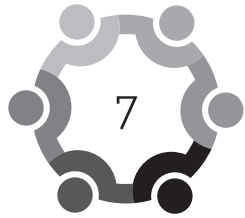

Digital Preservation Should Be More Holistic

A Digital Stewardship Approach

Somaya Langley

A considerable amount of creative, cultural, and research output is expressed in digital form. It is imperative that the memory sector rapidly improve its capability and capacity for handling digital content in all forms, including complex data. It took gallery, library, archive, and museum (GLAM) sector institutions somewhere between decades and centuries to implement systems for managing physical collections. Those who work with digital content are acutely aware that the same time frames are not afforded when it comes to saving our digital cultural heritage. The "fragility" of content produced from computing environments—thanks to the rapid churn of technological innovation and obsolescence[1]—means that even acquiring, preserving, and providing sustained access to a seemingly "simple" stand-alone file can take considerable effort. The interdependencies and limitations of the computing platforms, software, hardware, and other peripherals (whether mass-manufactured or custom-developed) bring a level of complexity that is typically not experienced with physical collections.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

For over two decades, the challenges of managing digital content have been discussed and debated. For organizations yet to establish acquisition and preservation programs for born-digital and digitized content, it is imperative to do so soon. To support this work and ensure success, a robust methodology is required. This methodology must be one that incorporates and creates an exchange between different professional paradigms in order to guide and enhance this work. In some narratives, digital preservation is bounded by certain tasks, with other activities undertaken under the umbrella of "digital archiving" or "digital curation."[2] Yet the lines between these disciplines are neither distinct nor clear-cut. If the aim is to provide long-term, sustained access to digital content, then preservation work is a necessity. To ensure that the critical steps needed for handling digital content do not slip into the gaps between disciplines, framing the management of digital content under the term *digital stewardship* provides a more holistic view of all the activities that need to be undertaken.

The "digital stewardship approach" is a concept the author has been forming over the past seven years, in collaboration with current and former colleagues at institutions in Australia and the United Kingdom. Born out of practical experience in digital acquisition and preservation activities, this work has developed as an attempt to narrow the gaps between current theoretical frameworks and day-to-day practical realities. A holistic approach also encourages collaboration, drawing together practitioners from different disciplines with complementary skill sets in order to enrich, streamline, and consolidate the management of the vast array of digital content.

The digital stewardship approach takes the form of a practice-based guide for handling digital content. This includes two practical tools: the *Digital Stewardship End-to-End Workflow Model* for visualizing the different stages that digital content passes through, and the *Digital Streams Matrix* to advise on the different pathways and tasks for processing digital content. Developed in order to meet a direct practical gap, these "alpha release" tools are intended to complement existing models and guidelines from the archiving, digital curation, and digital preservation disciplines.

As the prevalence of digital content increases, it is essential to strike a balance between the need to collect and manage digital content on the one hand, and the limited available resources the GLAM sector has to support this necessary work on the other. Working to a baseline of "good practice" is critical, and pragmatic approaches are essential.[3] A holistic, digital stewardship approach towards handling digital content provides the necessary perspective in order to consider where our work efforts are best directed.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Understanding Digital Content

For the custodians of digital content, it is accepted that context must be well understood.[4] In addition, it is essential to possess a solid awareness of how digital content is conceived and created, plus what has happened to it prior to transferring custody of it to a GLAM institution. This increases the likelihood that digital preservation activities can be undertaken with increased confidence that the content or meaning of this digital content will not be altered.

The digital preservation discipline often advocates for handling digital content via batch processes. Given limited resources, this is essential. But without fully understanding the digital content's context, or what has taken place prior to its arrival at a GLAM institution (due to lack of information captured), future digital preservation work may be compromised, resulting in undesired or potentially catastrophic outcomes. Adopting a digital stewardship approach and focusing on capturing metadata and other information as early as possible in the "life cycle" of digital content will greatly assist preservation activities.

Where the Digital Preservation Work Starts

The term *digital preservation* is still somewhat difficult to define, particularly when identifying tasks that fall within or outside of its scope. In 2006, the Joint Information Systems Committee (JISC) defined digital preservation as "the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value."[5] In the context of digital preservation and research data management (RDM), assumptions are often made that this work starts once data is already controlled and located in a local networked server environment. Bundling and ingesting digital content into a digital preservation system (Archivematica,[6] Preservica,[7] RODA,[8] Rosetta,[9] etc.) or digital repository (DSpace,[10] Fedora,[11] etc.) are frequently seen as some of the first steps that must be taken. File format identification, characterization, validation, checksum generation, and virus-checking are accepted as actions that fit within the digital preservation remit, as are preservation actions such as migration and normalization. The provision of access via emulation or virtualization is also considered digital preservation work.[12]

Tasks needing to be carried out prior to these "accepted" digital preservation activities, including transferring data from donors or researchers, transferring content off physical format digital carriers (e.g., floppy disks, optical media,

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

portable external hard disk drives, USB flash drives, etc.), and organizing this data are seen as the responsibility of archivists, curators, and librarians. For institutions that have yet to employ digitally skilled archivists, curators, or librarians (who possess both the necessary GLAM and information and communications technology knowledge), there is the risk that traditionally trained staff (as well as some research data managers) do not yet have the requisite technical skill sets or necessary in-depth understanding of certain types of digital content. While these staff possess valuable professional expertise and experience that are essential for working in GLAM institutions, undertaking the appropriate transfer and handling of digital content demands considerable technical knowledge. As a result, files can be transferred to an institution in inadequate ways (via drag and drop, uploaded via web interfaces, etc.) without an appreciation of how data and metadata may be modified or lost (e.g., the date last modified), particularly if this data needs to be relied upon in the future.[13] In the worst cases, lack of understanding of a carrier can result in the data being deemed corrupted, rather than it being assessed as an inability to access the data (e.g., an uncommon disk file system).[14]

Digital preservation activities need to commence far earlier in the life cycle of digital content, not immediately prior to the content being ingested into a digital preservation system or digital repository. Instead, this work needs to begin when the digital content is being conceived, and it should be "baked in" to all processes from the beginning.[15] From an RDM perspective, a "sheer curation" approach would support the production of better-quality digital content.[16] Digital preservation responsibilities can no longer be considered as separate from the effort it takes to advise on the creation and acquisition of digital content, and vice versa. Embracing holistic approaches to managing digital cultural and research content, such as digital stewardship, should be encouraged.

What Is Digital Stewardship?

In 2011, Butch Lazorchak published a post on the Library of Congress's blog *The Signal.* In it, he broadly describes the differences between the terms *digital preservation*, *digital curation,* and *digital stewardship*.[17] For those working in GLAM contexts, *digital preservation* is the most common of these terms encountered. However, as has been discussed, some necessary tasks for managing digital content may be overlooked, if we are only viewing this work through a digital preservation "lens." The term *digital curation* began to be used to describe the management of data within research contexts.[18]

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

The Digital Curation Centre's Curation Lifecycle Model (first published in 2007) made visible the range of different stages that digital content needs to progress through in order to be adequately managed. While digital curation provides a clearer illustration of what is required to support digital content through all the stages of its creation, preservation, and ongoing access, there are still considerable gaps between this approach and the "hands-on" practicalities of being a custodian of digital content. The notion of digital stewardship encompasses both digital curation and digital preservation.[19] In the United States, the National Digital Stewardship Alliance (NDSA) was launched in 2010, and this initiative framed the approach to managing digital content within the government, educational, and not-for-profit sectors.[20] Outside the United States, the concept of digital stewardship, while acknowledged in some contexts, is yet to be adopted.

Why Consider Digital Stewardship?

In order to prolong the life span of digital content, preservation activities need to be factored in right from conception, rather than consisting of action taken merely when an issue or problem is discovered. If preservation is only thought of when digital content is "finished" (and custody of it has been transferred), it is possible that specific details—which are important for ensuring the digital content's authenticity, provenance, and the long-term access to it—will not have been captured. Alternatively, these details may have been intentionally or unintentionally modified, or discarded.

It may be far too expensive to address preservation issues only after problems are encountered or when obsolescence is imminent. Digital stewardship provides a much-needed overarching perspective for digital content creators and custodians.

Unlike paper-based archival materials, validating the authenticity and provenance of unmanaged digital content is nearly impossible. An "original" of a file is a misnomer. What is possible is to identify a set of files and obtain the "earliest" available metadata and other information related to those files, in order to capture the "best possible" representation and details about the digital content's context.[21] The later in time that data and metadata are captured, the greater the risk of inaccuracies being introduced or incorrect assumptions being made. Comprehending digital content, and attesting to its authenticity and provenance, as well as accurately presenting and contextualizing it, are all critical in the digital environment.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Existing Models

There is a range of different models[22] that are used in digital preservation and digital curation work. Given the breadth and complexity of digital content, a range of models to suit different purposes and functions is a necessity. Because there is a multitude of ways of approaching the management of digital content, relying on only one model is unlikely to meet the needs of all scenarios.

