

# Electronic Records Task Force Final Report

## *Documentation of the University of Minnesota Electronic Records Task Force's Initial Year of Activities*

September 11, 2015

### *Task Force Sponsors*

John Butler, Associate University Librarian for Data & Technology; Kris Kiesling, Elmer L. Andersen Director of Archives and Special Collections

### *Task Force Members*

Carol Kussmann (co-chair), Arvid Nelsen (co-chair), Lisa Calahan, Kevin Dyke, Lara Friedman-Shedlov, Lisa Johnston, Mary Miller, Erik Moore, Jon Nichols, Justin Schell, and Mike Sutliff

### Abstract:

The University of Minnesota Libraries sponsored the Electronic Records Task Force to explore and develop capacity of the Libraries to preserve and provide access to electronic records. Focusing on the records in the Archives and Special Collections, the Task Force was charged with developing initial capacity for ingesting and processing electronic records; defining tasks and workflows for staff; and developing ingest and processing workflows. This report documents the work of the Task Force during its first year and recommendations for moving forward.

Suggested Citation: University of Minnesota Libraries. (2015). Electronic Records Task Force Final Report. Retrieved from the University of Minnesota Digital Conservancy, <http://hdl.handle.net/11299/174097>.

## Table of Contents

[Executive Summary](#)

[Summary of Task Force Activities](#)

[Detailed Description of Task Force Activities](#)

[Phase I Activities](#)

[Develop Initial Capacity for Electronic Records Ingest and Processing](#)

[Education and Knowledge Capacity Building](#)

[Main Workstation \(Windows\)](#)

[Macintosh Workstation](#)

[Ubuntu Operating System \(Windows Workstation\)](#)

[Other Equipment](#)

[Software Programs](#)

[Security](#)

[Digital Storage / Working Space](#)

[Phase II and III Activities](#)

[Define Tasks and Workflow for Staff and Develop Ingest and Processing Workflows](#)

[Selection of Task Force Members to Form Workstation Working Group](#)

[Minimal Task Selection](#)

[BitCurator Software Exploration](#)

[Reviewing Tools / Choosing Tools for Initial Workflow](#)

[Testing of Initial Tools](#)

[Drafting of Workflows](#)

[Modifications of Workflows](#)

[Summary of Ingest Work Completed](#)

[Documents Supporting Workflows and Tasks](#)

[Development of Donor Guides and Transfer Information Sheet](#)

[Development of Policies/Procedures/Guides](#)

[Drafting of Addendum to Deed of Gift](#)

[Determine Short and Long-Term Human Resource Requirements for Performance of Technical Work](#)

[Communications and Outreach](#)

[Survey Working Group](#)

[Contact with Archives and Special Collections Staff](#)

[Working with DMCI](#)

[Presentations](#)

[Conclusion](#)

[Next Steps](#)

[Ongoing Development](#)

[Production Environment for Ingest](#)

[What We Need to Get This Done](#)

- [Appendix 1: Charter for Electronic Records Task Force](#)
- [Appendix 2: Hardware and Software](#)
- [Appendix 3: Workflow Images](#)
- [Appendix 4: Master Processing/Ingest Workflow Instructions](#)
- [Appendix 5: Policies](#)
- [Appendix 6: Accession Log and Instructions](#)
- [Appendix 7: Three Phase Work Plan](#)
- [Appendix 8: Donor Guides](#)
- [Appendix 9: E-Records Transfer Sheet](#)
- [Appendix 10: Deed of Gift Addendum for Electronic Records](#)

## Executive Summary

One of the Libraries' roles is to collect, preserve, and provide access to materials of interest whether licensed or owned, broadly published or rare or unique. The materials for which the University Libraries is responsible has increasingly included electronic records -- unique unpublished materials coming into the Libraries' stewardship in a variety of electronic formats and on various media. Electronic records are routinely integrated into our daily lives and the Libraries must be able to support the ingest, curation, and preservation of such materials for the goal of long-term access and usability. Addressing the needs of these materials has most often been the responsibility of the Libraries' Archives and Special Collections (ASC) Department, to which materials are frequently entrusted, often (but not always) as donations.

The Electronic Records Task Force (ERTF) was initiated to explore the first steps of working with and caring for electronic materials. First steps included developing staff expertise, building technological capacity to ingest and process electronic materials, establishing an initial workflow for and policies related to the ingest process, and exploring the roles and responsibilities of staff who bring in and/or process electronic records. As will be revealed in the report, the group's thinking about workflows, policies, and procedures evolved significantly over the course of the year.

