassanoff and Colin Post

EDITORS

Katherine Skinner (Lead Editor), Jessica Farrell, Brandon Locke, Caitlin Perry (Copyeditor), Kari Smith, Hannah Wang

CONTRIBUTORS

Christopher A. Lee, Sam Meister, Jessica Meyerson, Andrew Rabkin, Yinglong Zhang

DESIGNER

Hannah Ballard

EDUCOPIA INSTITUTE

Community Cultivators

TABLE OF CONTENTS

INTRODUCTION

By Katherine Skinner

The purpose of the *Guide to Documenting Born-Digital Archival Workflows* is to encourage and assist collecting institutions of all shapes, sizes, and types to begin documenting their born-digital workflows.

In our experience, both through the OSSArcFlow project (Educopia, 2017-2020) and in consultation with many collecting institutions, the vast majority of today's born-digital archiving activity is not well documented. Most collecting institutions believe that their born-digital archiving workflows are still too ad hoc or nascent to deserve formal documentation, and the lack of formal documentation keeps collecting institutions from being able to see, share, compare, and build upon their collective successes, failures, gaps, challenges, and opportunities.

In the **Guide**, we aim to make the daunting task of selecting, implementing, and refining born-digital archiving workflows more achievable. To do so, we first document and describe the steps that are commonly included in digital curation workflows, from acquisition to access and preservation. We then provide a process map and detailed guidance to help you produce your own born-digital workflows documentation in both visual and descriptive formats. Finally, we provide a set of use cases for this visual and descriptive documentation, illustrating with case studies and examples how you can use it to help your institution improve its born-digital archiving practices over time.

The **Guide** includes four main sections:
1. **Introduction** - provides a brief background of the project, the research questions that have driven our inquiry, and how to use this *Guide* in your own work as an archivist and curator of born-digital collections.
2. **Common Steps in OSS Born-Digital Archival Workflows** - provides brief descriptions of each of the main steps in born-digital archiving (13 in total) and what tools are commonly used to accomplish each of these steps today.

3. **Documenting Born-Digital Workflows** - provides detailed guidance to help you use the OSSArcFlow survey, interview questions, and visualization model to document and depict your own workflow.
4. **Using Workflows -** provides guidance and examples of how an institution can use its existing workflows to identify growth/maturity goals, to advocate for resources, and to transform roles/relationships as needed to improve its born-digital archiving practices.

We are grateful to the Institute of Museum and Library Services for its generous support of this project work, and to our project partners for willingly and excitedly participating in every step of this work with us.

Background

Achieving open-source software (OSS) combinations and integrations at scale and in diverse institutional environments is a critical issue in libraries and archives. It has also been a core goal of the Open-Source Software Archival Workflows (OSSArcFlow) project led by Educopia Institute and the University of North Carolina at Chapel Hill - School of Information and Library Science, and generously funded by the Institute of Museum and Library Services (2017-2020). Our research has sought to understand and document some of the common decision points, challenges, and barriers in OSS tool and system integration through studying and supporting the iterative workflow development undertaken by collecting institutions today, including libraries, archives, and museums.

Our OSSArcFlow research team, comprising leaders from each of three leading OSS technologies, has worked with 12 partner institutions to research, devise, test, and document various strategies for implementing digital archival workflows within institutions of multiple sizes and types, including public libraries (DC Public Library and New York Public Library), research institutions (Stanford University, Duke University, Emory University, Rice University, and Massachusetts Institute of Technology - MIT), a historical society (Kansas Historical Society), a college consortium (Atlanta University Center, Robert W. Woodruff Library), a research center (Odum Institute for Research in Social Science) and a liberal arts college (Mount Holyoke College).

Each project partner committed to integrating a set of common OSS technologies —BitCurator, ArchivesSpace, and Archivematica—during the project period. Each partner also worked with a range of other tools and environments, and each institution was grounded by its own specific aims and abilities. Involving this intentionally diverse range of library and archives partners has helped our team to see and document myriad workflow manifestations and challenges, and to understand how these were borne out of the unique combination of technical and social factors in each institution—including local levels of staffing and expertise, budgets/resources, institutional policies and guidelines, departmental relationships, stability of key roles (e.g., administrators and archivists), and the library/archives' positioning relative to IT. It has also helped us to fine-tune ways of describing and visualizing workflows so that they can be shared and compared.

The project has resulted in documentation of the integration pathways - including the complex array of decisions made by humans along the way - so that the experiences of these 12 institutions now may benefit others.

OSSArcFlow partners at in-person meeting in Chapel Hill, North Carolina (December 2017)

PARTNERS

Atlanta University Center
D.C. Public Library
Duke University
Emory University
Kansas Historical Society
M.IT. Libraries
Mount Holyoke College
New York Public Library
Rice University
Stanford University
New York University
Odum Institute

Born-digital archiving implementation challenges

Since the late 20th century, records and other knowledge objects and outputs across science, art, social sciences, and the humanities have been created predominantly in digital form and stored on a mix of hard drives, floppy disks, optical disks, tapes, and other media containers. Archives, libraries, museums, and other collecting institutions increasingly serve as the stewards of these born-digital materials and must implement digital curation workflows to support their acquisition and care.

Modular OSS tools supporting the curation of born-digital content have matured greatly in recent years, and many OSS applications now have solid user communities, stable code bases, and documentation. However, collecting institutions frequently experience difficulties when attempting to synchronize these OSS tools to enable efficient, effective, and scalable curation workflows.

STEPS

1. Gather information before acquisition
2. Transfer materials to institution
3. Create disk image
4. Run virus checks
5. File identification & format characterization
6. Check file integrity & ensure fixity
7. Create accession record
8. Analyze & identify sensitive content
9. Analyze forensic/technical metadata
10. Create/extract digital object metdata
11. Assemble AIP
12. Assemble DIP
13. Transfer AIP to preservation environment

Thus far, the array of functions necessary to curate myriad content types has not been available in any single or "end-to-end" solution. Instead, most libraries and archives engaged in digital collections management must adopt and integrate separate systems for different functions or "steps" in the born-digital curation process. Sometimes software bundles a few of these steps or functions into an automated process, but no workflow, or chain of steps, is entirely automated, and more often than not, many different systems are in use for a single workflow.

Each time a distinct system or tool is employed, it generates particular forms of data and metadata, and these do not always (or often) match up or interoperate gracefully.

Institutions report that they have to manage significant gaps and overlaps between different tools and environments. ***Gaps*** between tools and repository requirements can make it difficult to push content through a workflow. For example, the output from the first tool/function in a curation workflow may have to be transformed before it is compatible with the next tool. This means that instead of spending time studying the content of the collections in order to produce the best description and access pathways possible, a large portion of time is spent massaging data and metadata so that it can interface and interoperate with different systems.

GAPS
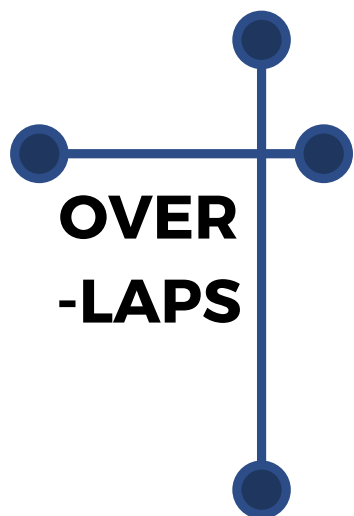
OVER -LAPS

Overlaps between tools also challenge curators to make decisions about when and where to complete a particular function or hand off the content to another staff member to complete the next step. An institution may deploy two different OSS tools that contain some of the same modular scripts and functionality, such as printing a directory or checking for viruses. In such cases, the institution must decide when and where in the workflow it makes most sense to enact a specific function. Choosing one tool over another for a particular step might impact its workflow choices and timing for this particular function.

Many institutions are stymied in their born-digital collecting due to these gaps and overlaps. Solving them requires time, not just from one or two archivists, but from a wide array of archivists, metadata specialists, software developers, and systems administrators. The successful coordination of the systems and tools requires coordination of stakeholders across departments and units. Even the partner institutions in the OSSArcFlow project entered the project work with a low level of faith that their born-digital archiving work could be mapped in its current, half-built state.

Over the course of the project, we demonstrated as a team that even nascent, underdefined, and emerging born-digital archival workflows are well worth documenting, as the very process of documenting them—including the gaps and frozen-in-midstream activities—helps them to take shape.

Documenting Workflows

As we have demonstrated in the OSSArcFlow project, collecting institutions typically have highly localized processes and technical infrastructures supporting digital content management, making the implementation of a "one size fits most" workflow a challenging and impractical endeavor. Institutions instead have craved independence to select from a smorgasbord of OSS tool options that best suit their local needs. Interoperability between most OSS tools in born-digital curation is still an unrealized dream, and collecting institutions encounter barriers and frustrations as they try to stitch together data flows across applications. To understand and overcome these barriers, an institution needs to first document its workflows and then analyze the gaps and overlaps that occur between the different tools and environments that it uses to curate born-digital objects and collections.

Once an institution completes the documentation and analysis of its current ("as-is") practices, it can develop a plan for improving its own workflow and processes over time. The workflow documentation quickly becomes useful to the institution in other ways as well. The institution can share and compare its workflow documentation with that of its peers in order to further refine and optimize its work. It can also use these comparisons to help it to demonstrate to administrators, tool developers, and funders how different investments might be used to streamline and improve dramatically the institution's curatorial work. The workflow documentation provides a baseline and historical record of the evolution of the institution's work and processes as well.

Documented workflows can also have a significant impact at the field level. For institutions that have not yet implemented born-digital archiving workflows, such documentation provides concrete examples to follow, which in turn simplifies the workflow-development process.

For developers and OSS communities, these documented workflows show precisely how different institutions are using their tools, and this can help to surface where tool improvements and new features are most needed. Workflow documentation also has the potential to provide real-time information about desired tool integrations, incentivizing collaborations between tools to improve interoperability over time. Finally, as historical artifacts, these workflows can help us understand how born-digital archiving morphs and changes over time in terms of roles, responsibilities, processes, tools, and other factors.

We produced detailed documentation of 12 collecting institutions' workflows with our formal project partners, and then we completed an additional 11 workflows in collaboration with the BitCurator Consortium in 2019, in order to both test and expand our methodology and framework. In each case, institutions cited this work as transformative for the ways they can now see, show, and envision both their current workflows (many of which became much more formalized as a direct result of the workflow documentation process) and the ways that they can improve those workflows in the future.