While the available models provide a good foundation for digital preservation and digital curation work, some gaps have been identified. This chapter attempts to address these gaps, but in doing so, it is crucial that selected models from several different disciplines are discussed. What follows is by no means the full breadth of models that are available for use in managing digital cultural and research content.[23]

Open Archival Information System Reference Model

The draft recommendations of the Open Archival Information System (OAIS) Reference Model were released in 1999.[24] Since that time, it has been published as an ISO standard (ISO 14721:2012).[25] The model has been used widely in digital preservation as the foundation for conceptualizing the management of digital content.[26] In the OAIS Reference Model paradigm, a "digital object" is made up of both files and metadata (which typically contains crucial information about the files). Since the release of the OAIS Reference Model, a range of digital preservation systems—such as Archivematica, Preservica, RODA, and Rosetta—assert that they comply with the OAIS Reference Model.

The OAIS Reference Model provides a high-level overview of how digital files need to be prepared, ingested, and managed within a digital preservation system.[27] Not only are digital preservation systems and repositories conceptualized around this model, but it is also used as a foundation for preservation education.[28] And yet, no specifics are provided on how data should be processed in order to wrangle it into each of the different states (i.e., a Submission Information Package, an Archival Information Package, or a Dissemination Information Package). It should also be acknowledged that because the OAIS Reference Model was developed to manage space science data, it does not necessarily suit all digital content contexts.[29]

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Digital Curation Centre—Curation Lifecycle Model

The Digital Curation Centre's (DCC) Curation Lifecycle Model (first released as a draft in 2007)[30] was developed for use in the newly emerging field of digital curation.[31] The Curation Lifecycle Model provides a more thorough illustration of how digital content must progress through a series of stages in order to be adequately managed. The model also acknowledges that the process of managing digital content is continuous. So as to reflect this, the visual representation of the model is as a continuous "lifecycle." In addition, the model recognizes that working with digital content is rarely linear, and certain activities are likely to be iterative.

The DCC Curation Lifecycle Model was developed to support both the GLAM and research sectors.[32] In developing the DCC Curation Lifecycle Model, it was noted that complete control over the whole life cycle of digital content would be the ideal scenario, but this is rarely possible.[33] In RDM practices, anecdotal accounts indicate that there tends to be less focus—compared to the archival domain—on maintaining the authenticity and provenance of digital content prior to its being ingested into a digital repository.[34] Whereas for archivists working with born-digital personal and corporate records,[35] controlling both the data and metadata in order to ensure that no intentional or unintentional changes occur is of crucial importance.

While the Curation Lifecycle Model is intended to be more practical, gaps become evident when using it to guide operational work. For born-digital content, and particularly born-digital personal and corporate records, additional stages are required in order to manage and transfer the custody of digital content.[36]

JISC—Research360 Institutional Research Lifecycle Concept

The Research360 Institutional Research Lifecycle Concept provides a high-level set of stages for managing research data.[37] It is based on two other models (the Idealized Scientific Research Activity Lifecycle Model and the UK Data Archive Lifecycle Model) that were developed to manage scientific data in the United Kingdom.[38] While these models factor in important aspects of managing research data in the U.K. context (e.g., the Research Excellence Framework),

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

they do not reflect the actual operational tasks of managing digital content that can guide RDM and GLAM staff in handling research data.

Other Tools to Support Digital Content Management Workflows

While the following are not defined as models, information provided in these resources supports digital content management workflows and is useful for comparison and contextualization of the Digital Stewardship End-to-End Workflow Model.

Digital Preservation Outreach and Education Baseline Digital Preservation Curriculum

Modeled on the OAIS, the overarching headings of the Digital Preservation Outreach and Education (DPOE) Baseline Digital Preservation Curriculum provide a way of viewing the broad stages that digital content needs to progress through in order to be managed.[39] The curriculum stages are:

1. Identify
2. Select
3. Store
4. Protect
5. Manage
6. Provide

These curriculum stages are typically used to train GLAM staff who are entering the digital preservation space for the first time, and so they only provide a conceptual outline. While the DPOE curriculum is a useful framework for beginners, it doesn't provide suitable guidance for staff who need to undertake operational work with digital content.

Preserving (Digital) Objects With Restricted Resources Tool Grid

The Preserving (Digital) Objects With Restricted Resources (POWRR) Tool Grid is not intended as a workflow model. However, in order to select tools to undertake different digital preservation activities, various overarching stages are defined, grouping tools by the different types of functions they can carry out.[40] These overarching stages are comparable with the stages in the Digital Stewardship End-to-End Workflow Model.

The POWRR Tool Grid also provides more granular information on the types of tasks typically carried out at each of these stages.[41] These include:

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

- Ingest (Copy, Fixity Check, Virus Scan, File Dedupe, Auto Unique ID)
- Processing (Auto Metadata Creation, Auto Metadata Harvest, Manual Metadata, Rights Management, Package Metadata, Auto Submission Information Package Creation)
- Access (Public Interface, Auto Dissemination Information Package Creation)
- Storage (Auto Archival Information Package Creation, Reliable Long-Term Bit Preservation, Redundancy, Geographically Dispersed Data Storage Model, Exit Strategy)
- Maintenance (Migration, Monitoring, Auto Recovery)

Further information is also provided under an "Other" category.[42]

An Inter-Institutional Model for Stewardship—Four Functions of Stewardship

An Inter-Institutional Model for Stewardship (AIMS) was a collaborative project between the University of Hull Library, Stanford University Libraries, the University of Virginia Libraries, and Yale University Library.[43] Running from 2009 to 2011, the project developed the framework called "The Four Functions of Stewardship," which comprises

- Collection Development
- Accessioning
- Arrangement and Description
- Discovery and Access

An outcome of the AIMS project was the AIMS Digital Material Survey for Personal Digital Archives.[44] The AIMS project acknowledges its similarities with the Personal Archives Accessible in DIGital Media (PARADIGM) Workbook (discussed later in this chapter), with the intention that the AIMS framework and the PARADIGM Workbook be used in conjunction with each other.[45]

Records Continuum Model

To briefly contextualize this chapter in terms of the available models in related disciplines, it is important to mention that there are numerous models suitable for use in digital preservation used in the archival context, including several "continuum models." One of the better known of these is the Records

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Continuum Model, which was developed by Frank Upward (of Monash University, Australia) just over two decades ago.[46] Rather than a workflow or life cycle, this model is visualized as a set of "concentric rings."[47] These are represented as

1. Create
2. Capture
3. Organize
4. Pluralize

As is the case with other digital preservation and digital curation models, this doesn't provide the guidance that archivists, curators, librarians, and research data managers are seeking in order to undertake operational work with digital content.

National Digital Stewardship Alliance Levels of Digital Preservation

The National Digital Stewardship Alliance's (NDSA) Levels of Digital Preservation are a set of recommendations[48] that are often used for maturity modeling and risk management.[49] The NDSA was established in 2010 out of the Library of Congress's National Digital Information Infrastructure and Preservation Program (NDIIPP). In 2012, the NDSA Levels of Digital Preservation guidelines were published (as Release Candidate One).[50] With regard to the name of the NDSA itself, it is positive to see the term *digital stewardship* being adopted into mainstream usage in the memory sector. While the NDSA Levels of Digital Preservation are intended for a different purpose, it is worth noting their alignment with the tasks identified in the POWRR Tools Grid.

Archival Processes

In discussing born-digital personal and corporate records in GLAM and research contexts, it is essential to discuss archival handling processes. In transitioning to the digital context, it is important not to reinvent the wheel. Rather, adapting and building upon existing principles is a pragmatic strategy.

Archival Theory and Practice

Archivists are adept at working in the context of core archival functions to process archival collections.[51] Areas of this work include negotiation and donor relations (including facilitating donor agreements), appraisal and selection, acquisition (including the transfer of records), accessioning, rights management

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

(including the application of access restrictions) to records, and arrangement and description.[52]

Fundamental to archival work is the maintenance of authenticity and provenance. For digital content, it is important that robust processes are established to support the retention and protection of the data and metadata, in order to support the verification of authenticity and provenance for the digital content. This is necessary due to the ease with which digital content (and its metadata) can be intentionally or unintentionally altered. In the shift to the digital environment, the need to ensure that both the data and metadata have not been modified is no longer only a concern for archivists. For curators, librarians, and research data managers, being able to verify at any point in time the authenticity and provenance of digital content (including published digital content, digitized content, and research outputs) is critical. Adopting archival principles can facilitate the improved management of digital content in other GLAM and research contexts.