The Task Force, along with Libraries Data & Technology Division staff, built a digital processing workstation based on recommendations made by the Task Force and, as the project progressed, increased physical space and computational capacity as experience demonstrated needs and identified tools. The initial workstation consisted of a Windows machine with hardware and software specifically designed to ingest and process electronic materials according to standards developed by the archival community and digital forensics specialists. A Macintosh workstation and additional external write-blockers and hard drives were added to meet new needs. Drives for accessing legacy media (5 ¼ and 3 ½ floppy disks, Zip disks) had been purchased by ASC previously and/or were salvaged from storage rooms in order to read, ingest, and care for both new and existing-collections.

This workstation provided a clean, secure computer environment to work with electronic materials that is required to assure the authenticity of documents by minimizing the alterations and corruptions that normally occur in the daily, non-forensic processes of viewing and moving files. Using this workstation, a workflow was developed to address initial ingest processes in a consistent manner. Prior to developing the ingest workflow, Task Force members discussed goals and requirements and reviewed best practices and standards. A sound ingest process requires understanding the original storage media, determining the method of transfer best suited to the media and file types, having a secure storage location within the Libraries, and running multiple programs against source and destination records for quality control and to establish a preservation baseline. Electronic records acquisition and preservation also require

refined selection decisions that establish priority levels; these are among the recommended tasks for moving forward as they were not part of this charter. The Conclusion and Next Steps sections of this report provide additional details.

The Task Force developed policies and procedures to guide electronic records ingest. These will evolve over time as new software and new document types emerge. Consistency in applying electronic record protocols requires familiarity with technical processes, which is gained and retained by regular practice and growth of expertise through experience. This will need to be an ongoing commitment by the Libraries.

Staff roles were explored as workflows were developed. After several modifications to the workflow and associated role assignments, it became clear that electronic records ingest and processing is *not* a task that is best distributed to multiple staff. Just as the current Central Processing function in ASC provides consistent oversight in working with paper records, a similar function needs to be created for electronic records.

### **Recommendations**

- Establish a “phase 2” task force to develop policies and procedures for the appraisal, description, and preservation of, and access to, electronic records.
- Carry over funds remaining from “phase 1” of the Task Force to support potential hardware, software, and other related needs of the “phase 2” effort. (~6K)
- Staffing -- hire immediately a half-time electronic records archivist to support current ingest processes. By mid calendar 2016, have a full-time electronic records archivist in place to handle ingest and processing of electronic records (This supersedes the half-time position. If feasible, recruit for the full-time position immediately.) Such a position will require a new staff line.

## Summary of Task Force Activities

The Electronic Records Task Force was officially chartered and launched in May 2014 to address the immediate need to safely and securely acquire, transfer and stabilize unique electronic records, in a multitude of file formats and from a host of contemporary and legacy media. The Task Force members were: Lisa Calahan, Kevin Dyke, Lara Friedman-Shedlov, Carol Kussmann (co-chair), Mary Miller, Erik Moore, Arvid Nelsen (co-chair), and Justin Schell. Resource personnel consisted of Lisa Johnston, Jon Nichols, and Mike Sutliff. John Butler and Kris Kiesling were the Task Force sponsors. ([Appendix 1](#))

As a whole, the Task Force met monthly or bi-monthly. Working groups formed to address specific issues met more frequently. Working group members additionally consulted with colleagues outside of the Task Force, most often but not exclusively Archives and Special Collections (ASC) staff, to discuss unit concerns pertaining to electronic records in their care. The Task Force co-chairs met with the sponsors monthly or bi-monthly.

In order to achieve accomplishable goals within the one-year duration of the charter and to address immediate and emergent needs, specifically the acquisition, transfer, and stabilization of electronic records on multiple media, some of which is volatile<sup>1</sup>, the charter divided the work of the Task Force into three phases. Phase I pertained to building initial capacity for the ingest and processing of electronic records. This consisted of the identification and acquisition of necessary hardware and software, and included some specialized training. This task was addressed discretely and completed early on, with modifications made over time. Phases II and III addressed the development of tasks for staff and workflows for processing, respectively. The phases were iterative and highly interdependent, thus the Task Force engaged in them simultaneously and continuously throughout the year. Workflows, policies, and procedures were developed but the group expects that these will require continual evolution. Each electronic record set consists of multiple and varying formats and every set differs from every other. Some file formats require special handling and while record sets typically contain many “standard” formats, it is not uncommon for new formats to be discovered during ingest. These may require – and have required – the use of tools and the development of processes specific to those formats. Older collections contain records in legacy formats that may be unfamiliar to us today. Meanwhile, computer technologies continue to change, promising the ongoing evolution of today’s “standard” formats and the creation of entirely new ones. Electronic records are an ever changing and moving target, but each encounter with a new record set adds to our experience and capabilities.