We have produced this **Guide** as a synthesis of our framework and a mechanism to enable more institutions to produce and share their workflow documentation. We hope it will catalyze efforts across the library and archives fields to study and address common challenges faced across the common steps of born-digital archiving. As we see more examples of what works and does not work, and as institutions identify common gaps and opportunities across documented workflows, our field will be better equipped to build the scripts and binding and standards necessary to foster efficient and effective digital curation programs that ensure ongoing access to our increasingly born-digital legacy at scale.
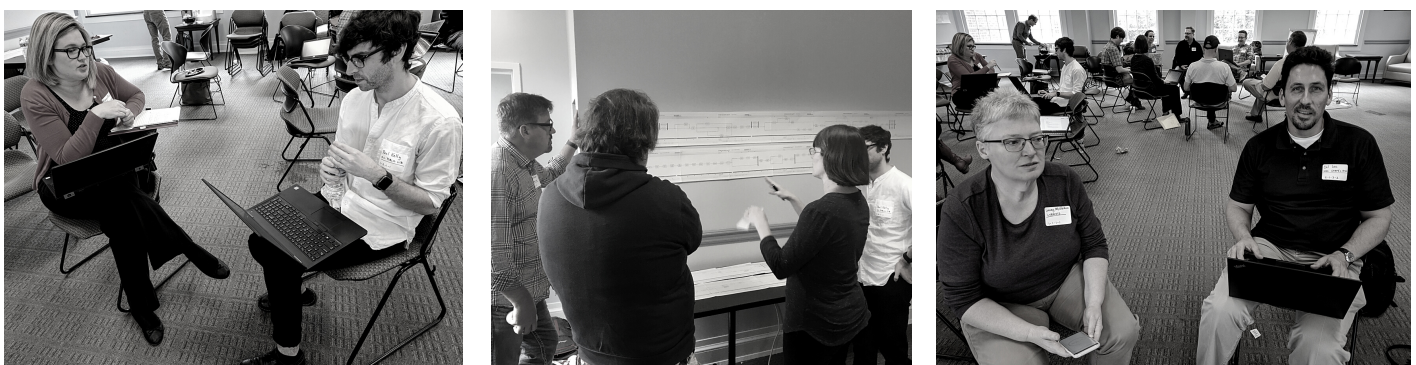
OSSArcFlow Partners at in-person meeting in Chapel Hill, North Carolina (December 2017). Pictured (L to R) Megan Rohleder, Paul Kelly, Michael Olson, Don Mennerich, Kari Smith, Paul Kelly, Laney McGlohon, Christopher Lee

COMMON STEPS IN OSS BORN-DIGITAL ARCHIVAL WORKFLOWS

By Alexandra Chassanoff

This section identifies and describes 13 common workflow steps used in born-digital archiving, as identified in the OSSArcFlow project (Educopia Institute, 2017-2020). Each step is described in terms of specific activities and tools and methods that might be deployed.

These steps should not be considered to be prescriptive or linear—as evidenced by our project partners, these steps are often assembled in different ways for different collections and units even within one institution. Each step documented below includes an overview describing the activity, a case study, implementation recommendations, and a few examples of open-source software tools that institutions might use to accomplish the task.

We highly recommend using free, open-source software solutions that are well supported by a community of practice wherever possible. Please do note that the tools we suggest herein are not the only software options available for these tasks. We recommend that you consult with your IT department or staff to help you identify solutions that can work well for your institution.

Workflow Comparison at OSSArcFlow in-person meeting (December 2017)

1. Gather information before acquisition

As the institution and donor work together to develop a donor agreement or to arrange transfer of a new accession to an existing collection, much information about the material and the nature of its creation can be gleaned that will inform the collection's or accession's lifecycle.

During this step, the staff member collects information from the donor about what known types of material are present in the collection, deposit, or new accession; what the donor perceives as priority material; and —if the creator is living or is the donor—how they used technology in their day-to-day records creation processes. If the institution creates transfer, preservation, processing, or access plans for a collection, these plans are usually developed during this step.

Implementation recommendations
- Conduct pre-appraisal of born-digital materials (where possible) to aid in decision-making about next steps for acquisition and transfer, as well as to inform choices about preservation and access packaging.
  - Consider what packaging formats are compatible with the tools you use in your current or ideal processing workflow.
  - Consider storage needs and current resources, both during initial transfer and for ongoing maintenance and storage. Try to address these prior to transfer.
- Work closely with the donor during the pre-acquisition process to produce documentation that is as complete as possible about media devices or files, the transfer process, and the donor's intentions regarding access to the collection.
  - Seek permission from the donor to collect materials that are being captured through web crawling or fetched from networked filesystems.
  - Conduct recorded interviews with the donor, discussing access restrictions or concerns for materials.
  - Ask the donor for clarity on how personally sensitive information should be handled, and what near-term restrictions are needed.

2. Transfer materials to institution

Materials may be transferred either over a network or by transfer of physical media. Arrangement for transfer of materials usually includes documenting the disposition of records before and after the transfer, and much of the information required to plan the transfer is gathered in the pre-acquisition step.

CASE STUDY: EMORY UNIVERSITY

Pictured (L to R): Nick Krabbenhoeft and Dorothy Waugh

"The Digital Archivist conducts or is involved in a Donor interview in order to gather information about the nature and extent of the born-digital materials. Information gathered during the interview informs the method of transfer for the materials. Some of the options for transfer include on-site forensic disk imaging, transfer of hardware storage media and computational devices to the archives for forensic imaging and analysis, and/or transfer of donor selected files." (2018)

Implementation recommendations
- Use packaging and transfer tools that identify file characteristics, create checksums, and run virus checks.
- Transfer content to non-networked locations for initial triage and assessment.
  - For removable media, use write blockers in transfer procedures to safeguard content from alteration.
- Where possible, use tools that create data that is useful for accession records.
- Clearly document your local best practices for producing accession records that describe born-digital material at your institution.

OSS Tools/Resources:

- Exactly: Developed by AVP, Exactly is an open-source transfer tool that enables secure sending and receipt of digital objects and metadata. The tool employs BagIt, works for SFTP/FTP transfer and standard network transfers, and integrates into desktop-based file sharing applications (e.g., Google Drive).
- Data Accessioner: Initially developed at Duke University to extract content off disks and onto server storage, Data Accessioner creates and validates checksums, gathers metadata, and produces XML.
- The BagIt File Packaging Format is a specification for the storage and transfer of digital content initially developed by the Library of Congress and California Digital Library.

3. Create disk image

Institutions acquiring whole computers or removable storage media (like hard drives or floppy disks) often create disk images as part of their digital curation workflows. Disk images provide "sector by sector copies of the data from physical storage media" (Woods, Lee, and Garfinkel 2011, 57), including potentially "hidden" files and contents of unallocated sectors (partially or fully recoverable content of deleted files). Institutions can create either raw or forensically packaged disk images (usually Expert Witness Format, also known as E01), depending on goals, storage availability, and local needs. Making an exact copy of accessioned content helps to protect the integrity of the original source, which can be required when the files have evidential value. Institutions can then work with and perform specific curation actions on the copy in a staging location, where possible.

CASE STUDY: NEW YORK UNIVERSITY

"For heterogeneous born-digital collections, NYU's digital archivist creates a forensic disk image. For organizational, structured transfers and donor-appraised curated files, NYU creates a logical copy [of files and folders]." (2018)

Pictured (L to R): [Matthew] Farrell and Don Mennerich

Implementation recommendations:

- Disk imaging is the best way to capture environments that you expect to emulate in the future.
- To provide additional context about acquisition imaging, metadata can be added before the disk image is created.
- Disk images are considered the gold standard choice for ensuring and preserving bit-level integrity during the accessioning process. In deciding whether to create disk images or simply extract and copy/transfer files, institutions should evaluate the cost of creating and maintaining disk images as preservation objects (including staffing resources, ongoing maintenance activities, and environmental footprint).
  - Consider what you will do with disk images after processing them. While disk imaging is often a low-barrier way to capture and stabilize born-digital material, it does capture deleted files and other unappraised material. As a result, it can pose risks to the institution or the donor to manage these files over time.
  - Disk images also have a significant environmental footprint and should not be retained unless the need has been identified. Be sure to assess these risks and needs for each collection you choose to image. You can even develop a disk retention policy that attempts to address some of these issues long term.

OSS Tools/Resources

- Guymager: Free disk imaging program that runs on Linux and is bundled into the BitCurator environment. Guymager generates disk images (raw or forensically packaged) from removable media (e.g., flash drives, hard drives, floppy disks).
- Forensics Wiki Imaging Tools: Resource list of imaging tools (some proprietary).
- Archivists' Guide to Kryoflux: A group of archivists created this guide to provide clear documentation and instructions for using the Kryoflux, a floppy disk controller that creates specialized disk images based on the magnetic flux transitions from a drive.
- Working with Legacy Born-Digital Materials: 2013 Code4Lib presentation that discusses disk image format strategies.

4. Run virus check(s)

Depending on the nature of the accession, your storage infrastructure, and the complexity of your workflow, virus checking may happen many times throughout a born-digital object's lifecycle.

An institution often maintains anti-virus software with regular updates in the staging area. All newly acquired contents are scanned for viruses and any infected items are moved to a designated "quarantine." Quarantined items may be discarded, held in quarantine indefinitely, or they may be treated, depending on the importance of the items to the collecting institution. Information about the virus scan is often stored in the digital object's metadata, and a key feature your OSS virus checking software should include is the ability to export your virus checking results.

Implementation recommendations

- For removable media, use a write blocker, also sometimes called a forensic disk controller, in order to have read-only access and avoid damaging the drive's contents.
- Try to include virus checking in your workflow before content is transferred from networked or distributed filesystems. Some distributed filesystems automatically run a virus check before files are downloaded from it, others do not.
- Consider if you need to save the results of the virus scans along with your accession documentation. These results are useful for confirming that no viruses have appeared during subsequent steps in the workflow, especially when unzipping containers such as disk images, zip files, and tar files.
- Scan content on a non-networked location during initial triage and assessment to help safeguard against viruses.
- You may choose to quarantine infected files only or the entire disk image. Consider your quarantine workflow in each of these scenarios.

- Once a file or disk image is in quarantine, it often requires further review from additional staff. Sometimes the curatorial decision-makers do not have the technical knowledge to engage with the format—the disk image or quarantined file(s). There are many handoffs during this process that can create bottlenecks. Consider what you will do next after a file or disk image is quarantined, and what will be required of the collaborators in that workflow to finish processing infected content.

OSS Tools/Resources
- <u>Clam AV</u>: Clam AV is a virus checker that can be run independently or as part of either <u>Archivematica</u> or the <u>BitCurator environment</u>. The results of Clam AV checks can be exported as metadata.

5. File identification & format characterization

In this step, an institution uses tools to identify files and characterize their formats. This can be done by extracting information that is included in the files' structure and metadata. Digital files possess many characteristics that are of potential interest to archives, library, and museum professionals and researchers, such as file formats, file sizes, and file creation information. This information can impact digital curation decision-making related to accessioning, description, batch migration, preservation, emulation, and long-term access.