Archival Processing Tools

Two significant tools that have been developed to support the management of born-digital personal and corporate records are discussed here.[53] Staff working in digital curation, digital preservation, and RDM would benefit from considering approaches that are derived from archival practice, particularly when negotiating with researchers and donors.

The PARADIGM Digital Private Papers Workbook

The Personal Archives Accessible in DIGital Media (PARADIGM) project was a collaboration between the John Rylands University Library at the University of Manchester and Bodleian Libraries, University of Oxford, with some funding support from the JISC. The project took place between 2005 and 2007 and was seminal in identifying a range of challenges regarding private papers in the digital context.[54]

One significant outcome of the project was the "Workbook on Digital Private Papers." This guide provides in-depth guidance for handling and processing aspects of born-digital personal and corporate records.[55]

University of British Colombia Library—Donor Survey Instrument

Developed in 2011 as part of the Persistent Digital Collections Strategy by the University of British Columbia (UBC) in Canada, the Donor Survey

Instrument was based on the work of the PARADIGM project and the AIMS Digital Material Survey for Personal Digital Archives.[56] The UBC Donor Survey Instrument facilitates discussion between archivists and donors in order to obtain information about a donor's digital content and the technologies used.

The Digital Stewardship Approach

A digital stewardship approach to support the acquisition, preservation, and access to digital (born-digital and digitized) content is being devised at Cambridge University Library, taking place under the banner of the Digital Preservation at Oxford and Cambridge (DPOC) project.[57] The digital stewardship approach grows out of the author's firsthand experience in acquiring born-digital collections for institutions such as the National Library of Australia, the National Film and Sound Archive of Australia, and the State Library of New South Wales in Australia.[58] Looking at the management of digital content holistically was seen as essential to being able to understand digital content in GLAM institutions' collections.

While in its relative infancy, the digital stewardship approach is made up of two (alpha release) tools. The Digital Stewardship End-to-End Workflow Model is a practice-based workflow model for guiding the process of digital content management from start to finish. This process is conceived as fourteen stages. The accompanying Digital Streams Matrix reflects a list of tasks or approaches for each stage of the Digital Stewardship End-to-End Workflow Model, mapped against a series of classes (and subclasses) of digital content. Together, these two tools can be used as a guide for processing digital content in order to know "what to do next." For staff making the shift from handling paper-based material to handling digital content, a step-by-step guide was seen as the most practical form of assistance to support operational work. These tools assist in decision-making by providing high-level information on what actions or approaches may or may not be suitable (based on institutional policies and/or technical limitations) for each subclass of content.[59]

As has been mentioned, commonly used digital preservation and digital curation models do not provide enough guidance to support operational work. It is for this reason that the Digital Stewardship End-to-End Workflow Model and the Digital Streams Matrix have been developed. Because both of these tools are considered alpha releases, over time they will be refined and extended as Cambridge University Library looks to operationalize support for

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

the acquisition and preservation of digital content, particularly born-digital personal and corporate records.

These tools also attempt to fill the existing gaps in known models, as well as lobbying for digital preservation activities to be undertaken earlier in the life cycle of digital content. The intention is not to reinvent the wheel (or duplicate current work efforts taking place elsewhere), but rather to build upon previous work produced by the digital preservation, digital curation, research, and archival communities. So as to make these tools easier to integrate into existing processes in these disciplines, current concepts and terminology have been retained and reused even where the terms are problematic. While the workflow model presented in figure 7.1 supports the end-to-end processing of a digital collection, it should never be seen as a straightforward linear process. A life cycle or continuum approach is always necessary because managing digital content is a constant active and iterative activity.[60]

The Digital Stewardship End-to-End Workflow Model

The alpha release of the Digital Stewardship End-to-End Workflow Model is conceived as fourteen stages, as illustrated in figure 7.1.

The Digital Stewardship End-to-End Workflow Model has been developed based on fifteen years of real-world experience in acquiring, preserving, and providing access to digital content. Although visualized as a linear workflow, it is important to re-emphasize that, as already mentioned, handling digital content is rarely straightforward. It is common for stages to need repeating or an earlier stage returned to. In addition, for certain classes of digital content (depending on institutional policies), some stages may be skipped altogether.

For each stage, there may be a series of substages. Prepare (Stage 2) involves planning how to create digital content (by the content producer), whereas GLAM staff need to prepare template documents to support the acquisition of digital content. Acquire (Stage 6) includes the transfer of custody of digital

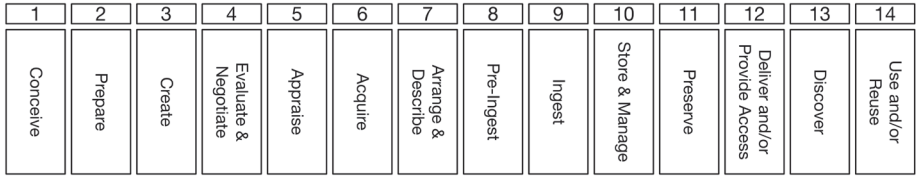

Figure 7.1 • Digital Stewardship End-to-End Workflow Model

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019). © 2019 American Library Association.

content from a donor or researcher to a collecting institution. Either institutional staff or the donor may carry out the actions required at this stage; the goal is to ensure that both the data and metadata are transferred without being modified. Pre-Ingest (Stage 8) contains a number of different substages: Preconditioning, Technical Analysis, and Generate Submission Information Package (both Preconditioning and Technical Analysis are discussed later in this chapter). Preserve (Stage 11) is where Preservation Actions take place, which includes assessing the content in relation to its immediate or near-future inaccessibility, and migration or normalization activities based on this risk assessment. It should be noted that it is the view of the author that digital content should never be stored and then just left to languish. Digital content must be actively managed. It is for this reason that Store and Manage (Stage 10) is represented as a single stage; it is an iterative action that must take place at regular intervals.[61]

Mapping the Digital Stewardship End-to-End Workflow Model

When used in isolation, existing digital preservation and digital curation models fall short of describing all the activity required for managing digital content. The digital curation and RDM disciplines acknowledge that conceiving and creating content is a critical part of the digital content life cycle, and that steps need to be taken to bring digital content under control (particularly when transferring custody of it). However, it must be acknowledged that how digital content is made and managed by creators and researchers can only be influenced, not controlled. The existing models from these disciplines are insufficient, since they do not fully represent the tasks that are undertaken in operational contexts. Other disciplines such as traditional archival and library science or the digital humanities also lack the holistic view of the activities required for managing the full life cycle of digital content. This is likely due to where work effort is typically focused.[62]

Figure 7.2 shows the Digital Stewardship End-to-End Workflow Model alongside various related disciplines (from the perspective of a research library).

Visualizing the disciplines in this way helps to draw attention to where the focus of the work of each discipline lies, as well as illustrating some of the gaps. By managing digital content from the perspective of only one discipline, or tolerating the "siloing" of disciplines, the potential for troublesome issues is introduced. Additionally, the staff establishing workflows to support born-digital content may not have all the guidance they require.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

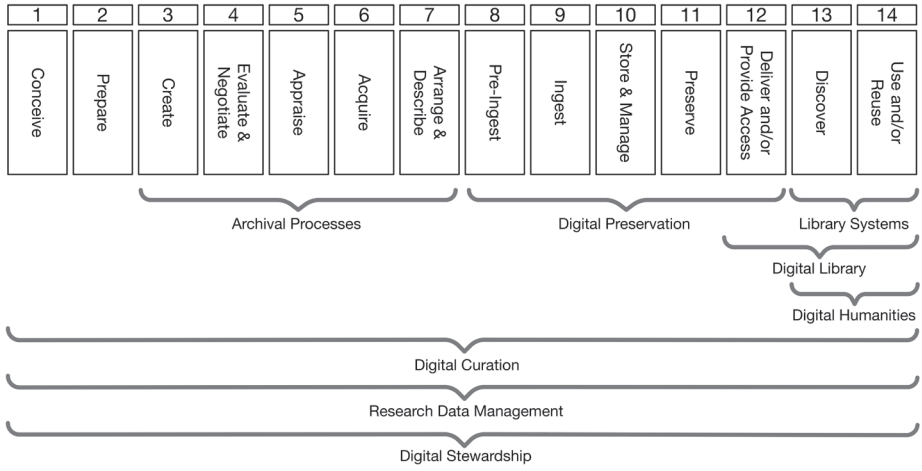

Figure 7.2 • Digital Stewardship End-to-End Workflow Model, also illustrating the focus areas of various GLAM and research disciplines

In order to fully support all aspects of digital content creation and management, taking a holistic approach is necessary. To truly inspect the current gaps, selected models from the archival, digital curation, and digital preservation disciplines must be seen alongside each other.