The Task Force completed basic deliverables for Phases II and III, specifically in respect to a secure ingest process and stabilization of records. More work is needed to inform our

---

<sup>1</sup> In this context, “volatile” refers to media at risk of losing its integrity, thus putting content at risk to loss or degradation.

understanding of what records within a set can and should be retained and which can and should be discarded, as well as which staff members are best able to make that determination. For example, it is not uncommon for electronic record sets to contain many duplicate records – actual and apparent. In some cases apparent duplicates may contain small but meaningful differences that demonstrate the evolution of a document. This may be highly significant in the papers of a famous author, though in other cases it may be irrelevant. Questions pertaining to retention and disposition of records include:

- When might apparent duplicates be significant?
- To what extent do processing time requirements and institutional storage capacities determine if or when extensive weeding should take place?
- How and when are retained files “migrated” – that is transformed to contemporary formats as technology continues to evolve, thus permitting ongoing access to their content?
- How do such conversions affect a student or researcher’s experience and understanding of the file?
- Are unit curators or central processing staff best able to make these assessments?

The answers to these questions will have critical impact on the further formation of University Libraries policies and/or determination of required human resources.

The Task Force set out to create a procedure for the initial transfer, or ingest of materials. This is a complex technical process requiring familiarity with special hardware and software as well as an understanding of what each software tool does. The Task Force developed these processes in the course of working with materials, adapting known best practices to our local needs and capabilities. The group originally envisioned that unit staff would complete post-ingest processing of their records. However, feedback and experience received after working with ASC staff one-on-one confirmed what the group already suspected to be true -- ASC staff do not have a sufficient amount of time to review electronic records at the necessary level of detail; they would prefer assistance -- similar to what now happens with the paper records as they move through Central Processing. Currently the Libraries do not have staff available to do this type of work with electronic records. The job of securely ingesting electronic records and preparing them for long-term preservation and access to students, faculty, and all of the communities that the University Libraries serve is a job that requires specialization. The successful steward of electronic records needs considerable basic technical training, the capacity to continually develop knowledge, and the focus to regularly and consistently exercise skills while developing local practice along the way. Staff who are able only intermittently to work with electronic records need to re-learn software and workflows and cannot keep up with new developments, which impedes the process.

Moving forward, a new Task Force needs to develop workflows and staff responsibilities for making decisions regarding selection, arrangement, and levels of access. This group also needs to work with Libraries staff towards a preservation repository to address long-term

preservation needs. As these things are done the Task Force expects that different types of storage media and files will be encountered, requiring new processes to be developed.

## Task Force Charge: Phase Descriptions and Related Deliverables

### **Phase I: Develop Initial Capacity for Electronic Records Ingest and Processing**

- Identify initial hardware and software needed for secure ingest and processing of electronic records.
- Procure initial hardware and software needed for secure ingest and processing of electronic records.
- Implement initial hardware and software needed for secure ingest and processing of electronic records.

**Deliverable:** *Acquisition and deployment of workstation and related hardware and software ([Appendix 2](#)) based on earlier recommendations for technical infrastructure / hardware to support policies and procedures.*

### **Phase II: Define Tasks and Workflow for Staff**

- Define tasks to be performed by staff engaged in technical functions (ingest and processing)
  - Assign role of ingest and processing to select members of the Task Force initially
- Define tasks to be performed by staff engaged in curatorial/selection functions.
  - Understand what information needs to be gathered and conveyed to workgroup by curators
  - Decisions to be made regarding selection, arrangement, levels of access, etc..
  - Information that can be conveyed to donors.
- Determine short and long-term human resource requirements for performance of technical work.

**Deliverable:** *Define tasks to be performed by technical and curatorial staff ([Appendix 3](#), roles drafted and redefined in Figures A5 and A6). Make recommendations for short and long-term personnel to support the policies and procedures (described in this report).*

### **Phase III: Develop Ingest and Processing Workflows**

- Develop workflows to support secure ingest of electronic records and the creation of Submission Information Packets (SIPs) and Archival Information Packets (AIPs)
- Survey electronic record sets in our holdings (focus on new acquisitions)
  - Develop instructions/recommendations on how to complete an inventory of electronic records



- Select sample electronic record sets with which to test workflows exhibiting best practices
- Establish institutional workflows by modifying test workflows according to lessons learned in testing, institutional priorities and capabilities.

**Deliverables:** Workflows ([Appendix 3](#)), procedures ([Appendix 4](#)), and policies ([Appendix 5](#)) as required, for accessioning and ingesting electronic records. Accessioning log and instructions ([Appendix 6](#)) were created as a method for inventorying collections.

## Detailed Description of Task Force Activities

The following section documents the intricacies of how the Task Force addressed the charge, worked through challenges, accomplished tasks, and began to explore next steps that include addressing new questions.

### Phase I Activities

#### *Develop Initial Capacity for Electronic Records Ingest and Processing*

Building initial capacity for ingest and processing required procurement of physical machines and equipment, software programs, local and remote storage, and physical and virtual working spaces. An equipment list ([Appendix 2](#)) provides documents and statistics about current physical capacity investments. Developing capacity also included an education component for Task Force members. Details about how the physical and educational capacity was built are provided below.