CASE STUDY: NEW YORK UNIVERSITY

Pictured (L to R): Andrew Rabkin, Shaun Trujillo, Paul Kelly, Kari Smith, Don Mennerich

"[We] work with curators in the field, working with tools to get an understanding about the materials that donors are intending to transfer with us. [We're] thinking about how the transfer will happen based on the content. Working with the tools to do file type identification on file directories, [we can] get a sense of what's there." (2018)

Implementation recommendations:

- Characterize file formats early in the workflow to understand whether files/content require specific preservation actions.
- Consider where else in the workflow this information is important, how this information will be captured over the course of processing, and how it can be used to trigger next steps in the file's lifecycle.
- Consider how file and format analysis can be used to inform institutional priorities. Born-digital collections often include such a diversity of formats that preservation beyond the bit-level is impossible for 100% of the collection. If you collect this data in a way that can be aggregated, you can use it to help prioritize access and preservation for commonly collected formats and characteristics or to prioritize formats that have particularly critical information to preserve.
- Consult guides such as the Library of Congress "Recommended Formats Statement" to inform decisions about recommended formats for preservation and access.

OSS Tools/Resources

- Siegfried is a signature-based file format identification tool.
- DROID performs batch identification of file formats.
- ExifTool Batch Processor: This script runs ExifTool to read and write metadata information at the directory level and creates both a CSV and XML report of ExifTool analyses.
- FITS is a tool for identifying, validating, and extracting technical metadata from files.

OSSARCFLOW PARTNER REFLECTIONS

"Participation in the OSSArcFlow project has helped us start the long process of capturing and improving our workflows in a more systematic way. It has inspired us to improve and clarify what we do and put that in a format that is readable and understandable." (2018)

-Josh Hogan (pictured right; also pictured Megan Rohleder)

6. Check file integrity & ensure fixity

Fixity measures can be used to verify that there have been no undocumented changes to digital objects.

A commonly used example of a fixity measure is a cryptographic hash (also called a checksum). In this step, an institution uses a tool that calculates a numerical value (checksum) for each digital file, and then outputs and/or stores that value. In the future, the institution can use the same tool, or another tool that uses the same algorithm, to calculate a numerical value for each digital file, and then compare the new value to the old. If the values do not match, then further action may be required to investigate what caused the change, restore the file, document the change, or other next steps that vary based on local institutional practice.

Checking file integrity using the checksum may happen as an institution migrates the digital file from one storage location to another, or as it replicates a file (e.g., for preservation storage), or on a fixed schedule to identify file degradation or compromises. It may also be helpful to provide researchers with the checksum of a file along with the file when it is available for download, or to include the checksum in the descriptive metadata, in order to demonstrate that there was no corruption in file transfer.

CASE STUDY: DUKE UNIVERSITY

"Fixity is monitored by a homegrown system called File Tracker, and replicated copies are created in house, with copies remaining onsite online, copies offline and off site (but local to NC) and another copy offline that we intend to send to a geographically separate site but have not identified the site yet." (2018)

Implementation recommendations
- There are a few different algorithms for creating cryptographic hashes, and these can be distinguished by their names, for example MD5 or SHA-512. Some institutions go through the process of researching different algorithms to inform their decision regarding which algorithm to use.

- Similar to decisions like tool selection or preservation formats, the most important thing is that you use the same algorithm consistently and clearly document your decision to use that algorithm.
- To validate file integrity, check fixity before and after the transfer of content at different stages in the born-digital processing workflow (e.g., acquisition, staging, ingest, permanent storage).
- If a file's checksum is found to have changed during a data integrity check, further decisions around documentation and recovery options should be made. These will depend on local policies and procedures.
- While some tools store checksum values, many export them as reports that are run at the time the tool is used. If you implement a tool that exports but does not store checksums, consider what additional micro-workflows you may need to add to this step in order to use that report in the future.
- Some attributes, such as metadata like filename or date created, can be changed without the checksum changing, but the checksum verifies that the contents of the file (the bits that make it up) have not changed. Rely on data gathered during other parts of the process, such as pre-acquisition or creating the accession record, to inform the true creation dates for files.

OSS Tools/Resources

- Fixity is a tool for monitoring file integrity, developed by AVP. It provides automated monitoring and reporting of stored files' checksums. It also provides reports on file attendance, which notifies the institution if files have been added, removed, or have had filename changes.
- What is Fixity and When Should I be Checking It? This resource from the National Digital Stewardship Alliance provides an overview of fixity.

OSSARCFLOW PARTNER REFLECTIONS

"The ability to reference the work of collaborating institutions has informed decision making for updating born-digital workflows here at Mount Holyoke College. The community of OSSArcFlow partners has been a great resource, both through direct communication and the monthly group meetings conducted by Educopia." (2018)

-Shaun Trujillo (pictured left; also pictured Kari Smith)

7. Create accession record

Institutions create accession records by logging key information (metadata) about physical media or digital objects in a database, registry tool, or content management system.

This step is crucial for preserving information about an objects' acquisition and provenance and maintaining it over time. Accession records typically capture baseline descriptive and contextual information about the materials, such as: title, accession number, the date materials were created, the extent, copyright information, and processing status. Ideally, an institution establishes a common set of metadata fields (a schema) to guide the information that it records for every accession, and this metadata conforms to, or at least maps to, known metadata standards.

CASE STUDY: DC PUBLIC LIBRARY

"We use ArchivesSpace for creating accessions, creating access points to materials...whenever something comes in, accession data is entered in Google Sheets format and then after that, the sheets are used as a buffer for entering data. That information is double checked and then, and only then, does it make its way into our ArchivesSpace instance." (2018)

CASE STUDY: KANSAS HISTORICAL SOCIETY

"Most description happens at the point of accessioning—there aren't really multiple points along the way where the initial metadata is enriched by different staff like processing archivists. Outside of a schema—in DART they also include information like is it a state agency or which agency, etc. On the newspaper side, most of the metadata is in DART but there is also a microfilm database that is part of it." (2018)

Implementation recommendations

- Consider how to document the rights transferred or given to the institution that allow for management, access, and dissemination of content. Check donor agreements, purchase agreements, or transfer agreements for copyright, property rights, and license information. This information can all be recorded in the accession record.

- Enhance your accession record with scripts or tools that automate links between the initial transfer, accession records, and access/preservation copies.
- Use a single unit of measurement for describing the extent of accessions (e.g., megabytes) and use multiple "extent" data points if you need to.
- Decide and document when accessioning is "complete" for your institutional context. Relationship-building, buy-in, communication, and clear documentation are needed in order to make clear where handoffs occur and to systematize workflows.

OSS Tools/Resources
- ArchivesSpace is an open-source archives information management application for managing and providing web access to archives, manuscripts, and digital objects.
- Cornell's Copyright Information Center provides helpful documentation that can help you determine copyright duration and its implications for accessioning content.

8. Analyze and identify sensitive content

Institutions that collect born-digital material often acquire files that contain sensitive content, including personally identifiable information (PII) and personal health information (PHI).

Examples of PII and PHI range from credit card and Social Security numbers to health information, employee information, and student records. Analyzing directory structures or disk images to identify sensitive content is a critical step in the born-digital workflow. Once a digital object is known to contain sensitive information, the collecting institution can make informed decisions about how to handle that PII, including through restricting item usage or redacting the PII. Actions taken regarding PII (e.g., redacting or deleting/obscuring the PII) are often recorded in the digital object's metadata.

Implementation recommendations

- Determine what is sensitive content at your institution; this may be informed by people outside of your unit.
- Analyze, identify, and review the digital material for sensitive content.
- Generate a redaction workflow that results in disk images/files without accessible sensitive information, capturing information about the redaction event alongside preservation copies.
- Consider de-accessioning material that has low informational value and would be risky to the institution or donor to manage over time.
- Consider where and how restricted material will be stored until it can be released.
- If restricting content, consider where metadata developed during this step needs to be stored to ensure that future access providers have the information they need to open the content. Also consider the possibility of using that metadata as a trigger to automate lifting digital access restrictions.

OSS Tools/Resources

- Bulk Extractor Viewer is a graphical interface for bulk_extractor; together, these applications scan a disk image, a file, or a directory of files; extract useful information without parsing the file system or file system structures; and then enable you to view the results of extracted information.
- Bulk Reviewer is a tool that builds on bulk_extractor, adding features for review and removal of sensitive files and PII in directories and disk images.

9. Analyze forensic/technical metadata

Forensic analysis of files provides crucial insight into how and when files were created or modified, as well as into the systems and environments in which files were created.

A range of important information can be extracted including the file type, the file creator, its creation date, and when and how it has been accessed or changed. Information gleaned from forensic analysis can be exported from several tools in the form of Digital Forensics XML (Garfinkel 2012) which can be used in the arrangement and description of digital archival collections.

Implementation recommendations
- Understand aspects of files that impact description, arrangement, and preservation activities so you can determine what types of forensic analysis you want to perform.
- Forensic analysis can be labor intensive and time consuming for staff, but is often necessary for understanding how files were created or how they should be preserved. Consider for which steps and which types of content forensic analysis is appropriate to create a minimally viable product.
- Forensic analysis may be a necessary step if you need to ensure that PII/PHI is not included in digital objects made available in your institution.

OSS Tools/Resources
- Fiwalk collects metadata about the file system, capturing information about the dates that files were last accessed or modified, file type, and creator information. Fiwalk can be run on the command line or as part of the BitCurator reporting tool.
- Brunnhilde is a Siegfried-based reporting tool that can be used to analyze directories and disk images.

10. Create/extract digital object metadata

Digital object metadata helps institutions facilitate discovery, access, preservation and use options for born-digital materials. Depending on the original resource format, some kinds of metadata can be extracted or added manually at different points in the workflow. Determining what metadata an institution wants to maintain, and establishing a schema and a registry or database location in which to maintain it, are critical in born-digital archiving. This step points to the need to plan, understand, and document how information about the digital object is generated in almost every other step chronicled above (e.g., file identification and format characterization, run a virus check, analyze forensic/technical metadata, analyze and identify sensitive content, check file integrity and ensure fixity, create an accession record).

CASE STUDY: DUKE UNIVERSITY

"For description, we take pieces of metadata...output from Brunnhilde, Siegfried characteristics, duplicates, unidentified files...we use pieces of metadata from those reports and may look at individual files to figure out what they are, what content is if the reports don't indicate. We use this to create scope notes (to item or disk level); extent notes (number of files, aggregate file size); general physical description note - types of files - '100 document files in PDF and Microsoft Word format'; coverage date (fiwalk or Siegfried reports). Writing notes might occur outside of ArchivesSpace, but ends up in ArchivesSpace." (2018)

Implementation recommendations
- Develop use case scenarios to help guide decision making around access and preservation choices. For example, consider how researchers may search for materials and complete descriptive metadata to facilitate discovery, access, and use.
- Advocate for automation support. Some metadata handoffs may be automated, but developing the technical infrastructure to support automation must be a collaborative and carefully planned project rather than the task of a single staff member.

- Consider that metadata can be interoperable with other tools, such as output from BitCurator being input into ArchiveSpace accession records.