Filling the Gaps

Supporting born-digital personal and corporate records holistically requires a large proportion of the work to be done up-front; this runs contrary to the mistaken assumption that digital preservation efforts are merely work to be undertaken at a later stage (e.g., creating a Submission Information Package and ingesting this into a digital preservation system).

Figure 7.3 aligns several of the models used in digital preservation, digital curation, RDM, and archival practice that were mentioned earlier in this chapter, mapping these against the Digital Stewardship End-to-End Workflow Model.

Gaps are particularly noticeable in earlier stages of the workflow. Staff actively acquiring and preserving born-digital personal and corporate records will attest to the fact that considerable work must be undertaken prior to ingesting digital content into a digital preservation system. In recent years, cross-disciplinary approaches—borrowing from different GLAM practices—have

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

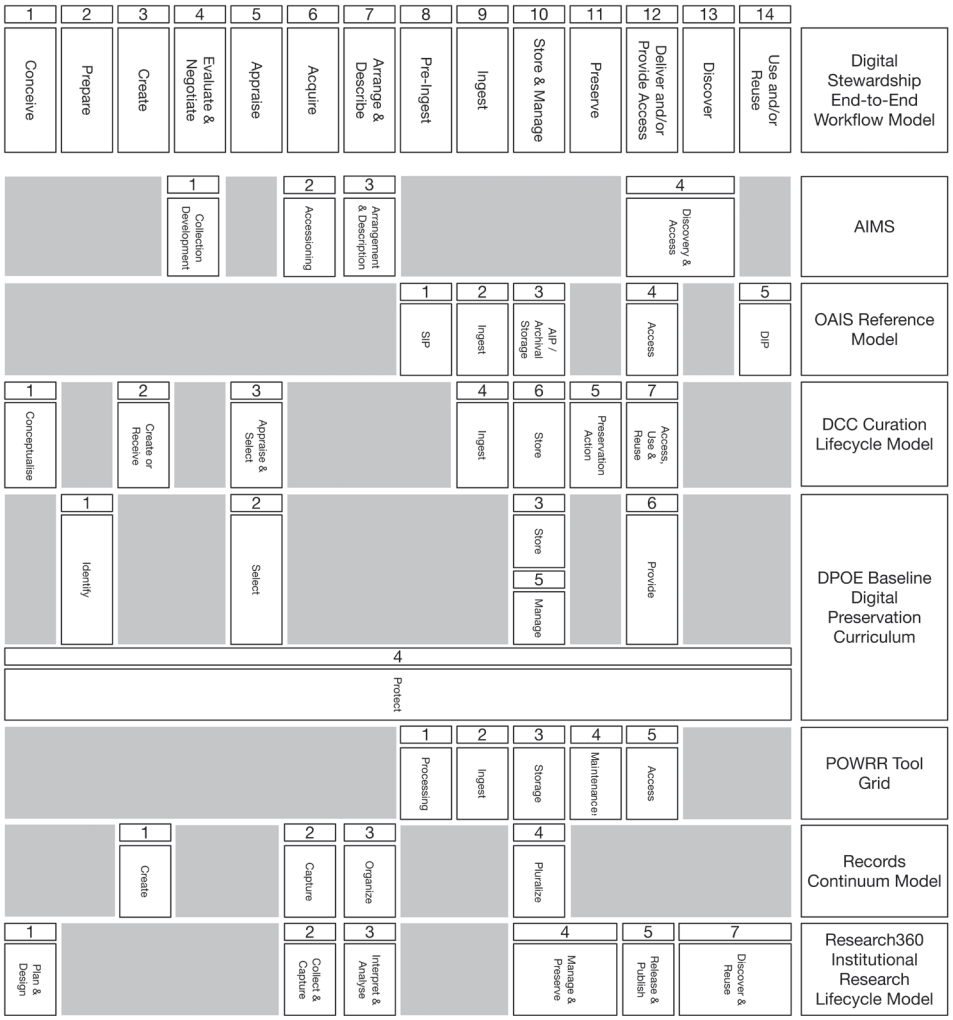

Figure 7.3 • Digital Stewardship End-to-End Workflow Model, alongside selected models

been encouraged. One example is the work on "Repurposing Archival Theory in the Practice of Data Curation," which was presented at the 2014 International Digital Curation Conference.[63] There are benefits to combining several models and approaches, as well as introducing complementary skill sets from similar disciplines. Libraries, galleries, museums, and RDM should look more frequently to the archival discipline for guidance.[64]

There are also a multitude of other considerations that must be factored in when handling digital content. In order to cement these workflow processes, other concerns need to be taken into account, including quality control,[65]

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

monitoring and reporting,[66] managing preservation risks (including preservation planning), data security,[67] applying access restrictions, retention schedules, deaccessioning (including deletion of digital content), configuration of systems and tools, quality assurance of systems and processes, and integration between other institutional systems and tools, as well as managing available resources and adequate resourcing, and so on.[68] Of course, this list of additional factors is not exhaustive. As a side note, while retention isn't explicitly described in the Digital Stewardship End-to-End Workflow Model, the author believes it falls under the umbrella of the Store and Manage (Stage 10) stage.[69]

Digital Collection Classes at Cambridge University Library

The types of digital content held in Cambridge University Library's collection is fairly typical of research libraries around the world. As part of the DPOC project deliverables, a collection survey was carried out (in late 2016 and the first half of 2017). An initial set of five "classes" of collection material had been previously defined, but these five were insufficient.[70] A further two classes were identified: in-house-created digital content and audiovisual content.[71] The seven classes are defined as:

1. ***Born-digital personal and corporate records***—digital archives of significant individuals or institutions
2. ***Born-digital university archives***—selected records of the University of Cambridge
3. ***Research outputs***—research data and publications[72]
4. ***Published born-digital content***—e-books, web archives, digital maps and music, copies of electronic subscriptions (archival and/or access copies, as permitted by agreements), other published born-digital content held on physical format digital carriers (floppy disks, optical media, portable external hard disk drives or USB flash drives), and so on
5. ***Digitized image content***—2D photography and 3D imaging
6. ***In-house-created content***—photography and videography of events, lectures, photos of conservation treatments, and so on
7. ***Digital (and analog) audiovisual content***—moving image (film and video) and sound recordings

However, a workflow model and a set of classes by themselves are not enough to guide staff; developing another tool to support decision-making is necessary.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

The Digital Streams Matrix

As a means of understanding how digital content should be acquired, managed, preserved, and made available the Digital Streams Matrix was developed. Like the Digital Stewardship End-to-End Workflow Model, the Digital Streams Matrix is an alpha release.[73] This matrix will continue to evolve alongside current DPOC project work at Cambridge University Library. However, the focus here is not the full matrix; rather, it is the identified "gap" areas that are present. The following two tables illustrate two stages where in-depth and critical work is required. The amount of effort required at these two stages, Acquire (Stage 6) and Pre-Ingest (Stage 8), is often underestimated.

The Acquire (Stage 6) can be complex with many methods of capture or transfer required, depending on the collecting scenario. Likewise, effort needed during the Pre-Ingest (Stage 8) is significant as this is where technical issues, that may not have been picked up during the Appraise (Stage 5) or Arrange and Describe (Stage 7) stages, will need to be addressed.

For certain stages, typically at either ends of the workflow—particularly for Acquire (Stage 6; shown in table 7.1) and Deliver and/or Provide Access (Stage 12), specific methods of transfer or delivery are required. In the central stages of the workflow, such as Ingest (Stage 9) and Preserve (Stage 11), more homogenization occurs. A long list of tasks may need to be carried out in the more central stages—such as Pre-Ingest (Stage 8; shown in table 7.2) and Preserve (Stage 11). However, the same tasks are likely to be carried out across the majority of digital content. This allows for the streamlining of overall digital collection management processes.[74] Batch processes are critical to digital preservation, since the work effort must be scalable. It is imperative that the context is understood before batch processes are applied, or undesired effects may result (e.g., "broken" digital content).