#### **Education and Knowledge Capacity Building**

Prior to the establishment of this Task Force, ASC staff and others had engaged in a year-long initiative to provide overarching education in the basics of born-digital archival curation and management. This comprised a series of classes offered by the Society of American Archivists (SAA) Digital Archives Specialist (DAS) program, sponsored by the University Libraries in 2013.

In fall 2013, in response to a call for volunteers by members of its development team, the group also participated in pilot testing of the first official release of BitCurator, a digital forensics ingest and processing tool. During Phase I, the Task Force determined the need for specific, additional higher-level training in the area of digital forensics, directed primarily toward members of the Task Force, although opportunities were open to others as well. In June 2014, the University Libraries hosted another week of courses co-sponsored by SAA and the Association of Research Libraries (ARL), including a two-day Digital Forensics course focusing heavily on BitCurator, which several Task Force members attended.

Understanding that BitCurator could play an important role in ingesting and processing electronic materials, the Task Force took advantage of an opportunity for individualized BitCurator training and submitted a successful application for an on-site BitCurator workshop offered by the BitCurator development team. In July 2014, Porter Olsen, a BitCurator developer, came to visit for a two-day, small group, hands-on workshop that allowed the ERTF to experiment with the tool in the context of the materials in our holdings, and allowed for the exploration and evaluation of digital forensic workflows and the BitCurator tool itself. Additionally the group invited representatives from various local repositories to attend a public presentation given by the developer. This experience greatly enhanced the group's knowledge of the software and its functionality.

Throughout the year the group took advantage of other educational opportunities, including conferences and webinars on topics such as born-digital programs, tools, and standards.

### **Main Workstation (Windows)**

In November 2013, the informal ASC electronic records working group drafted a Born Digital Workstation Requirements list based on research and input from a variety of sources. Review of these requirements led to the suggestion of and specifications for a digital records processing workstation.

The ERTF worked with resource member Mike Sutliff to discuss practical options for building a workstation with the Libraries' specific needs in mind. Following the specifications outlined by the ASC electronic records working group, Mike purchased the specified computer, write blockers, extra internal hard drive, and additional equipment and built the workstation himself. The workstation was set up in the Charles Babbage Institute (CBI) for use by all ERTF members.

### **Macintosh Workstation**

After the Windows based workstation was up and running, it became clear that a more direct way to work with Macintosh files and Macintosh hard drives was needed. Many of the materials being reviewed were created on or transferred to external media by Macintosh computers, which increased the time and effort required to transfer and ingest materials using the Windows workstation. The Task Force discussed possible use cases and requirements for a Macintosh computer workstation with Mike, who came up with a proposal for a machine. This workstation was to be simple; something that provided a means to view and transfer files from incoming Macintosh media. Most of the processing itself would still be done on the Windows workstation. While the cost was low to purchase a new Macintosh machine (around \$500.00 plus the cost of a monitor) the Task Force was able to acquire a Macintosh computer that was previously used in the Libraries. Because of this, if the machine breaks down, it will not automatically be replaced. However, the group felt that having an older machine was an advantage when working with older digital materials as many older drives would not connect directly to a newer machine, requiring the purchase of additional components.

### **Ubuntu Operating System (Windows Workstation)**

Possible workflows for processing electronic materials include using the BitCurator software package. This software can be run in a virtual machine environment or installed locally using the Ubuntu operating system. Earlier testing and experimentation with the virtual machine environment led us to believe using the software in Ubuntu would be the better option.

With assistance from Mike again, the 500GB internal hard drive on the Windows workstation was partitioned, giving ½ of the space to the Windows operating system and ½ to the Ubuntu

operating system. This allows us to run BitCurator in its native environment. More information on BitCurator software can be found below in the BitCurator Software Exploration section.

### **Other Equipment**

While working with the initial setup, the group also realized that there were times when an external write blocker rather than the one built into the Windows workstation tower would be the most appropriate tool to use. Two external write blockers were purchased to fulfill this need. One write blocker works with USB devices while the other can connect to a wide variety of older media. In addition to assisting in implementing best practices for exploring content on electronic media, these write blockers allow, on an as needed basis, to safely and securely provide users access to content knowing that the records on devices connected through the write blocker can't be modified, only viewed.

### **Software Programs**

Many of the tools the Task Force utilized were free computer programs. However there are a few that the group decided to purchase that are now part of the main workflow. Quick View Plus, TeraCopy, and Data Discovery (Identity Finder) assist with viewing files for which the Libraries may not have the native program, moving/copying files with verification of the process, and identifying private information. Additional software programs that may assist in the ingest and accessioning of files are added to a list of tools as they are identified. Identifying new tools is an ongoing process as new needs or issues arise and new tools are developed.