OSS Tools/Resources

- <u>UC Guidelines for Born-Digital Archival Description</u>, created by a working group of digital archivists under the aegis of the UC Born-Digital Content Common Knowledge Group (CKG), is a UC-wide descriptive standard for born-digital archival material.
- <u>Fiwalk</u> collects metadata about the file system, capturing information about the dates that files were last accessed or modified, file type, and creator information. Fiwalk can be run on the command line or <u>as part of the BitCurator reporting tool</u>.
- <u>Siegfried</u> is a signature-based file format identification tool.
- <u>Brunnhilde</u> is a Siegfried-based reporting tool that can be used to analyze directories and disk images.

11. Assemble AIP

The term "AIP" refers to an Archival Information Package, a central concept in the Open Archival Information Systems (OAIS) Reference model (Consultative Committee for Space Data Systems 2012).

AIPs typically contain the digital files in a collection, metadata about those files, and any other information needed to understand and use the files. Institutions build these packages to facilitate digital preservation, essentially bundling together all of the information that a repository needs in order to plan and implement the long-term curation and storage of a digital object.

CASE STUDY: RICE UNIVERSITY

"We create AIPs that include extracted metadata (output from DROID and EXIFTool), a text file with descriptive metadata, reports, and collection material using internally defined structure. The AIP folder structure is based on required elements...not all our AIPs are bagged, but they are all in a formal AIP format. Some AIPs proved hard to bag for various reasons, such as containing long file names or deeply nested files with long names, tripping up the bag." (2018)

25

Implementation recommendations:

- Collocate all information needed to use files in the future along with the files.

OSS Tools/Resources

- <u>Archivematica</u> is a web- and standards-based, open-source application which allows institutions to produce AIPs that can be stored and preserved for long-term access.
- <u>Bagger</u> is an application that packages data files according to the BagIt specification.
- <u>Exactly</u> is a tool created by AVP that uses the BagIt specification to create packages that include information about the package transfer and the original directory structure.
- <u>Overview for using Bagger</u> provides documentation regarding how to implement Bagger for assembling AIPs for submission into a preservation repository.

12. Assemble DIP

DIPs are Dissemination Information Packages. In the OAIS Reference Model, a DIP converts all or a subset of the AIP (the package stored within a digital preservation environment) into an access copy that is provided to a system to deliver to end users. The DIP has to include all of the information an end user needs in order to understand and render the package, for example including an inventory and perhaps a link to the description of the objects or a summary of how and why they were created. Some institutions assemble DIPs on-the-fly in direct response to user requests for materials, while other institutions assemble DIPs as a routine workflow step and maintain these information packages along with the collection.

CASE STUDY: KANSAS HISTORICAL SOCIETY

"For the born-digital newspaper collection, the ultimate goal … is for those files to be on their servers and to have metadata associated with them so we know what the files are and can provide public access to those files in the research room. [This is not yet] as well developed in the sense that they aren't yet generated as an [information package] that is stored in our robust preservation environment." (2018)

   -Megan Rohleder (pictured left; also pictured Josh Hogan)

Implementation recommendations:

- Maintain AIPs separately from DIPs to retain archival/master copies in digital preservation storage that are not changed, similar to how master copies of analog materials might be kept in a vault or limited access storage area.
- Consider adding enhancements to digital objects included in DIPs to increase accessibility and usability of materials, such as sound quality improvements.
- Consider whether the DIP will be delivered directly to a researcher through ad-hoc access workflows or a direct download of the package itself, or if the DIP will be parsed by a system such as the institution's digital repository. This can help you make decisions on what packaging format to use for DIPs.
- Consider whether a staff member will mediate access to the DIP or if access will be non-mediated. For example, if your institution has a request system for archival materials, how will the born-digital material be requested? This may inform decisions about DIP packaging formats, repositories, and other parts of the born-digital access workflow.

OSS Tools/Resources

- Automated DIP creation in Archivematica generates access copies of digital objects and packages that can be uploaded to an access system to be parsed or stored for future use.

13. Transfer AIP to preservation environment

After digital collections have been processed, the resulting AIPs can be maintained long term in dedicated preservation environments. These environments can take many forms—from dedicated digital preservation storage with multiple copies and services to public-facing institutional repositories—and can exhibit a range of functionalities like regular fixity checking, file repairs, and other preservation activities. Information about all curation activities that are undertaken on the package while it is in the preservation environment (e.g., running a fixity check and comparing the checksum against the original checksum) are recorded in the AIP's metadata.

"[For the repository], we do all of our own depositing as well as the metadata for it. System adds PREMIS [metadata], which is partially automated, [and is also] doing fixity checks and ingesting metadata that we supply. Once a SIP goes in and becomes an AIP, the support of that software is done by IT, as well as backups in Isilon storage managed by the IT department." (2018)

Implementation recommendations

- Similar to transfering born-digital materials from the donor to the archive, use a tool for transferring to the preservation environment that validates the transfer by verifying file integrity before and after the transfer.
- Perform regular fixity checks at different intervals to help ensure the continued integrity of materials and to identify potential technological failure in digital preservation storage environments.
- In order to protect individuals and organizations represented in the collection and fulfill contractual promises to the donor, consider how sensitive materials in AIPs will remain secure over any restriction periods outlined in the donor agreement.

OSS Tools/Resources

- Islandora is an OSS framework designed to help institutions and organizations and their audiences collaboratively manage and discover digital assets.
- LOCKSS is an OSS preservation software that makes a designated number of copies of each file (or package) it receives, stores them in distributed locations across a network, and performs regular fixity checks on these copies, using hash values to detect degradation, loss, or change.

Other resources/guidance for born-digital workflows

- The Digital Processing Framework is designed as a "minimum processing standard for digital archival content." The framework integrates archival practice with standard digital preservation activities.

- <u>Community-Owned Digital Preservation Tool Registry</u> (COPTR) describes tools useful for digital preservation. It is a finding and evaluation tool to help practitioners discover the tools they need to perform particular preservation tasks.
- <u>Digital Preservation Resources</u> compiles information for members of the international digital preservation community, pooling together knowledge and resources for better preservation efforts.
- The <u>BitCurator Consortium wiki</u> includes documentation for taking many of the steps in this guide using the BitCurator environment.

DOCUMENTING BORN-DIGITAL WORKFLOWS

By Colin Post

This section describes the process for creating both text-based documentation and visual diagrams of born-digital workflows in libraries, archives, and museums.

As previously described in Common Steps in OSS Born-Digital Archiving, *born-digital workflows encompass activities from accessioning born-digital materials to providing end users with access to those materials. Workflows can be completely informal and ad hoc or very formalized (though notably, most institutions do not yet have fully formalized workflows in place for born-digital archiving). In other cases, significant portions of workflows may be formalized but only known tacitly by current staff and not explicitly documented. This section of the Guide provides clear advice and concrete procedures for producing workflow documentation, a crucial step in both formalizing ad hoc processes and generating knowledge artifacts to make a record of current practices.*

There is no single "correct" way to arrange a workflow; instead, various technical and organizational factors shape how institutions work with born-digital materials. This section of the Guide *will demonstrate how to document and make sense of an institution's current, "as is" practices. By following the instructions below, you and your team will create a narrative description and a visual diagram of your workflows. The next section,* Using Workflows, *helps you to build on this information by describing ways to use your workflow documentation to expand and mature your born-digital archiving practices.*

Many institutions have multiple or diverging workflows based on the various types of materials or sources of acquisition. For instance, institutions may have different workflows for internal university records versus manuscript collections from outside donors, or special processes for legacy storage media like floppy disks. Consider first mapping out these workflows separately, and, if applicable, seek ways to integrate those multiple pathways into your visual diagram. However, it is fairly common for institutions to have two or more parallel work processes for different kinds of materials or when diverse sets of tools are required for complex collections, so do not worry about synthesizing pathways in the documentation that are actually distinct in practice.

Assessing Workflows

Thoroughly assessing your current, "as is" practices is the first step in documenting born-digital workflows. If you are worried that you do not have strong born-digital archiving practices to document, you are in good company. In fact, a key finding of the OSSArcFlow project was that very few institutions have well-established workflows in place. Our partners found that the very process of documenting current practices helped them to make sense of what previously felt like largely ad hoc work.

Notably, many collecting institutions would describe their born-digital archiving activities as nascent, in development, or still emerging. Most do not have fully formalized workflows in place for processing and working with their born-digital materials. A formalized workflow is not a prerequisite for documenting your current digital curation practices or for filling out the questionnaire we share below.

Consult with stakeholders

As a first step, we recommend that you identify and gather all of the staff members and stakeholders who are involved in your born-digital archiving work. You may identify roles and stakeholders through the process of documenting the workflows that you had not previously considered. As we emphasize throughout this section, the documentation process is iterative and benefits from multiple passes, so you can bring these additional stakeholders in for their feedback at later stages as well. Due to a range of factors like changing organizational structures or shifting staff responsibilities, workflows often cut across departments and involve staff in a variety of roles. Individual staff members likely have only partial knowledge of the overall workflows. Undertaking workflow documentation in a small team representative of all the staff and units involved in digital curation will provide a range of perspectives and help surface activities and steps that might otherwise remain invisible.

Coming together (virtually or in person) to assess your workflows is a constructive exercise that will foster your shared awareness of the work currently being done and provide you with a space to build consensus around your future goals and priorities for born-digital archiving.

Digital curation practices may change quickly, and so this documentation serves as a living artifact of an ongoing process. It also serves as a record of how your organization worked with born-digital material at different points in time, which will be useful for downstream preservation and dissemination activities. As we reiterate below, this team can work together on all steps of the documentation described in this section.

Use the Assessment Questionnaire

The OSSArcFlow project team has developed an Assessment Questionnaire (see Appendix A) to facilitate the information gathering process. The Assessment Questionnaire will guide you to explore and describe the beginning and end points of your workflows and all major activities in between. Involve everyone in your institution that engages with born-digital archiving, either by jointly filling out the questionnaire, or by consolidating responses after each person fills out the questionnaire independently.

Using succinct prompts, the questionnaire will help you and your colleagues to fully characterize each step in your born-digital workflows. In the responses, provide as much information as possible, including vital details like the staff roles, technologies, and techniques involved in these processes. We will show you how to derive the subsequent workflow documentation from the questionnaire to streamline this material into a more usable form, but this documentation will be more accurate and offer more insight if it is based on a wealth of information. Do not try to make your workflow seem "cleaner" than it is; try to include every decision point, action point, and roadblock that you encounter.

The questionnaire has been adapted from an interview protocol first developed for the OSSArcFlow project and later honed for consultations with BitCurator Consortium members. In both iterations, our research team used these interviews to guide staff responsible for various aspects of digital curation at their institutions through the process of articulating their workflow activities.

CASE STUDY: RICE UNIVERSITY

Rice University, a private research university in Houston, Texas, was one of the 12 partner institutions for the OSSArcFlow project. At the time of the project, Rice had five archivists who all spent some time on digital curation activities, although no individual staff member specialized exclusively as a "digital archivist." The Woodson Research Center at Rice collects digital materials in a variety of formats and media and undertakes many of the digital curation steps outlined in Common

Pictured (L to R): Rebecca Russell and Michael Olson

Steps in OSS Born-Digital Archival Workflows. *As a smaller research institution with an established digital collecting program, Rice offers a good example for other institutions just getting started documenting their born-digital workflows. Throughout this section, we will use examples and excerpts from Rice's 2018 born-digital workflow to illustrate the documentation process.*

An archivist at Rice, Rebecca Russell was the principal partner who participated in the project activities. At the initial stage of the project, Russell completed a preliminary "snapshot" providing an overview of digital curation practices at Rice. She then participated in an in-depth interview with OSSArcFlow project team members Andrew Rabkin and Jessica Meyerson.