Applying the Digital Stewardship Approach

In early 2017, over forty digital collections were nominated as potential candidates for case studies as part of Cambridge University Library's DPOC work. Through a thorough selection process, three case studies were chosen, based on their complexity and other parameters.[75] Carrying out case studies was intended as a means of informing digital strategy and policy, and planning for training and skills development in managing digital content, as well as developing requirements for tools, equipment, infrastructure, and the broader digital preservation business case as a whole.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Table 7.1 • Digital Streams Matrix—Acquire (Stage 6)

	BORN-DIGITAL PERSONAL AND CORPORATE RECORDS		BORN-DIGITAL UNIVERSITY ARCHIVES		RESEARCH OUTPUTS			PUBLISHED BORN-DIGITAL CONTENT				DIGITIZED IMAGE CONTENT		IN-HOUSE CREATED CONTENT		DIGITAL (AND ANALOG) AUDIO-VISUAL CONTENT	
	Selected records only	*Whole digital collection (e.g., 'deceased estate')*	*University records*	*University databases*	*Research publications*	*Research Data*	*Software and/or scripts*	*E-resources (e.g., subscription service digital content)*	*Donated published born-digital content (e.g., eBooks)*	*Physical format digital carrier (e.g., floppy disks, optical media, portable external hard disk drives, USB flash drives)*	*Websites*	*Images*	*3D*	*In-house created content (e.g., for public use)*	*In-house created content (e.g., documentation of physical collection items from conservation treatments etc.)*	*Digital audio-visual carriers and files*	*Analog audio-visual carriers*
Optical disk imaging	X	X	X			X				X						X	
Web portal upload					X	X	X		X								
Web archiving (harvesting)	X	X	X		X	X					X						
'Deep web' archiving	X	X				X					X						
Electronic Legal Deposit web archiving	X	X	X						X		X						
Web recorder (Rhizome)	X	X				X					X						
One off 'ad hoc' deposit	X		X		X	X			X	X						X	X
Still image digitization												X	X				
Video digitization																X	
Film digitization																	X
Audio digitization																X	X
Floppy disk imaging	X	X	X			X				X							
Data on tape						X		X				X				X	
Export from cloud storage	X					X										X	
FTP and network file share transfers	X		X			X	X		X			X	X			X	

Table 7.2 • Digital Streams Matrix—Pre-Ingest (Stage 8)

	BORN-DIGITAL PERSONAL AND CORPORATE RECORDS		BORN-DIGITAL UNIVERSITY ARCHIVES		RESEARCH OUTPUTS			PUBLISHED BORN-DIGITAL CONTENT				DIGITIZED IMAGE CONTENT		IN-HOUSE CREATED CONTENT		DIGITAL (AND ANALOG) AUDIO-VISUAL CONTENT	
	Selected records only	*Whole digital collection (e.g., 'deceased estate')*	*University records*	*University databases*	*Research publications*	*Research Data*	*Software and/or scripts*	*E-resources (e.g., subscription service digital content)*	*Donated published born-digital content (e.g., eBooks)*	*Physical format digital carrier (e.g., floppy disks, optical media, portable external hard disk drives, USB flash drives)*	*Websites*	*Images*	*3D*	*In-house created content (e.g., for public use)*	*In-house created content (e.g., documentation of physical collection items from conservation treatments etc.)*	*Digital audio-visual carriers and files*	*Analog audio-visual carriers*
File format identification	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
File format validation	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
File format characterization	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Quarantine	X	X	X	X													X
Virus check	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Checksum hash generation or validation	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Ensure content renders	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Check files meet minimum quality baselines	X					X		X				X	X	X		X	
Generate or verify Technical Manifest	X	X	X			X		X				X	X			X	
Flag 'junk' (or other) files for removal	X	X	X			X				X							
Filename 'normalization' (including documenting 'original' filename)	X	X	X		X	X			X	X		X	X	X	X	X	

Table 7.2 • Digital Streams Matrix—Pre-Ingest (Stage 8) (cont.)

Identify files that require access restrictions	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	
Flag files for further analysis	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Flag files for reversible changes (and fixes)	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	
Flag files for future reversible destructive changes																	
Flag files for future preservation action work	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Create provenance notes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Descriptive metadata creation, generation, extraction or capture	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Structural metadata generation	X	X	X	X		X	X	X		X	X	X	X	X	X	X	
Technical and preservation metadata generation or extraction	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Batch preconditioning	X	X	X		X		X	X	X	X		X	X	X	X	X	
Individual file preconditioning	X	X	X	X		X	X		X	X	X	X	X			X	
Generate access and/or delivery copies	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X

The DPOC's intended strategy for digital content management at Cambridge University Library is to embed governance, operational, and workflow processes that reflect the digital stewardship approach. This is being piloted as part of the Born-Digital Case Study.

Born-Digital Case Study

The Born-Digital Case Study is the first focused, in-depth work on born-digital personal and corporate records to take place at Cambridge University Library. Several previous digital transfer attempts have been undertaken (e.g., with born-digital university records), while research outputs (born-digital research publications, research data, and digital theses) are being submitted to Cambridge University Library's Apollo Open Access Repository.[76] For the most part, born-digital acquisitions arriving at Cambridge University Library remain on their physical format digital carriers (such as USB flash drives, optical media, etc.) at present.

The digital stewardship approach is a means of establishing requirements while demonstrating—to operational staff and senior management alike—suitable processes for handling born-digital personal and corporate records. The Born-Digital Case Study allows for closer inspection of selected workflow stages. For all three Cambridge University Library case studies, a subset of the workflow stages were selected (due to time restrictions). The Born-Digital Case Study project is defining "bare minimum" mandatory tasks at each stage (for each class and subclass of digital content), as well as documenting other optional tasks.

The content nominated for the Born-Digital Case Study is the digital content of a deceased estate containing personal records (e-mails and documents), research data (software code), published born-digital content (presentations and websites), and so on. As is typically the case with archival acquisitions (regardless of whether they are paper-based, digital, or a hybrid of both), negotiations take considerable time and are highly sensitive.

Since the Cambridge University Library is right at the beginning of implementing robust born-digital acquisition and preservation processes, it is an opportune time to focus on the earlier stages of the Digital Stewardship End-to-End Workflow Model.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Contextualizing the Born-Digital Case Study

Many leading GLAM institutions have already commenced the development of processes and workflows for acquiring and managing born-digital content. As a result of several projects—the PARADIGM project (2005–2007) and the future Arch project (2008–2012)[77]—the DPOC's partner organization, Bodleian Libraries, University of Oxford, was able to establish a born-digital manuscripts processing lab: the Bodleian Electronic Archives and Manuscripts (BEAM) service.[78]

The Born-Digital Case Study is informed by the author's firsthand involvement in establishing programs for the acquisition of born-digital content. In 2011, the National Library of Australia began conceptualizing and preparing to manage digital transfers of personal and organizational papers, and in 2012 their Digital Transfer Pilot project took place. Information about the approach and the challenges and lessons learned from this pilot were shared during the Born Digital Appraisal, Ingest, and Processing workshop at the 2014 iPres conference, alongside similar efforts from the Gates Archive and the National Library of New Zealand.[79] Work undertaken at the National Library of Australia, the National Film and Sound Archive of Australia, and the State Library of New South Wales in Australia, as well as information generously provided by staff from Archives New Zealand and the National Library of New Zealand, informs Cambridge University Library's Born-Digital Case Study.[80]

Pre-Ingest

A range of different functions for born-digital content (particularly for personal and corporate records) should be carried out at the earliest possible stage within the Digital Stewardship End-to-End Workflow. However, if this work isn't undertaken during the Conceive (Stage 1), Create (Stage 3), or Acquire (Stage 6) stages, then it is essential that it occurs during the Pre-Ingest (Stage 8) stage.[81] Ideally, these tasks should take place early on, with outputs only needing to be verified at subsequent stages.

While literature is beginning to appear that documents pre-ingest activities, there is no single definitive guide outlining all the possible tasks to be undertaken at this stage (or earlier). The Pre-Ingest (Stage 8) stage is critical, because digital content received from a donor or researcher is highly unlikely to come with all of the metadata required or to arrive as a Submission Information Package, or SIP (meeting an organization's specific SIP structure).

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

In order to build a SIP that may be successfully ingested into a digital preservation system or digital repository, pre-ingest work is required. Depending on the system, the number and types of files, and how they are arranged, pre-ingest work can range from being quick and easily automated to taking months for a staff member to manually process the digital content.

Most pre-ingest tasks are mandatory when processing born-digital personal and corporate records. A more limited set of tasks may be mandatory for trusted content creators, such as in-house digitization or in-house-created content. However, keep in mind that performing a thorough set of checks is likely to result in a higher degree of confidence in the digital content.[82] In order to guide pre-ingest work, a documented set of principles and actions (which should be outlined in an organization's policy, standards, guidelines, and procedures) is required. Mandatory and optional tasks should be clearly established for each class and subclass of digital content. Typical checks include format identification, characterization and validation, as well as virus-checking and checksum hash generation (if this hasn't already been performed as part of an acquisition process). In addition, structural information (such as a technical manifest) that includes full file names and file paths should also be produced.

As a means of further illustrating the degree of work that may be required for each task, two pre-ingest substages are discussed below.