### **Security**

Protecting digital assets is a high priority for the Libraries and the University. Guided by the University's Information Security Framework, a multifaceted approach to securing digital assets includes:

- Physical security for the storage media temporarily in custody of the Libraries for accessioning purposes
- Physical security for storage media owned by the Libraries
- Security of the digital files while being accessioned and processed
- Security of the servers where digital files are permanently retained
- Identifying and determining disposition of digital files that contain Private Data, as classified by the University in accordance with policy and law
- Permissions, for both staff and researchers

To address the physical security of media in the Libraries' custody while digital files are being ingested, the Task Force identified the need for a locking cabinet and sought to purchase one with the funds provided. An appropriate cabinet was not found, but the search continues. Currently after ingest, any physical media that is owned by the Libraries will be stored in a secure location following typical ASC collection storage security protocols.

Access to the digital files on storage media or servers during accessioning is limited to specific staff, usually the curatorial staff for the specific collection and ERTF members or electronic

records staff who are carrying out the ingest. It is important to ensure that files are not inadvertently changed or deleted, so staff trained in using appropriate software should be responsible for the ingest process. The content of some digital files poses a potential risk, including financial records, Social Security numbers, health information, and other information considered to be “Private-Highly Restricted” or “Private-Restricted” data by the University’s Data Security Classification<sup>2</sup> policy.

The Libraries has developed a strong working relationship with University Information Security Services (UIS), which is addressing risk management across the University. The University’s new Information Security Framework developed by UIS provides units (e.g., the Libraries) with the responsibility of assessing and managing the risks it assumes. Within this approach, the Libraries will follow established security standards to mitigate risk or accept some level of risk based on the potential benefit to business needs. UIS has developed a basic understanding of the Libraries’ data issues and those surrounding electronic records. Due to the potential risk of private data being present at ingest, the Task Force adjusted its workflow to keep such files from being transferred to a network drive. The Office of Information Technology (OIT) is planning to offer secure storage services that may permit transfer of these files to a network drive in the future, a development that will bear watching.

### **Digital Storage / Working Space**

The workstation computer was equipped with an extra internal 4TB hard drive for additional storage space to be used during the ingest and processing of electronic records. This is a non-networked location accessible only from the workstation computer, referred to as the “Data Drive.”

In addition to limiting the number of people with physical access to the workstation and the contents on its internal drive, this non-networked storage location provides a secure space in which to search for private and sensitive information not allowed on University network drives.

Initially it was hoped that ASC curatorial staff would be directly involved in appraising electronic records after ingest occurred. However, with access to the files limited to one workstation and due to curatorial time constraints, this proved unworkable. To address the workstation access issue, the Task Force worked with resource personnel Jon Nichols and Mike Sutliff to secure 10 TB of network storage space, referred to as the “Q Drive,” which would allow curatorial staff to work with pre-processed records from their own computers on their own time. The drive is organized by ASC unit in order to support better integration into current workflows and assist with security controls. Access to this additional 10 TB of space is controlled by X500 user log in. Members of the Task Force have access to the entire drive, while specific ASC staff have access only to their unit’s collections. Limiting access to the Q Drive increases the security of the files while they are being reviewed.

---

<sup>2</sup> <http://policy.umn.edu/it/dataclassification>

Working with both the Data Drive and Q Drive proved to be challenging. The Workstation Working Group needed to create clear and standardized workflows so that pertinent staff would know where to put data, when to move data, what data could or could not be in a particular location, and how the data was being backed up. There is significant room for improvement in this process, as it does not address long term preservation problems, especially when it comes to ensuring that electronic records are being backed up.

In addition to computer and network storage space, the Task Force purchased two 2TB external hard drives to use for transferring files from donors. One of these drives is Macintosh formatted, the other is Windows. These drives have also provided the means to provide access to files for researchers on an as needed basis, as other access mechanisms are not yet available.

## Phase II and III Activities

### *Define Tasks and Workflow for Staff and Develop Ingest and Processing Workflows*

While defining tasks and developing workflows for ingest and processing electronic records is and will continue to be an ongoing activity, Phases II and III allowed the Task Force to develop a foundation on which to build future work. Workflows and processes will change over time as tools and technology used to address electronic records ingest and processing change. The following explorations were completed in relation to these two phases this past year.

#### **Selection of Task Force Members to Form Workstation Working Group**

As the work of Phase II and III would require more hands-on work with the workstation and an additional commitment of time, the Workstation Working Group was formed. This group included members of the Task Force who were most closely involved with the accessioning of materials in the Archives and Special Collections units: Lisa Calahan, Lara Friedman-Shedlov, Carol Kussmann, Erik Moore, and Arvid Nelsen. The Workstation Working Group developed a three phase work plan ([Appendix 7](#)) and met as necessary to address tool selection and workflow drafts as well as work through questions and issues that arose through testing programs or while ingesting collections. Group members also worked together as necessary when transferring and processing sets of electronic records.