In the interview, Russell described Rice's "as is" digital curation practices (as of 2018) from the point when materials enter the archives all the way through to providing access to researchers. Russell discussed BitCurator, among other tools and systems used in the workflow, highlighting both the successes and challenges in using these tools. For BitCurator specifically, Russell noted the very labor- and time-intensive activities involved in processing legacy media. Russell also described more general challenges, such as a helpful IT department albeit with limited time and resources for installation, troubleshooting, and other OSS IT tasks. For open-source tools like BitCurator, the archivists at Rice initially have had to work through implementation issues largely on their own.

OSSARCFLOW PARTNER REFLECTIONS
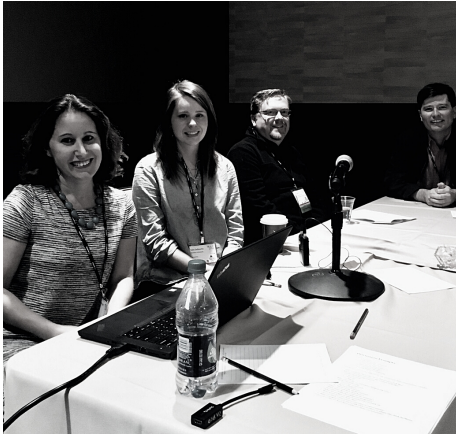
"The OSSArcFlow project has given us a more holistic view of our workflows and provided an opportunity to step back and analyze our processes. Participating in this project has also provided us with a larger network of colleagues to reach out to for advice or issues we encounter."

Quote by Rebecca Russell
Pictured (L to R): Alex Chassanoff, Sally DeBauche, Josh Hogan, and Shaun Trujillo at 2018 DLF Forum panel on OSSArcFlow

Describing Workflows

The next steps will help you to shape the information gathered by the questionnaire into a more streamlined description that delineates the steps making up the workflow. This will necessarily involve some simplification and will abstract away some of the rich detail solicited by the questionnaire, but the resulting textual descriptions and visual diagram will provide overviews that help to make sense of how all the various steps fit together—and to identify gaps where steps do not quite fit together yet.

Producing Text-based Workflow Documentation

We suggest two approaches to describing workflows: generating a narrative description in an outline format and a tabular description in a spreadsheet (templates for both of these are included in Appendix B). Although institutions can choose to adopt one approach over the other, we recommend that you use both in tandem. The narrative outline and the spreadsheet each capture slightly different aspects of the workflow, and each can be instrumental in creating the visual representation of the workflow.

For the narrative description, comb through the questionnaire responses and break these down into discrete steps. Each step description should include an agent (the staff or staff members carrying out the activity), the action carried out, the tool or technologies used, and any other relevant information.

The narrative description preserves much of the detail from the questionnaire while ordering activities in an easy-to-follow linear progression. If applicable, the outline can be further organized into distinct workflow stages or phases (e.g. accessioning, processing, description, preservation, etc.).

EXAMPLE: EXCERPT FROM RICE UNIVERSITY'S 2018 NARRATIVE DESCRIPTION

- **ACCESSIONING**:
    - IF there is no existing accession record for the collection THEN
        - [Archivist] creates accession record in ArchivesSpace
    - Store materials in folders by call #/accession # on Fondren Library project server (fonlibstor) in the WRC folder which is redundantly backed up by IT
    - Track the material in the Nearline tracking sheet

- **PROCESSING**:
    - Born digital materials on storage media
        - [Archivist] uses the BitCurator environment to create forensic disk images and run reports to locate Personally Identifiable Information (PII)
            - IF PII is found THEN: (Rice has not encountered this scenario yet and does not yet have a protocol)
                - [Archivist] will stop processing, add restricted note in ArchivesSpace

In this example, you can see Russell working through issues and gaps in the workflow. Russell knew they would need a protocol in place for when personally identifiable information is found in a collection, but they had not yet dealt with this situation at Rice.

For the tabular description, put the steps outlined in the narrative description into a spreadsheet. We recommend including the following elements for each step:
- the broader phase or stage in which the step is carried out;
- any conditions that impact how, if, or when a step is carried out;
- a brief description of the activity involved in the step;

- the software used during the step (if any);
- the hardware used (if any); and
- the staff member or staff role responsible for carrying out the step (this can be included in the description or as a separate element).

However, your institution can expand or contract the extent of the spreadsheet as you see fit.

EXAMPLE: EXCERPT FROM RICE UNIVERSITY'S TABULAR DESCRIPTION

	A phase	B	C step	D description	E software
1	phase		step	description	software
4	accessioning		accession record creation	[Archivist] creates accession record	ArchivesSpace
5	accessioning		store materials	Store materials in folders by call #/accession # on Fondren Library project server (fonlibstor) in the WRC folder which is redundantly backed up by IT	
6	accessioning		begin tracking materials	Track the material in the Nearline tracking sheet	
7			forensic imaging	If digital content is on storage media, [Archivist] uses the BitCurator Environment to create forensic disk images and run reports to locate Personally Identifiable Information (PII)	Guymager; BulkExtractor
8			encountering PII	If PII is found, [Archivist] will stop processing the collection and add a "restricted" note in ArchivesSpace	BulkExtractor
9			Obtain fixity information	Run BulkExtractor reports to obtain file formats, MAC times and fixity information	BulkExtractor, fiwalk
10	processing	if on storage media	AIP creation	[Archivist] create AIPs that include extracted metadata, reports and collection material using internally defined structure	
				Normalize files as needed (depending	

Here you can see Russell separating out different pathways for materials on storage media versus files acquired via a network transfer. The storage media undergo additional steps like forensic imaging that are not necessary for files sent over email or a cloud service.

In both the narrative and tabular descriptions, steps should be defined clearly and concisely. It is important to think about how you define your workflow steps at this stage because this language will feed directly into your visual workflow diagram. First, consider how finely you want to divide your workflow activities into discrete steps.

Generally speaking, each step consists of a single activity with some readily delimited goal, aim, or outcome. Examples of these steps are discussed at length in **Common Steps in OSS Born-Digital Archival Workflows,** and you can also see how other institutions have defined these steps by reviewing the OSSArcFlow workflow diagram examples.

Applying clear and consistent terminology throughout your workflow documentation will help you use these documents to communicate across departments within your institution and to compare workflows developed by your peer institutions. Different departments may have varying understandings of a term, and likewise, institutions may use the same term to describe different activities or use different terms to describe the same activity. Donors, researchers, and other user communities may also use entirely different vocabulary. Along with the steps and the OSSArcFlow project examples discussed in this **Guide**, we highly recommend "The Digital Processing Framework" (Faulder et al. 2018) as a strong set of shared terms for digital curation activities. Your workflow documents should ultimately speak to the practices of your particular institution, though, and as such, you may need to balance local terminology with more standardized terms used in the broader field.

Reviewing Your Text-based Workflow Documentation

As you develop your narrative workflows, you will benefit from taking several iterative passes. Going over the questionnaire responses multiple times and making adjustments to the descriptive documents will help you to ensure that details of your workflows are fully reflected. Likewise, comparing the narrative and tabular descriptions against one another can improve the accuracy of your documentation and provide further insights. Something represented well in the narrative description may not be reflected in the tabular description. Make sure both sets of documents align with each other, even if each provides a slightly different perspective on the work carried out on the ground.

This synthesis step will benefit tremendously from the involvement of and/or review by all staff and stakeholders involved in different parts of the workflow. At this point in the process, you likely have already identified gaps and challenges in your workflows, or surfaced points of confusion where workflow processes cut across departments.

This documentation process can help to reveal some of those challenges and potentially clarify those points of confusion, but you need not resolve all of these issues right away. Ongoing challenges can be reflected and annotated in the textual documentation and the visual diagram, and these can all be used as tools to address these issues moving forward.

Consider sharing these documents for feedback before moving onto the visual representation of the workflow, since this diagram is generated directly from the descriptive documents. Along with the visual diagram, both the narrative and tabular workflow descriptions are valuable knowledge artifacts that can be used independently to make sense of current practices, identify gaps, and expand workflow activities.

Representing Workflows Visually

Visual workflow representations are powerful documents that offer both a commanding overview of the entire workflow and detailed depictions of each constituent component or step.

Essentially, these visual representations comprise boxes for each step and arrows that connect the steps sequentially, although a great deal of detail and customization can be added to this basic formula. Complementing the descriptive documentation discussed above, these visual representations can be used to assess current practices, communicate between staff and across departments, and plan for the future.

Developing Visual Workflow Diagrams

When constructing a visual workflow representation, there are a number of stylistic choices and practical decisions that need to be made. These choices will impact both the look and usability of the resulting document. Choosing a workflow modeling tool is the first step in this process. Many such tools exist at different price points and with varying features, and so it is important to choose an option that best suits the needs of your institution. The OSSArcFlow project adopted LucidChart, a web-based application that enabled collaboration for our distributed team; many other institutions have followed suit, resulting in a growing body of LucidChart-based diagrams.

The features and settings of the modeling software will influence the construction of the visual workflow representation, but there are certain practical considerations to weigh regardless of the particular tool. Whether to orient the diagram horizontally or vertically is a significant decision that greatly affects how the document is read and used.

The OSSArcFlow workflow diagrams are oriented horizontally, suggesting a linear progression from start to finish. However, project members also recognized the potential utility of vertically oriented diagrams, more closely resembling the top-to-bottom progression of information common to flow charts. Both orientations have benefits and disadvantages, although both can be interacted with as digital files or analog printouts. As workflow modeling becomes a more widespread technique for digital curation, the community will likely arrive at a clearer consensus around this issue.

The other important stylistic choice regards the symbols and signs used in the workflow diagram. The modeling software will likely come with a set of default options as well as the ability to adapt these or create new symbols. As noted above, the basic structure we have used in our visual representations are boxes for each workflow activity connected in sequential order by arrows, but various symbols and shapes can add nuance to the diagram. The OSSArcFlow diagrams include symbols to indicate steps that happen concurrently, decision points where workflows branch off, and "pain points" where certain aspects of current workflows could be updated or improved. Though additional symbols increase the information communicated by the diagram, these need to be applied consistently and in a balanced way so that the document can still be read and understood easily. Find the template used for the OSSArcFlow project in Appendix C. This template includes a legend for all the symbols used in partners' workflow diagrams. You can reuse or adapt these symbols for your own workflow diagrams.