Technical Analysis

In order to undertake any preconditioning activities, a technical analysis of a digital content must be performed.[83] This would include (but is not limited to) the following actions:

1. Generate or verify checksum hashes[84]
2. Confirm content renders[85]
3. Confirm that the content meets minimum quality standards or requirements
4. Generate a technical manifest (if one does not already exist), containing:
   - Structural metadata (including file paths and file names)
   - File format identification (such as PRONOM Unique Identifiers, or PUIDs)[86]
   - Date-time stamps (such as date last modified)
5. Undertake virus scan, ensuring no changes are made (and document results)

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

6. Identify unwanted or "junk" files (and document the decisions regarding the intended removal of these)

Underlying principles should form the basis for a technical analysis and further preconditioning work.[87] While each GLAM institution must develop its own set of principles (in line with the organization's context), it is the author's view that until files are fully ingested and managed by a digital preservation system, this digital content remains "at risk."[88]

For work carried out at the Pre-Ingest (Stage 8) stage (or earlier), the following principles are proposed:

- Nondestructive, reversible changes[89]
- Actions are undertaken in automated ways (or failing this, manually)
- Documentation of all actions and decisions is recorded[90]

Destructive or nonreversible changes should be avoided where possible, and should only occur as a preservation action (once the digital content is managed by a digital preservation system). This allows for files to be versioned and all changes recorded as PREMIS events.[91] While this is the view of the author, it is acknowledged that organizations may choose to take a different approach.

Preconditioning

Preconditioning work can be finicky and time-consuming. Moran and Gattuso's 2015 iPres paper "Beyond the Binary: Pre-Ingest Preservation of Metadata"[92] and Rosin's article "Applying Theoretical Archival Principles and Policies to Actual Born-Digital Collections" describe real-world examples that illustrate solid preconditioning efforts.[93]

The preconditioning work to be carried out should fall under the following categories, including (but not limited to):

Checksums

- Checksum hash verification (if preconditioning work does not take place immediately after a technical analysis)
- Document (and address) any issues

Metadata

- Ensure that the available structural metadata provides information on all folder structures (including full file names and file paths), and generates structural metadata if not available

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

- Capture or extract existing administrative and technical metadata (from file headers, sidecar files, or "associated materials")
- Generate any additional preservation and technical metadata that is required
- Document preconditioning actions as provenance notes
- Create, generate, or extract descriptive metadata

File names and file paths

- Address issues with "problematic" characters in file names (e.g., diacritics, characters not recognized by available character sets, etc.)
- Record "original" and "modified" file names and file paths

Files for acquisition decisions

- Inspect for viruses (using a virus scan report generated during the technical analysis)
- Files to remove, due to not meeting acquisition parameters[94]
- Remove unwanted files and "junk" files (operating system index and store files, software cache files, etc.)
- Document decisions related to files that are not selected (and reasons why)

Files requiring work

- Flag files for further analysis, including files that are not able to be identified, characterized, or validated
- Files that can be fixed using reversible changes (e.g., incorrect information in the file header, or in the wrong location)

Further analysis or preservation work

- Flag files that require preservation actions (once inside a digital preservation system)

Access restrictions

- Identify files that require various access restrictions to be applied
- Generate rights metadata

Access and/or delivery copies

- Generate access and/or delivery copies (as per the organization's policies, standards, guidelines, and procedures)

Much of this information should be included as part of the SIP metadata. It should be noted that digital content requires considerable preconditioning; this is labor-intensive and is, at best, no small feat to undertake. At the time of writing, no one tool is currently available to identify or address all of the preconditioning issues that are likely to be encountered in a single digital collection. Community effort is currently bridging this gap by developing python scripts.[95]

Conclusion

The digital stewardship approach, incorporating the Digital Stewardship End-to-End Workflow Model and the Digital Streams Matrix, provides GLAM and research practitioners who are working across a range of disciplines with further guidance for handling digital content. These models build on previous digital preservation and digital curation work, narrowing existing gaps. While there is no single approach to suit all scenarios, utilizing a holistic, cross-disciplinary strategy and methodology improves the management of digital content, and provides practitioners with a better understanding of the digital content in their custody. Undertaking preservation activities earlier in the life of digital content increases the chances of its availability over the long term. Underlying any model or preservation activity is the need for skill-sharing across disciplines, which is fundamental to the future success of the memory sector.

Digital preservation practitioners are well placed to advise on and advocate for the creation and acquisition of standardized, good-quality digital content and comprehensive related metadata. Meanwhile, archivists and curators can provide valuable experience and contextual insights. Reframing the long-term management and preservation of digital content as digital stewardship allows for a broader perspective, and a collaborative approach will ultimately benefit practitioners, users of digital content, and the longevity of our digital cultural heritage.

Acknowledgments and Dedication

The author would like to thank The Polonsky Foundation for its generous support in funding the DPOC project. Immense thanks are extended to former colleagues Emma Jolley, Douglas Elford, and Jonathan McCabe from the National Library of Australia, whose collaborative work efforts formed the

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

foundations for this chapter. Thanks to my Cambridge University Library Polonsky Fellow colleagues, David Gerrard and Lee Pretlove, for their involvement in developing the digital stewardship approach. Finally, thank you to Pascal Aeberhard for proofreading.

This chapter is dedicated to Elizabeth Caplice (https://skybetweenbranches.wordpress.com), who passed away in 2016 at the age of thirty-two. It is with you in mind, Liz, that this has been written.

NOTES

1. Apple considers any hardware older than five years after the end of the manufacturing date of a product to be "vintage." Replacement parts are no longer available (with some exceptions, for example, Turkey and the United States) for such products. "Vintage and Obsolete Products," Apple Inc., https://support.apple.com/en-gb/HT201624.
2. "Semantics: Digital Preservation vs. Digital Curation," Digital Curation Forum, https://groups.google.com/forum/#!topic/digital-curation/ehppkZT9XGs.
3. Jan Hutař, "Digital Preservation—From Theory to Practice?" Knihovna—Knihovnická revue, http://knihovnarevue-en.nkp.cz/archives/2015-2/library-and-information-at-home/digital-preservation-from-theory-to-practice.
4. Joan E. Beaudoin, "Context and Its Role in the Digital Preservation of Cultural Objects," *D-Lib Magazine* 18, no. 11/12 (2012), doi:10.1045/november2012-beaudoin1.
5. "Digital Preservation Briefing Paper," Joint Information Systems Committee, November 20, 2006, https://www.webarchive.org.uk/wayback/archive/20140614202005/www.jisc.ac.uk/publications/briefingpapers/2006/pub_digipreservationbp.aspx.
6. Archivematica, https://www.archivematica.org.
7. Preservica, https://preservica.com.
8. Keep Solutions, "RODA," https://www.keep.pt/en/produtos/roda.
9. Ex Libris, "Rosetta," https://www.exlibrisgroup.com/category/RosettaOverview.
10. DSpace, https://duraspace.org/dspace.
11. Fedora, https://getfedora.org.
12. Digital Preservation Coalition, "Preservation Actions," in *Digital Preservation Handbook,* 2nd ed., Digital Preservation Coalition, 2015, https://www.dpconline.org/handbook/organisational-activities/preservation-action.
13. Tony Knutson, "Filesystem Timestamps: What Makes Them Tick?" SANS Institute, https://sans.org/reading-room/whitepapers/forensics/filesystem-timestamps-tick-36842.
14. Ben Fino-Radin, "It Takes a Village to Save a Hard Drive" (blog), September 12, 2013, http://notepad.benfinoradin.info/2013/09/12/it-takes-a-village-to-save-a-hard-drive.
15. Dave Thompson, "Why Digital Preservation Is or Isn't Business as Usual," *Digital Preservation Coalition* (blog), February 17, 2017, https://www.dpconline.org/blog/why-digital-preservation-is-or-isn-t-business-as-usual.