#### **Minimal Task Selection**

There are many potential steps and procedures that can and arguably should be undertaken to ingest born digital material according to the best practices as currently established, and they continue to evolve in this emerging field. Many practitioners feel overwhelmed to the point of paralysis. Rather than let the perfect be the enemy of the good, the Task Force's strategy was

to start by identifying a basic, conceptual workflow and a set of tasks that the group would regard as the minimal level of intervention required to get started with ingesting materials.

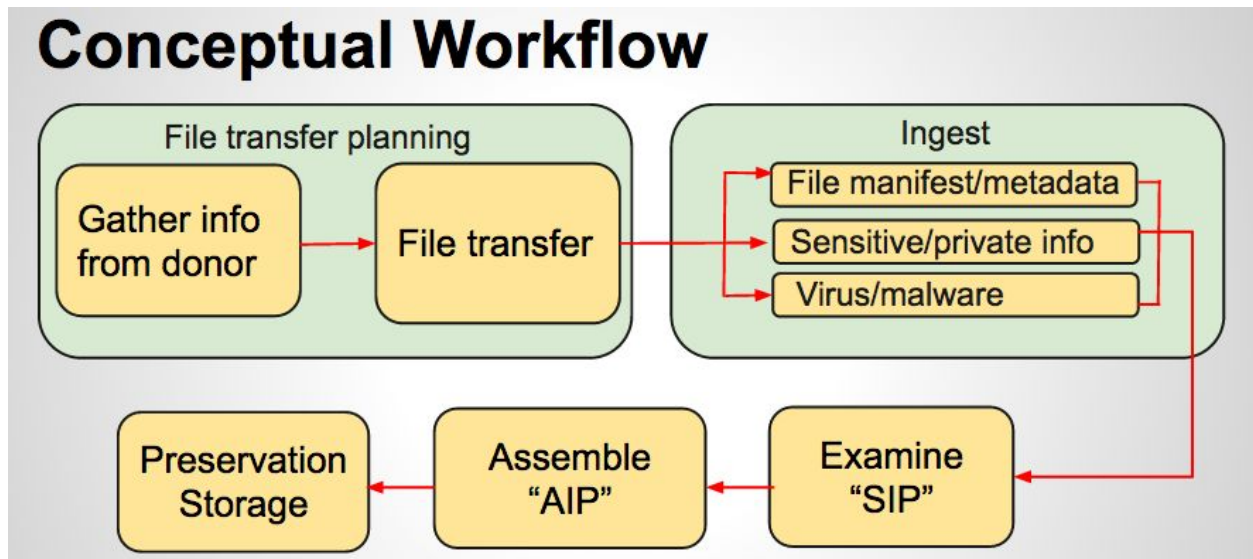


Figure 1: Conceptual Workflow

The Conceptual Workflow (Figure 1) builds on more complex models like the Open Archival Information System (OAIS), but provides a simplified overview of the most critical parts of the process from the Task Force's perspective, beginning with planning for file transfer, followed by the actual ingest of records and creation of a Submission Information Packet (SIP), and finally the transformation of the SIP to and Archival Information Packet (AIP) for long-term storage. The Conceptual Workflow defines a minimal set of ingest tasks and provides a framework for identifying a basic set of tools that could be used to accomplish these tasks.

The goals established for the ingest process were:

- Transfer of files to the workspace verified by checksums
- Generate a file manifest that lists filename and path, file size, and file type/format
- Identify potentially sensitive electronic information (SEI) and personally identifiable information (PII)
- Identify duplicate files

This information, along with the files themselves, would become the SIP that could then be reviewed by the curating staff member to determine which files should remain part of the AIP and whether any reorganization or redaction of the files would be necessary.

### BitCurator Software Exploration

As discussed in the Education and Knowledge Capacity Building section of this report, Task Force members were familiar with BitCurator software as a result of various trainings. BitCurator is a set of open source forensic tools designed to be used by archivists processing

born digital material. It is used to create and work with a disk image, allowing the archivist to make a complete and authentic copy of a piece of media, explore and analyze its contents, identify duplicate and/or sensitive data, and export data as needed to create an Archival Information Package (AIP).

When working with BitCurator, the process of creating a disk image is the first step. This was documented by the Task Force, along with instructions on how to analyze the contents of the disk image. A disk image is a bit-for-bit copy of a single piece of media. For example, a full 256MB flash drive is captured as a 256MB file, while a 2TB drive with 1TB of data on it is captured as a 2TB file - every bit, used or not, is captured. Not only can this take up valuable space, but the time it takes to image a large drive is often prohibitive. Various tools within the BitCurator environment are used to explore and identify files and content on a disk image. One thing to note is that disk images may include files thought to be deleted from the drives. This could be used to recover accidentally deleted files, but it could also uncover files never meant to be transferred to the archives. If disk images become part of a standard workflow, a policy must be written to address discovery and/or recovery of deleted files.