EXAMPLE: EXCERPT FROM RICE UNIVERSITY'S 2018 VISUAL WORKFLOW REPRESENTATION
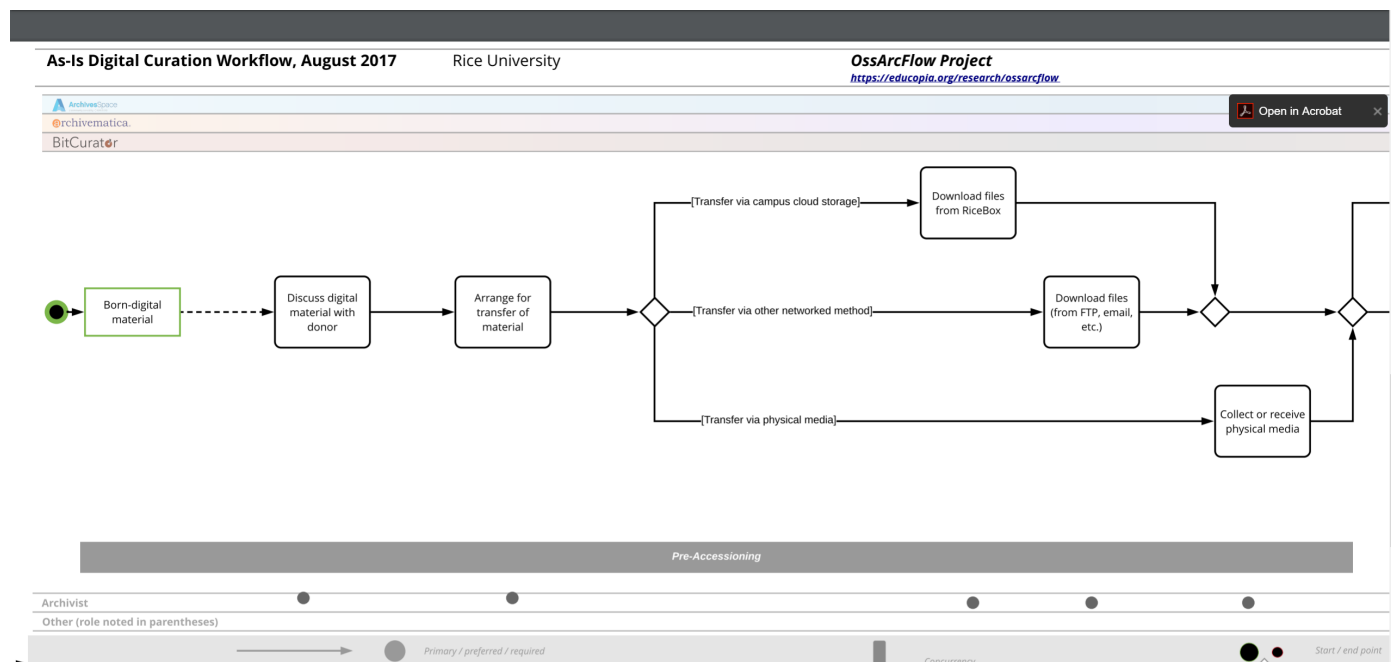
This excerpt from <u>Rice's visual diagram</u> illustrates the portions of the narrative and tabular descriptions included above. You can see how the visual diagram provides an overview of multiple diverging pathways, in this case the different ways materials are acquired by the institution.

The visual representation can be derived directly from the descriptive documents discussed above by translating each row in the tabular description or each bullet point in the narrative description into a step box in the diagram. The text included in each box should provide a concise and clear definition of the workflow step so that the diagram can be read quickly and easily. Arrows can be used to connect sequential steps, and decision points where multiple steps branch off or come together can be indicated with special symbols. In these cases, the arrows can also be labeled to indicate the conditions of the distinct workflow trajectories.

In addition to the workflow steps and activities, the visual representation can include other details and annotations from the descriptive documents. For example, the OSSArcFlow diagrams include tracks running parallel to the workflow steps to mark the staff role and the open-source tool involved in each activity.

Many of the OSSArcFlow diagrams include "sticky note" boxes that provide more extensive explanations of the activities, techniques, or tools involved in a particular step. LucidChart (and likely other workflow modeling tools) supports a variety of these kinds of annotations.

As with the text-based workflow documentation, constructing the visual diagram is also an iterative process that benefits from multiple passes and input from all staff and stakeholders involved in born-digital archiving. In putting this diagram together, the arrangement of steps might shift or new connections between steps might be made as the visual mode of thinking provides insight into aspects of the workflow that were not previously apparent. This is especially true for arranging multiple branching pathways for workflow activities since it is difficult to comprehensively depict these conditional steps in the textual descriptive documents.

There is no ideal born-digital workflow, and practices that suit one institution may not work for another.

There is no ideal born-digital workflow, and practices that suit one institution may not work for another. Similarly, no born-digital workflow will work perfectly all the time for a given institution. No matter how thorough your documentation is, some digital archival materials will need special attention and require ad hoc steps. Some of this can be anticipated and annotated in your workflow documentation, for instance recognizing that certain file formats or media types will need to be treated separately. Emory University, for example, has developed a tiered processing system that funnels a select group of materials into a specialized workflow at the outset based on a number of factors.

However, workflow documentation will never be fully complete or comprehensive. All of these documents are artifacts of a learning process, gleaned from digging deeply into and reflecting upon current digital curation practices. These are also living documents that should be updated regularly to reflect the dynamic nature of workflows, although older documentation should also be maintained to illustrate the history of changes in workflows. This ongoing documentation process is integral to the development of workflows. As we will discuss in **Using Workflows**, all of these workflow documents are tools for assessing current practices and advancing born-digital workflows to meet future goals and priorities.

USING WORKFLOWS

By Alexandra Chassanoff

This section provides guidance and examples of how libraries, archives, museums, and other collecting institutions can use their documented workflows to self-evaluate, compare against workflows from similar institutions, and identify potential areas of growth and goals for maturity over time.

This section opens by describing how workflows function as crucial artifacts in our still-emerging born-digital community of practice. It then describes ways that workflow diagrams developed using a common model (e.g., the process described in Documenting Born-Digital Workflows) can be compared and contrasted to see differences between institutions and also to identify where both technical factors (e.g., the operating system used, whether servers are local or in a cloud environment, the software used for different steps, and the programming knowledge and ability) and organizational factors (e.g., roles and responsibilities, expectations, historical connections and disconnections, departmental boundaries) intervene and lead to diverse instantiations of similar processes and toolsets in different institutions.

OSSARCFLOW PARTNER REFLECTIONS

"I have used the born-digital workflow documentation created as part of the OSSArcFlow project in internal conversations, as well as in conversations with other institutions, to illustrate the successes and pain points in our work. The project has helped us think more critically about what we do, when in the workflow we do it, and why." (2018)

-[Matthew] Farrell (pictured right; also pictured Kelly Stewart)

Workflows as artifacts

Many collecting institutions are in the beginning stages of making necessary changes to their workflows to accommodate the acquisition and processing of born-digital materials. The complexities introduced by new file formats, standards, and responsibilities can be significant hurdles for institutions implementing processes and programs that include born-digital collections. Workflow modeling is one approach that can aid cultural heritage organizations as they confront new challenges posed by managing digital collections (Collins 2009). Processing steps for born-digital materials are often less linear and far less visible than for analog materials, and so workflow representations serve as assessment tools for evaluating current practices and consideration of how software solutions can best be incorporated (Anderson 2014; Dowdy and Raeford 2014). The process of modeling workflows is itself instructive, providing space to iteratively reflect on digital curation and digital preservation practices as they currently exist, to identify the social and technical dimensions impacting these current processes, and to articulate future goals (Barbrow and Hartline 2015; Daines 2011).

The OSSArcFlow research project sought to address the challenges of implementing born-digital workflows by investigating and documenting current practices in a sample of 12 institutions. The project assembled a diverse group of partners, including an Ivy League research library, several mid-sized private and public research libraries, a library collaborative, a liberal arts college library, a research lab, several public libraries, and a historical society. Over the course of the project, the team encouraged institutions to engage with their own workflows as artifacts of current practice. Monthly meetings provided a landing space for partners to exchange tips on processing techniques, describe new frameworks and resources, and share institutional strategies. The project also developed smaller "breakout groups" to focus on particular discussion topics during breakout sessions that partners felt posed challenges. Prior to each session, partners completed a brief questionnaire on how their institutions were addressing different topics (see Appendix E). The information provided by partners can be used as a case study in anticipating common implementation challenges. Partners noted that documentation efforts played an essential role in helping them ideate and envision different opportunities for their own local archiving practices.

	Discussion topics
Cohort 1	Expanding staff roles, responsibilities and training
Cohort 2	Scaling up institutional capacity and infrastructure
Cohort 3	Automated versus manual processes
Cohort 4	Long term storage for larger collections
Cohort 5	Describing born-digital materials

Initially, the OSSArcFlow project sought to synchronize workflows across open-source software systems to understand how institutions moved from their existing practices toward more streamlined processing. However, in the process of documenting existing workflows the unique challenges facing individual institutions became clear. The vast discrepancies in processing steps and tools, as well as the contrasting manner in partners' implementation efforts, demonstrated that particular variables in existing workflows (what we ended up calling "pain points") played a significant role in how institutions expanded the robustness and capacity of their digital preservation programs. Exposing these pain points and then helping partners articulate potential paths forward became a central (if unexpected) achievement for the project.

OSSArcFlow Partners at in-person meeting in Chapel Hill, NC (December 2017)

Workflow Pain Points

As we discuss throughout the *Guide*, the process of creating the documentation itself was crucial in helping partners to recognize and articulate pain points, and to then take steps toward addressing those issues.

Throughout the OSSArcFlow project, the project team worked with partners to identify "pain points" in their born-digital workflows. These included all manner of issues and challenges that impacted digital curation work. These pain points were often precisely the obstacles that made it difficult for partners to streamline or advance their workflows in ways they had hoped to at the outset of the project. As we discuss throughout the Guide, the process of creating the documentation documentation itself was crucial in helping partners to recognize and articulate these pain points, and to then take steps toward addressing those issues. We grouped these pain points together in three main categories.

Handoff Between Systems
These pain points occurred when partners encountered difficulties taking data or metadata generated by one system and moving it into another. Integrating these kinds of system handoffs into workflows often requires labor- and time-intensive work. Metadata exported from one system/tool may have to be transformed significantly for successful input into another system/tool. Additionally, handoffs between systems for our partners introduced semantic challenges because they used localized approaches in their system/tool implementations. Even if tools are open source, a scripted handoff that works for one institution may not work well for another. For example, both Duke and Stanford expressed interest in automating transfer of metadata output generated by tools in the BitCurator environment into ArchivesSpace archival description fields. However, because both partners had custom implementations, the same script was not usable for both institutions.

System-specific Challenges
In addition to issues with handoffs between systems, partners had challenges implementing systems or integrating system functions into their workflows.

In some cases, tools may have limited functionality that falls short of an institution's needs. For instance, NYPL wished that Archivematica worked with a wider range of email file formats, such as mbox or EML. In other cases, tools may not work as advertised or may result in frequent technical issues. Duke reported issues with the Guymager tool in the BitCurator environment when attempting to image optical media. In general, open-source tools can also have steep learning curves, for instance requiring users to access certain functions through the command line.