16. "Sheer Curation," *UX Thesis* (blog), January 4, 2012, https:// web.archive.org/web/20160511181339/www.uxthesis.com/2012/sheer-curation.
17. Butch Lazorchak, "Digital Preservation, Digital Curation, Digital Stewardship: What's in (Some) Names?" *The Signal* (blog), August 23, 2011, https://blogs.loc.gov/thesignal/2011/08/digital-preservation-digital-curation-digital-stewardship-what%E2%80%99s-in-some-names.
18. "What Is Digital Curation?" Digital Curation Centre, www.dcc.ac.uk/digital-curation/what-digital-curation.
19. Jaime McCurry, "Digital Stewardship: The One with All the Definitions," *The Collation* (blog), April 2, 2014, https://collation.folger.edu/2014/04/digital-stewardship-the-one-with-all-the-definitions.
20. "National Digital Stewardship Alliance," Library of Congress, www.digitalpreservation.gov/ndsa/NDSAtoDLF.html.
21. In some cases, capturing the representation of the earliest copy of digital content may not be the desired goal. It may be the most recent version or a selection of versions that need to be retained; however, the same principles apply.
22. So as not to dilute the nature of the topic at hand, discussing maturity modeling is considered out of the scope of this chapter. Further information on maturity modeling is provided in a footnote later in this chapter, in the section titled "National Digital Stewardship Alliance Levels of Digital Preservation."
23. Bill LeFurgy, "Life Cycle Models for Digital Stewardship," *The Signal* (blog), February 21, 2012, https://blogs.loc.gov/thesignal/2012/02/life-cycle-models-for-digital-stewardship.
24. Brian Lavoie, "Meeting the Challenges of Digital Preservation: The OAIS Reference Model," Online Computer Library Center, https://www.oclc.org/research/publications/library/2000/lavoie-oais.html.
25. International Organization for Standardization, "ISO 14721:2012," September 2012, https://www.iso.org/standard/57284.html.
26. The OAIS Reference Model is based on the concept of "information packages." There are three main "states" an information package can take, in this reference model. These are the Submission Information Package (SIP) when files are being "submitted" or ingested into a digital preservation system or digital repository, an Archival Information Package (AIP) when the files are "archived" in the system or digital repository, and the Dissemination Information Package (DIP) when the files are being "disseminated" back out to end users.
27. To read more, see the Digital Preservation Coalition's 2014 Technology Watch report "The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)," https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file.
28. Susan Manus, "DPOE Interview with Austin Schulz of the Oregon State Archives," *The Signal* (blog), August 13, 2015, https://blogs.loc.gov/thesignal/category/dpoe-interview.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

29. To address insufficiencies in the OAIS Reference Model, this has led to the development of the Outer OAIS-Inner OAIS (OO-IO) Model. Eld Zierau, "OAIS and Distributed Digital Preservation in Practice: An exploration of Danish and other use cases that contributed to the development of the Outer OAIS-Inner OAIS Model for Distributed Digital Preservation," *Proceedings of the International Conference on Digital Preservation, Kyoto, Japan*, September 25-29 2017, https://ipres2017.jp/wp-content/uploads/14Eld-Zierau.pdf.
30. "DCC Curation Lifecycle Model," Digital Curation Centre, www.dcc.ac.uk/resources/curation-lifecycle-model.
31. Sarah Higgins, "Draft DCC Curation Lifecycle Model," *International Journal of Digital Curation* 2, no. 2 (2007), doi:10.2218/ijdc.v2i2.30.
32. Maureen Pennock, "Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information," *Library & Archives* no. 1 (2007), http://ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf.
33. Ibid.
34. Similar practices tend to occur around managing published born-digital and digitized content. Capturing and maintaining technical and preservation metadata as part of an acquisition or digitization process seem less of a concern when compared with managing born-digital personal and corporate records. This is not a view held by the author, but rather a reflection of current practices.
35. There are multiple terms used to describe this digital content, including "born-digital manuscripts," "born-digital unpublished materials," "born-digital personal and corporate papers," "born-digital papers and records," "born-digital special collections," and so on. For the purposes of this chapter, born-digital content of this type is referred to as "born-digital personal and corporate records." Given the physical connotations of terms such as "manuscripts" and "paper," these have been avoided.
36. The Digital Curation Centre is open to suggestions for improving the DCC Curation Lifecycle Model.
37. University of Bath, "Research360: Managing Data across the Institutional Research Lifecycle," 2011, www.dcc.ac.uk/sites/default/files/documents/IDCC11/photos/posters/Research360%20poster%20v3.pdf.
38. For further information about these other models, refer to the 2012 report "Review of Data Management Lifecycle Models," http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf.
39. "DPOE Baseline Digital Preservation Curriculum," Library of Congress, www.digitalpreservation.gov/education/curriculum.html.
40. "Tool Grid," Digital POWRR Project, http://digitalpowrr.niu.edu/digital-preservation-101/tool-grid.
41. The POWRR Tool Grid V2 (https://www.digipres.org/tools) is mapped against the DCC Curation Lifecycle Model stages.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

42. This includes details on whether each tool is open source and has clear documentation plus the cost.
43. AIMS Work Group, "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," January 2012, https://dcs.library.virginia.edu/files/2013/02/AIMS_final_text.pdf.
44. AIMS Work Group, "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," 2012, http://files.archivists.org/conference/dc2010/researchforum/MatienzoHandout.pdf.
45. Ibid.
46. Sarah J. A. Flynn, "The Records Continuum Model in Context and Its Implications for Archival Practice," *Journal of the Society of Archivists* 22, no. 1 (2001): 79–93, doi:10.1080/00379810120037522.
47. Wikipedia, "Records Continuum Model," https://en.wikipedia.org/wiki/Records_Continuum_Model.
48. "Levels of Digital Preservation," National Digital Stewardship Alliance, http://ndsa.org/activities/levels-of-digital-preservation.
49. It is important to note that there are a range of maturity models available for use in digital preservation, which can also inform the development of workflows for processing digital content. This includes Adrian Brown's Digital Preservation Maturity Model, found in his 2013 book *Practical Digital Preservation: A How-to Guide for Organizations of Any Size* (www.facetpublishing.co.uk/title.php?id=047555#.WhTCZoZpHeQ), Kenny and McGovern's Five Organizational Stages of Digital Preservation (https://quod.lib.umich.edu/cgi/t/text/text-idx?c=spobooks;idno=bbv9812.0001.001;rgn=div1;view=text;cc=spobooks;node=bbv9812.0001.001%3A11), and the NDSA's Levels of Digital Preservation (http://ndsa.org/activities/levels-of-digital-preservation). A further reference to digital preservation maturity modeling is Jefferson Bailey's blog post on "I Review 6 Digital Preservation Models So You Don't Have To" (www.jef fersonbailey.com/i-review-6-digital-preservation-models-so-you-dont-have-to). The Digital Curation Centre's Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) Roadmap Matrix (www.dcc.ac.uk/resources/tools/cardio) may also be of use in the GLAM sector. This is by no means an exhaustive list of maturity models that are in use, or that could be used to support digital archiving, digital curation, or digital preservation work. So as not to dilute the nature of the topic at hand, discussing maturity modeling is considered out of the scope of this chapter.
50. Areas of consideration represented in the NDSA Levels of Digital Preservation include: Storage and Geographic Location, File Fixity and Data Integrity, Information Security, and Metadata and File Formats. In 2016 it was suggested that "Access" should be added (https://blogs.loc.gov/thesignal/2016/04/expanding-ndsa-levels-of-preservation). Commencing in 2018, the NDSA Levels of Digital Preservation are undergoing a "reboot" (https://ndsa.org/working-groups/levels-of-preservation/).

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

51. Due to the need for brevity, a discussion of traditional archival theory and practice in any depth is excluded. The work presented in this chapter has been informed by archival processes and their application in operational contexts.
52. Archives Association of British Columbia, *A Manual for Small Archives* (Vancouver: Archives Association of British Columbia, 1999), http://aabc.ca/media/6069/manualforsmallarchives.pdf.
53. In recent years, the lack of standards and best practices for processing born-digital content has been identified. To address this gap, collaborative cross-institutional work is currently being carried out in the United States in order to develop a tiered approach to processing born-digital content (https://archives2017.sched.com/event/ABGm/201-what-we-talk-about-when-we-talk-about-processing-born-digital-building-a-framework-for-shared-practice).
54. Susan Thomas and Janette Martin, "Using the Papers of Contemporary British Politicians as a Testbed for the Preservation of Digital Personal Archives," *Journal of the Society of Archivists* 27, no. 1 (2006), doi:10.1080/00039810600691254.
55. Susan Thomas et al., "Workbook on Digital Private Papers," Personal Archives Accessible in DIGital Media (PARADIGM), October 1, 2008, www.paradigm.ac.uk/workbook.
56. Bronwen Sprout and Sarah Romkey, "A Persistent Digital Collections Strategy for UBC Library," *Proceedings of the Memory of the World in the Digital Age: Digitization and Preservation*, Vancouver, British Columbia, Canada, September 26–28, 2012, pp. 256–68.
57. The DPOC (https://www.dpoc.ac.uk) project is a two-year collaborative project between Bodleian Libraries, University of Oxford and Cambridge University Library, University of Cambridge, generously funded by The Polonsky Foundation.
58. This work was undertaken during the period from 2011 to 2016, and included the acquisition of born-digital content. The development of born-digital workflows and configuring ingest processes for a digital preservation system also informed this research.
59. Further work to develop procedures is required, so as to advise on exact steps for how tasks should be carried out, which software tools to use, and integration with Cambridge University Library's information and communications technology infrastructure. This should be positioned alongside an organization's strategy, policy, standards, and guidelines for digital content.
60. Nikolaos Lagos, Simon Waddington, and Jean-Yves Vion-Dury, "On the Preservation of Evolving Digital Content—The Continuum Approach and Relevant Metadata Models," presentation at the Metadata and Semantics Research conference, 2015, doi:10.1007/978-3-319-24129-6_2.
61. For paper-based archives, these are typically placed in compact shelving or stacks with natively managed and standards-based environmental and pest controls. Digital content must be taken care of in similarly controlled ways.
62. Collection management (including environmental management, security, storage, etc.) lies within the realm of traditional GLAM collection management responsibilities.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Due to the brevity required, this chapter does not discuss the affinities between digital stewardship, collection management, and traditional conservation and preservation practices. A discussion of access to digital content is also considered to be outside the scope of this chapter.