While disk images provide a complete picture of the media it represents; not all media require the time and effort to do so. We want to transfer only materials associated with the collection, not extraneous files that are not deemed archival. The best use case for BitCurator may be when working with floppy disks where the content is at risk of not being immediately readable because the files were created with older software. The resulting disk image allows staff to review a bit-for-bit copy while protecting the content on the original media.

It is important to note the challenges of working with software that is still in active development. BitCurator was developed by the archival community as part of a grant-funded multi-year (2011-2014) project which is now managed by the BitCurator Consortium<sup>3</sup>. New releases of the software are pushed out as bugs are fixed or new features become available; at least eight updates were made available between July and September 2014 leading up to the Version 1.0 release in late September. The most recent version (V 1.3.5) was released in May 2015.

In addition to the frequency and unscheduled timing of releases, one of the most challenging aspects of the software is that updates currently require full reinstallation of the Ubuntu operating system on the workstation. Since the installation of updates requires completely rewriting the Ubuntu section of the hard drive, any files we wish to retain must first be copied elsewhere, since they would otherwise be erased. The complexity of this process initially required the Task Force to coordinate schedules with IT, but after multiple reinstallations, IT staff worked with the Task Force to document the procedure so subsequent versions could be installed by Task Force members. However, more often than not, new installations cause something to not work the same way as it did previously or other things to work not at all. Time is required to determine if that is because of a setting in the workstation environment, or if there

---

<sup>3</sup> <http://www.bitcurator.net/bitcurator-consortium/>



is something wrong with the release. Many others in the BitCurator user community share our frustration with having to completely reinstall BitCurator rather than just update it or reinstall a certain program within it and have voiced their concerns to the development team, so we are hopeful this problem will be addressed in future releases.

During this past year the Task Force explored BitCurator's features and capabilities but did not put it into use with any specific accessions. Features tested and issues encountered were documented. The software should be included in the Libraries' toolkit and explored as a potential part of workflow under certain circumstances; creating a disk image would be particularly useful in cases where a collection is intended to provide a comprehensive picture of the original working environment of the donor or when transferring materials off legacy materials such as floppy disks.

Moving forward, it is expected that members of the Task Force who are part of the BitCurator Google/Users Group will share applicable news or software development with the rest of the Task Force, which may assist workflow development. For example, the current focus of BitCurator developers is to provide web-based access to disk images as well as other methods to make the content on disk images more accessible to a larger community. Developments in this area may increase the use of the tool.

### **Reviewing Tools / Choosing Tools for Initial Workflow**

The first objective for the Workstation Working Group was to review software tools and programs that could assist in performing necessary tasks such as transferring files, creating checksums, finding duplicates, capturing technical metadata (date created, date last modified, file type, file size, etc), and understanding file structure/organization. Software tools selected for testing were identified from suggestions made during SAA Digital Archives Specialist courses, various webinars and presentations, research on what other repositories are using, and general Internet searches. To accurately document the results of the testing process, identify the pros and cons of each tool, and document outputs and reporting features, the group created a "Tools to Test" spreadsheet.

A template was created to document the group's experiences for each tool tested<sup>4</sup>. Each person who tested a tool used this template to document their individual experience in order to capture different points of view for practical application. By the end of the testing phase, eighteen experiences were documented for fifteen different applications.

Based on early testing, the Workstation Working Group reviewed the list of possible tools and decided which tools to test in initial workflows. The group determined which applications, when used together, would accomplish a minimal set of requirements for ingesting records, and would

---

<sup>4</sup> This template documented the name of the tool, the date and the name of the person testing the program, the testing goal, the material the tool was being tested on, what steps were followed, the observation, and any questions or comments that came up as part of the test.

be simple enough for any staff member with appropriate training to utilize. Using these parameters, it was determined that Data Accessioner would be used to transfer files and HashMyFiles would be used to capture multiple checksum values and other technical metadata on incoming files. Other applications such as DROID might be utilized, however Data Accessioner and HashMyFiles should accomplish the basic tasks for most accessions.

[\(Appendix 2\)](#)

Note: The reviewing and testing of tools is an ongoing process. Many of the programs in the toolkit were designed for a specific purpose - a program may search for and remove empty directories or convert video files to another format - tasks which will not need to be done on every collection. The toolkit is a Google spreadsheet accessible to Task Force members that contains a list of programs, their purposes, links to documentation the Task Force or someone else created, and information on where it can be found on the workstation computer.

### **Testing of Initial Tools**

Data Accessioner and HashMyFiles were used on selected accessions listed on the ASC Electronic Records Interim Accession Log, which records current and incoming electronic records. Workstation Working Group members contacted unit staff to begin transferring identified electronic record content to the Data Drive on the workstation computer. In addition to identifying content for testing, this initial contact with ASC unit staff also helped provide a better understanding of which activities might be assigned to which role (unit staff or Task Force member) over the course of the workflow. After some successful and unsuccessful transfers using these tools, a draft workflow with more in depth instructions was created.