Partner-specific Issues

This group of pain points stemmed from factors specific to particular institutions. While these issues have technical aspects, social or organizational factors exacerbate these technological challenges. For example, Stanford needed to run the bulk_extractor tool from the BitCurator environment on a dark server due to policy decisions around the handling of sensitive data. This additional layer of security, though, compounded the steps in their workflow. NYPL collects materials from visual and performing artists, many of whom use specialized audiovisual software. These collections are difficult to curate using standardized processes, often requiring emulation or migration pathways. Many institutions described backlogs of digital materials and anticipated this problem to amplify in the future with the continued influx of larger and larger digital collections. Across the board, partners worried that current workflows would not scale up to meet the demands of these larger collections.

For more information on pain points analysis in the OSSArcFlow project, please view the video panels our team recorded in December 2017: Introduction (Cal Lee), Panel 1 (Joshua Hogan, Atlanta University Center, Robert W. Woodruff Library; Megan Rohleder, Kansas Historical Society; Michael Olson, Stanford University), Panel 2 (Jonathan Crabtree, Odum Institute; Matthew Farrell, Duke University; Nick Krabbenhoft, New York Public Library; Rebecca Russell, Woodson Research Center, Rice University), and Panel 3 Part One and Part Two (Paul Kelly, D.C. Public Library; Don Mennerich, New York University; Kari Smith, Massachusetts Institute of Technology; Shaun Trujilo, Mt. Holyoke College)

Comparing and analyzing workflows

The partner workflows created by the OSSArcFlow project capture the wide range of activities, tools, standards, software, and methodologies at play in contemporary born-digital archival practices.

As artifacts, the partner workflows identify emerging trends in the field while also showcasing potential common pain points among partners at similar stages of maturity. Using existing community practices to ground institutional choices is a strategy advocated by current archival practitioners, who note, "In the face of many unknowns about stewarding digital materials, archivists can no longer work in silos but must collaborate via participatory networks" (Clemens et al. 2020, 19).

As points of comparison, institutions can use partner workflows to evaluate their own capacity and guide decision making on implementation choices. For example, in our project, the majority of partner institutions created forensic disk images of physical media as part of their processing. The three institutions that did not do so during the project (AUC Woodruff, MIT, and UNC Odum) primarily received born-digital content through file transfer. Depending on institutional needs and desired outcomes, how born-digital materials are acquired may play an important role in determining workflow steps and advancement possibilities. Other important information can be gleaned from comparing and analyzing the specific tools that partners use in forensic image creation. For example, eight of the nine institutions that create disk images use the Guymager tool within the BitCurator environment. Identifying particular tools and the points at which they are used in the overall workflow process can help to lower implementation barriers for institutions. At the same time, knowing the environment(s) in which tool use takes place provides important context for automation efforts that might rely on metadata exchange.

Comparison and analysis of single institutions with different kinds of workflows can also help illustrate how the acquisition of different types of materials may direct processing steps. For example, at the time of the project New York University used three different workflow pathways for acquiring and processing materials depending on material type.

Noting similar processing steps and tool use across pathways may indicate that particular tools (e.g., MediaLog) are appropriate choices for institutions wanting to capture technical metadata about their born-digital materials.

CASE STUDY: NEW YORK UNIVERSITY

Workflow Pathway 1: Heterogenous Born Digital Materials
- Create forensic disk image
- Digital Archives Staff runs DROID to derive the type(s) of data and digital extent
- Digital Archives Staff hands off digital extent to Accessioning Archivist to create the accession record in ArchivesSpace
- MEDIALOG is used to capture technical information:
  - the type of storage medium
  - workflow management documentation
  - Google Drive folder is created if that folder does not already exist

Workflow Pathway 2: Donor-Appraised, Curated Set of Digital Files
- Create logical copy
- Digital Archives Staff runs DROID to derive the type(s) of data and digital extent
- Digital Archives Staff hands off digital extent to the Accessioning Archivist to create the accession record in ArchivesSpace
- MEDIALOG is used to capture technical information:
  - the type of storage medium
  - workflow management documentation
  - Google Drive folder is created if that folder does not already exist

Workflow Pathway 3: Organizational Partners—Collecting on an Ongoing Basis/Structured Transfer
- Self ingest if file size and scope of accession is under a certain size
- Logical copy
- MEDIALOG is used to capture technical information:
  - the type of storage medium
  - workflow management documentation
  - Google Drive folder is created if that folder does not already exist

Identifying gaps and challenges

A wide range of gaps and challenges, or "pain points" as we described them in the OSSArcFlow project, impact all stages of born-digital workflows. From demonstrating the difficulties involved in using new OSS tools to surfacing the missing points of communication between staff across departments, workflow modeling helps to reveal potentially inhibiting factors that can prevent institutions from advancing workflow practices. In fact, the act of generating documentation often forces institutions to confront these issues that may not have been readily apparent. For example, Emory University describes the difficulties in "figuring out where the points of hand-off are between manuscript and digital archivists. What falls under the purview of the manuscript archivist and at what point does it revert to the responsibility of the digital archivist?" Additionally, the use of formalized notations in visual workflows can capture important context about variables that may impede institutional streamlining of workflow processes.

EXAMPLE: EXCERPT FROM EMORY UNIVERSITY'S 2018 VISUAL WORKFLOW REPRESENTATION

In the example above, the presence of curatorial staff as both primary and secondary players in the pre-accessioning phase of processing shows how roles (and corresponding responsibilities) impact the execution of workflows in practice.

In many institutional environments, staff already have varying levels of technical fluency or expertise using the systems, tools, and scripts required for digital archiving. The value of tacit knowledge in workflow activities was noted by many participating institutions. Given high staff turnover, documentation efforts can help support the transfer of this knowledge. At the same time, institutions should be prepared to offer specific training for incoming staff as well as to establish baseline competencies expected for different stages of processing.

As one partner from Duke University noted, "Even if I understand and wish more of our staff were comfortable/used to command line work, there's a tension between a minimal viable product and a service that is 'easy to use' by people with all sorts of tech skills." Without staff who can help "glue" these tools together through the development of scripts, and professional support for staff to spend time learning and experimenting with building their own automations, the manual load for moving content between tools quickly grows too heavy to scale.

The need for allocated IT resources and dedicated staff support to run open-source software such as BitCurator can also prove to be an obstacle. Another barrier to implementation noted by project partners was the difficulty of using the command line interface. The DC Public Library pointed out, "While the command line is easy to get to grips with, it does present a barrier that will cause many archivists to simply ignore the tools entirely." Indeed, without adequate technical scaffolding, the realities of born-digital work can mean immediately shelving potentially unstable media for "future processing." Under-resourced institutions face particular challenges with implementation. In some cases, support may need to come from beyond the archives itself (e.g., from a different department or unit in a shared parent institution, or from a partner that has technical staffing).

Popularly used digital curation tools often exist within a variety of operating systems and environments. Workflows easily illustrate these complex tool chains, which often require adequate pipelines between system inputs and outputs to function correctly. For example, project partner Stanford University found that university-wide policies regarding high-risk data required the archives staff to be especially thorough in identifying personally identifiable information in incoming digital collections. Stanford was already using tools in the BitCurator environment to accomplish this task, but the existing hardware did not possess sufficient processing power to cull through large amounts of data in a time-efficient manner. Based on the workflow documentation and additional experimentation, Stanford established the amount of processing power they would need to efficiently process large disk images in accordance with the university-wide policies. The workflow documentation helped Stanford to identify a gap in their current workflow and take concrete steps to address and heal that issue.

Advancing aspirational goals

Workflow modeling can help lay the foundation for current practices while also helping to set the stage for workflow advancement.

Comparison and analysis of workflows helps institutions to identify and assess current gaps in emerging practices, while also helping to prompt articulation of future goals for institutions still actively building out their processing. As models, such artifacts present institutions with a universe of possible activities, roles, responsibilities, software, and decisions showing what born-digital processing actually looks like in action.

In the OSSArcFlow project, partners developed consensus on different born-digital strategies to help move processing steps forward. For example, a common shared pain point was that the description and arrangement practices for analog collections do not map well onto born-digital materials. Absent best practices, partners from Duke and MIT used a breakout discussion session to demonstrate how they use ArchivesSpace to describe their born-digital collections. The resulting discussion surfaced many of the pressing questions archivists will need to address as the profession develops these best practices. The documentation for both Duke and MIT served as important illustrations, helping all participants envision how ArchivesSpace can support emergent description and arrangement practices, and how these practices might fit into their own workflows.

In another example of shared consensus, project partners from Emory and Rice came together to discuss expanding staff roles and responsibilities in light of budget restrictions. Through this discussion, the partners identified concrete strategies they could use to improve their local advocacy efforts for expanding resource allocations to born-digital archiving. While most partners expressed the desire to improve workflow efficiency and processing times, shared institutional experiences provided the community with the necessary context for anticipating possible solutions for advancing workflows.

Partner workflows can be used to compare and contrast workflow steps, tool use, and decision-making that effectively shape how institutions carry out their born-digital processing.

Previously published research from the project demonstrates the important role that workflows play in developing digital curation processes (both the human labor and tools/technical processes involved). Yet the project also demonstrated the incompleteness of workflows, given the always evolving complicated scenario "on the ground" (Post et al. 2019). Institutions continually develop procedures as staff try out new tools and learn new standards. In seeking to develop robust workflows that further extend and streamline processing, institutions can draw on workflow artifacts and use provided templates to ground their own decision making and guidance. These resources provide a horizon of possibilities for developing strong, sustainable processing workflows and outcomes to which institutions can aspire.

APPENDICES

Appendix A: Assessment Questionnaire

Overview

1. Describe the primary kinds of born-digital content collected by your institution and how these materials are typically acquired (e.g., stored on hard disks, files transferred via email/cloud service).
2. Do these materials fall into any main categories (e.g., university archives, manuscripts/personal papers, electronic theses)?
3. Do you have an existing workflow (or multiple workflows) in place for working with this content? Briefly describe the overall workflow in broad strokes.
4. What (if any) are the primary tools or software suites used for working with born-digital materials (e.g., TeraCopy, Bagger, BitCurator, ArchivesSpace)? Don't feel the need to list every tool used throughout the workflow at this point.

Steps and activities

1. What is the starting point (or starting points) for the workflow (e.g., media drives or disks acquired, files received via email)?
2. What is the end point (or end points) of the workflow (e.g., AIP moved to long-term storage, access copies made available)?
3. What are the main (5-7) activities that are performed during the workflow?
4. If applicable, group these activities into broader phases or stages (e.g., pre-accessioning, accessioning, description and arrangement, preservation, access).

For each main activity:

1. Can these be broken down into more finely defined steps? Describe each step in a bit more detail.
2. What staff or staff roles are involved in these steps or activities?
3. What tools, technologies, systems, or techniques are involved in these steps or activities?