63. Elizabeth Rolando, Wendy Hagenmaier, and Susan Wells Parham, "Repurposing Archival Theory in the Practice of Data Curation," Georgia Institute of Technology, 2014, https://smartech.gatech.edu/bitstream/handle/1853/51321/IDCC14%2BPoster.pdf.
64. Jackie Dooley, "The Archival Advantage: Integrating Archival Expertise into Management of Born-Digital Library Materials," OCLC Research, 2015, https://www.oclc.org/research/publications/2015/oclcresearch-archival-advantage-2015.html.
65. While quality control (QC) is often found as a stage within a digital production workflow (or digital curation workflow), it is important to acknowledge that QC work should ideally be undertaken at each stage of the workflow. For this reason, where in other digital content workflows, a QC stage may appear, it has been intentionally omitted from the Digital Stewardship End-to-End Workflow Model. It is strongly recommended that QC work be embedded throughout the entire workflow process. The decision whether or not to undertake QC at each stage should be based on an organization's policies and "bare minimum" processes.
66. Monitoring and reporting are crucial throughout the entire workflow. Similar to QC checks, various monitoring and reporting should be established at selected stages. Again, a pragmatic approach that fits an organization's operational capabilities is necessary.
67. While the DPOE Baseline Digital Preservation Curriculum separates out "Protect" as an individual stage (Stage 4), it is the author's view that this should occur at every stage of a digital content processing workflow.
68. No individual practitioner working in the digital curation, digital preservation, or RDM fields will be able to handle every single scenario. For this reason, bringing together a number of different practitioners with complementary backgrounds, knowledge, and skill sets in order to collaboratively troubleshoot issues is necessary. Given the breadth and complexity of digital technologies, it is unreasonable to expect GLAM institution staff with traditional archival and librarianship training to rapidly acquire all of the skills and knowledge they will need in order to be able to fully manage digital content throughout an entire digital processing workflow. Likewise, staff with information and communications technology backgrounds are unlikely to have the requisite archival and preservation knowledge and skill sets such as donor negotiation, or be experts in the subject-matter domains of the content they are handling. Collaborative work and skill-sharing are crucial if robust born-digital programs are to succeed. One example of cross-disciplinary skill-sharing can be found at Purdue University: see Dearborn Carly, Amy Barton, and Neal Harmeyer, "The Purdue University Research Repository: Hubzero Customization for Dataset Publication and Digital Preservation," *OCLC Systems & Services: International Digital Library Perspectives* 30, no. 1: 15–27, doi:10.1108/OCLC-07-2013-0022.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

69. It is acknowledged that *manage* is not a particularly clear term, but this term is already in use across the digital preservation and digital curation disciplines.
70. The five initial classes were proposed by Grant Young and Jacky Cox in a 2014 Cambridge University Library internal briefing paper, "Towards a UL Strategy for the Acquisition and Preservation of Born-Digital Content," that was presented to the university librarian.
71. It should be noted that since there is a minimal amount of knowledge and skills related to identifying and handling audiovisual carriers, a decision was made by the author to include both analog and digital audiovisual carriers. This was due to staff being unable to differentiate between the analog and digital carriers. At the time of writing, there are no audiovisual specialist staff employed by Cambridge University Library, and there is only one staff member with audiovisual skills employed within the wider network of libraries across the University of Cambridge. Since there is no preservation strategy for audiovisual content at present, the audiovisual carriers—particularly video and audio—were deemed equally "at risk" as the digital content.
72. This includes supporting and meeting the compliance requirements of research funding bodies.
73. Somaya Langley, "Digital Streams Matrix," 2018, doi:10.17863/CAM.26363.
74. Due to the need for brevity, the granularity of documenting the various tasks associated with managing different classes of digital content is not indicated in this chapter.
75. Parameters included frequency and/or volume, significance, urgency, uniqueness, and value to users and/or stakeholders.
76. University of Cambridge, "Apollo" Open Access Repository, https://www.repository.cam.ac.uk.
77. Bodleian Libraries, University of Oxford, "futureArch project," https://www.bodleian.ox.ac.uk/beam/about/projects/futurearch-project.
78. The BEAM lab includes a Forensic Recovery of Evidence Device (FRED) workstation (https:// www.digitalintelligence.com/products/fred) and Forensic Toolkit (FTK) software (https://accessdata.com/products-services/forensic-toolkit-ftk) for use in processing born-digital content.
79. Jessica Moran and Leigh Rosin, "Born Digital Appraisal, Ingest, and Processing," *iPres 2014: Proceedings of the 11th International Conference on Digital Preservation*, State Library of Victoria, Melbourne, Australia, 2014, p. 305, https://ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf.
80. It should be noted that it is the view of the author that activities often placed under the banner of "digital forensics" should be considered part of everyday work when handling complex born-digital content or large born-digital acquisitions.
81. It should be noted that if tasks are only undertaken at the Pre-Ingest (Stage 8) stage, the loss or modification of data and metadata may have already taken place.
82. The most pragmatic way to currently perform necessary checks is by developing scripts, to automatically run across the content. Scripts should flag problematic content, producing meaningful reports so staff can effectively focus their work efforts.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

83. Similar to an appraisal of an archival collection to assess the content, a technical analysis is an appraisal of the digital content from a technical perspective, looking at the quality of the files as well as any apparent technical and preservation issues. A technical analysis should ensure that no changes to the files occurs. Ideally, files should be "read-only" when an analysis is undertaken. In some cases, this may take place only after a period of quarantine.
84. While there is no international digital preservation standard for checksum hashes, organizations typically use SHA-256, SHA-1, and MD5. Generating hashes, particularly on large files such as audiovisual content, can take a considerable amount of time. The audiovisual archiving industry tends to use the MD5 as the de facto standard. To ensure that no checksum collisions occur, some organizations opt for generating and verifying against two different checksum hashes.
85. In some cases, this will not be possible if you don't have the right software, hardware, or other equipment available, and if this is the case, this should be documented.
86. National Archives, "PRONOM," https://www.nationalarchives.gov.uk/PRONOM/Default.aspx.
87. Preconditioning means "to make changes to [a file] before it is ingested into the preservation system . . . [and to] . . . ensure that a version of transferred digital content will be able to be ingested into the digital preservation system without any issues or errors being presented by the system": see Peter McKinney and Euan Cochrane, "Digital Preservation Policy Manual," Archives New Zealand, National Library of New Zealand, Internal Affairs, 2012, http://ndha-wiki.natlib.govt.nz/assets/NDHA/About-Us/Strategic-Partnerships/Digital-Preservation-Policy-Manual.pdf.
88. Once the custody of digital content is transferred to an organization, such as during the Acquire (Stage 6), Arrange and Describe (Stage 7), Pre-Ingest (Stage 8), and Ingest (Stage 9) stages, the digital content must also be secured and managed.
89. The author's view on preconditioning principles has been heavily informed by Archives New Zealand and the National Library of New Zealand's Digital Content Preconditioning Policy, which is included in their "Digital Preservation Policy Manual."
90. This documentation should be able to be accessed at any point while the digital content is in the custody of a GLAM or research institution.
91. Library of Congress, "PREMIS," https://www.loc.gov/standards/premis.
92. Jessica Moran and Jay Gattuso, "Beyond the Binary: Pre-Ingest Preservation of Metadata," *iPres 2015: Proceedings of the 12th International Conference on Digital Preservation*, Chapel Hill, North Carolina, November 2–6, 2015, pp. 137–43, https://phaidra.univie.ac.at/view/o:429524.
93. Leigh Rosin, "Applying Theoretical Archival Principles and Policies to Actual Born-Digital Collections," Archive Journal, November 2014, https://www.archivejournal.net/notes/applying-theoretical-archival-principles-and-policies-to-actual-born-digital-collections.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

94. Deselection during the preconditioning stage may occur for a variety of reasons, including not meeting the collection development policy, not meeting minimum quality requirements, and so on.
95. One real-world example is Ross Spencer's "Heroes or Villains" python tool, which can be used to process DROID reports and produce a shortlist of "rogue" files for further investigation (http://openpreservation.org/blog/2015/08/25/hero-or-villain-a-tool-to-create-a-digital-preservation-rogues-gallery).