It may be helpful to fill out this portion of the questionnaire in a chart.

Activity	People/Role	System/tool
Create disk image	Digital Archivist	Guymager
Transfer to permanent storage	Digital Archivist	TeraCopy

Challenges and areas for improvement
1. What are the main goals and priorities for processing materials in the current workflow?
2. Are there overarching challenges or specific issues preventing the accomplishment of these goals and priorities?
   a. Are there challenges or limitations related to particular tools or software used in the workflow?
   b. Are there gaps or delays? Missing steps or missing connections between steps?
3. What are the goals and priorities for the ongoing development of the workflow?
   a. Are there specific steps or activities that you would ideally integrate into the workflow?
   b. Are there specific tools or technologies that would help to better achieve either present or future goals and priorities?
   c. Are there additional personnel (or additional training for existing staff) that would help to better achieve these priorities and goals?

Appendix B: Workflow Description Templates

Narrative Description Template

- **PRE-ACCESSIONING**
  - [Staff role] does [some step] using [this tool]
    - [Additional details, conditions, or issues related to this step]
- **ACCESSIONING**
  - [Staff role] does [some step] using [this tool]
    - [Additional details, conditions, or issues related to this step]
- **PROCESSING**
  - [Staff role] does [some step] using [this tool]
    - [Additional details, conditions, or issues related to this step]
- **ACCESS**
  - [Staff role] does [some step] using [this tool]
    - [Additional details, conditions, or issues related to this step]

**These phases can be changed or expanded to reflect the workflow at a particular institution. For instance, an institution might refer to pre-accessioning as appraisal or might have a separate preservation phase.
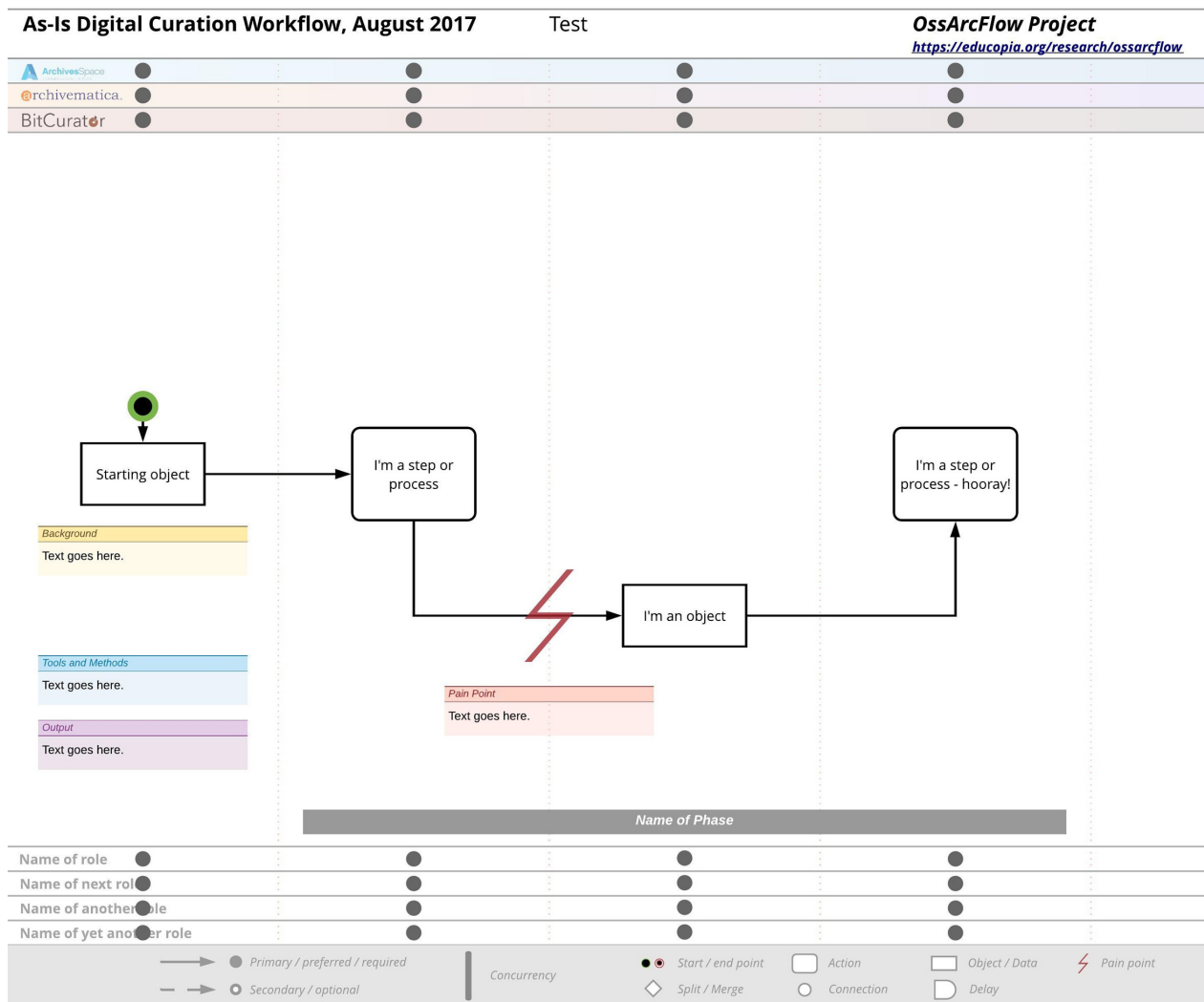
Access a copy of the template as a Google Doc here.

Tabular Description Template

Phase	Conditions	Step	Description	Software	Hardware
Pre-accessioning		Action carried out	[Staff role] does this action, described in more detail		This hardware used
Accessioning	If this is the case, then...	Action carried out	[Staff role] does this action, described in more detail	This software used	

Access a copy of the template as a Google Sheet here.

Appendix C: Workflow Diagram Template

Access a copy of the template in LucidChart <u>here</u>. Note: you may have to create a LucidChart account in order to access this template.

Appendix D: Use Cases for Workflow Documentation

Individual institutions can use their own workflows to:
- Capture dated snapshots of local practices as they morph and mature over time
- Engage with their own workflows as artifacts
- Provoke conversation and clarification of roles and processes
- Ideate and envision different opportunities for local archiving practices
- Expose pain points
- Prompt articulation of future goals
- Train new staff and archivists
- Illustrate how the acquisition of different types of materials may direct processing steps
- Reveal potentially inhibiting factors that can prevent institutions from advancing workflow practices
  - Demonstrating the difficulties involved in using new OSS tools
  - Demonstrating the impact of infrastructure changes (e.g., shifting from local servers to off-site and cloud environments)
  - Surfacing the missing points of communication between staff across departments
  - Showing where the institution meets challenges that lead to storage and postponing decisions (e.g., shelving potentially unstable media for "future processing")

Institutions can use each others' workflows to:
- Exchange tips on processing techniques
- Describe new frameworks and resources
- Share institutional strategies
- See how different staffing models and partnerships might work
- Expose common pain points and then collaboratively articulate potential paths forward
- Anticipate their own capacity needs by looking at their peers' workflows
- Guide decision making on implementation choices

Shared documentation can be used at the field level to:
- Identify emerging trends in the field
- Incentivize funders to invest in common needs
- Showcase pain points that are common among partners at similar stages of maturity
- Identify particular tools and the points at which they are used in the overall workflow process
- Lower implementation barriers for institutions
- Provide important context for automation efforts that might rely on metadata exchange
- Show what born-digital processing actually looks like in action
- Identify and assess current gaps in emerging practices

Appendix E: Discussion Group Case Studies

Name:

Institution:

Job Title/Role:

- What are the particular problem areas or challenges with regards to <DISCUSSION TOPIC>

- What are some of your goals for <DISCUSSION TOPIC>? What do you hope to accomplish in your workflow that you are not currently able to accomplish now?

- What are some of the institutional factors that impact this problem area (e.g., budget, staffing, organizational structure)?

- What are some of the technological factors that impact this problem area (e.g., hardware or software issues, particular aspects of IT systems at institution)?

- Are there any other factors that you see influencing this problem area?

- What are some of the ways that you have already addressed your problem area? What are some of the ways that you plan to address your problem area in the next 1-5 years?

- Have you experienced any obstacles as you have addressed this problem area? Do you foresee any further obstacles as you continue to address this problem area?

REFERENCES

Anderson, Elsa K. "Chapter 3: Workflow Analysis." *Library Technology Reports* 50, no. 3 (May 13, 2014): 23–29.

Barbrow, Sarah, and Megan Hartline. "Process Mapping as Organizational Assessment in Academic Libraries." Edited by Professor Alice L. Daugherty. *Performance Measurement and Metrics* 16, no. 1 (January 1, 2015): 34–47. https://doi.org/10.1108/PMM-11-2014-0040.

Clemens, Alison, Wendy Hagenmaier, Jessica Meyerson, and Rachel Appel. "Participatory Archival Research and Development: The Born-Digital Access Initiative." *Provenance, Journal of the Society of Georgia Archivists* 36, no. 1 (April 1, 2020). https://digitalcommons.kennesaw.edu/provenance/vol36/iss1/3.

Collins, Maria. "Evolving Workflows: Knowing When to Hold'em, Knowing When to Fold'em." *The Serials Librarian* 57, no. 3 (September 28, 2009): 261–71. https://doi.org/10.1080/03615260902877050.

Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)*. Washington, DC: Consultative Committee for Space Data Systems, 2012.

Daines, J. Gordon. "Re-Engineering Archives: Business Process Management (BPM) and the Quest for Archival Efficiency." *The American Archivist* 74, no. 1 (April 1, 2011): 123–57. https://doi.org/10.17723/aarc.74.1.h8159344u8331165.

Dowdy, Beverly, and Rosalyn Raeford. "Electronic Resources Workflow: Design, Analysis and Technologies for an Overdue Solution." *Serials Review* 40, no. 3 (July 3, 2014): 175–87. https://doi.org/10.1080/00987913.2014.950040.

Faulder, Erin, Susanne Annand, Sally DeBauche, Martin Gengenbach, Karla Irwin, Julie Musson, Shira Peltzman, Kate Tasker, Laura Uglean Jackson, and Dorothy Waugh. "Digital Processing Framework." Report, August 2018. https://ecommons.cornell.edu/handle/1813/57659.

Garfinkel, Simson. "Digital Forensics XML and the DFXML Toolset." *Digital Investigation* 8, no. 3 (February 1, 2012): 161–74. https://doi.org/10.1016/j.diin.2011.11.002.

Post, Colin, Alexandra Chassanoff, Christopher Lee, Andrew Rabkin, Yinglong Zhang, Katherine Skinner, and Sam Meister. "Digital Curation at Work: Modeling Workflows for Digital Archival Materials." In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 39–48, 2019. https://doi.org/10.1109/JCDL.2019.00016.

Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, 57–66. JCDL '11. Ottawa, Ontario, Canada: Association for Computing Machinery, 2011. https://doi.org/10.1145/1998076.1998088.