### **Drafting of Workflows**

Workflows were drafted to initially include a limited number of tools to help determine who would perform which tasks, and to begin to develop a step-by-step guide for staff to follow when ingesting and processing electronic records.

As the Workstation Working Group members ingested and captured information about the new accessions, it became clear that a method for sharing knowledge about the collection with unit staff and to document the success and challenges with the initial workflow was necessary. A Process Documentation template was created and used by Task Force members to understand and address issues that arose during transfer while the Inventory Report template was created

and used to share information with ASC unit staff about the accession itself<sup>5</sup>. These templates helped clarify the workflow; modifications were made accordingly.

### **Modifications of Workflows**

As the team began the process of testing the ingest workflow, issues arose regarding the selected tools themselves, where files were being moved to and how often, and what specific tasks should be completed to create the desired results. Working through these issues required a high level of problem solving skills and expertise to deal with the ever evolving set of tools and protocols for ingesting electronic records. Documenting the behavior of each tool was a key responsibility of Task Force members working in this arena. The following reflects selected high level modifications and notable examples of those modifications.

In the process of testing a proposed workflow for ingesting records, it was discovered that certain tools did not work in a uniform fashion on all file types. There were times when a tool failed to perform, or failed to recognize a file type, or inadvertently recognized “ghost files” (see below). Task Force members responsible for ingesting collections documented each issue and what type of file it occurred with, reported the error to the group, investigated the issue, and proposed a new tool or work around in order to properly ingest the entire collection.

Every collection brings with it the possibility of a new ingest scenario. As issues were discovered and new tools researched and tested, the ingest workflow was modified. In addition, a document with detailed instruction indicating how to use the particular tool and/or documenting inconsistencies was created or updated for every tool and linked to the workflow instructions.

The generation of “ghost files” was an issue frequently encountered when files transferred to external media from a Macintosh computer were ingested through the Windows workstation. While invisible on a Macintosh computer, the ghost files are displayed by the Windows operating system in what at first sight appears to be a set of duplicate files (. \_file1.pfd and file1.pdf). Deleting the ghost files one by one, although a simple enough task, is very time consuming. While it may be a reasonable option for small or simple collections, manually deleting thousands of files for multiple, large, or collections with complex folder structures is not scalable. As a result, software solutions were researched, tested, and installed to address and remove ghost files on an as needed basis. In order to make sure that future staff and workstation users are aware of the potential issues, a tool instruction document was created

---

<sup>5</sup> The Process Documentation template captured the name of the collection, the process that was used, capturing information on the documentation, general comments and the estimated time to complete the report and analyze the files. This report was intended for ERTF members. The Inventory Report documented similar information for the curators of the collection. Information included the collection name, original and current file location, basic information about the accession after it was analyzed, questions and discussion topics for the curator and what actions are needed to proceed. ERTF members feel that these two reports could be combined in some manner as they document similar things but for different audiences.

and linked to the workflow instructions. Reducing the likelihood of ingesting ghost files was one of the reasons a Macintosh workstation was acquired.

In addition to modifying the workflow, the ingest location in some instances was altered to provide a more secure environment for material that may include personally identifiable information. The Task Force originally suggested that all ingested collections would be copied directly to the networked Q drive for access by unit staff. However, after ingesting a few collections and considering the high level implications of security and collection management, it was determined that content would need to be placed initially on the Data Drive and then moved to the Q drive after reviewing content and identifying and restricting access to any private or sensitive information. More information on the use of the Q and Data drives can be found in the Security section under Phase I.

The sequence of securely ingesting accessions forced another workflow change. After content was transferred to the Q Drive, and an ERTF member met with unit staff to discuss initial findings, it was hoped that unit staff would review materials and inform ERTF of what they wanted to keep, a suggested organization of the accession, and concerns about access. However, more often than not, and for a variety of reasons, this envisioned process didn't happen. This has forced the Task Force to rethink initial appraisal and processing strategies, resulting in a suggested workflow similar to the current Central Processing workflow for paper records to address initial appraisal and processing. This workflow would also eliminate having multiple copies in various locations that may contain different files.

Based on experiences with changes made to the workflow during this process, it is clear that all future workflows will require clear documentation. In order to provide such documentation, the Workstation Working Group created and actively updates a list of instructions to follow for completing the ingest process, which also documents known discrepancies in tools. The current workflow document is called Processing Workflow Instructions\_MASTER ([Appendix 4](#)). Because the workflow is in a Google Doc that is collaboratively edited throughout the development process, prior versions do not exist as separate documents. In addition, graphical representations of workflows ([Appendix 3](#)) were created as visual guides to describe the workflow processes, and to assist in trying to understand who (unit staff / Task Force members) would be carrying out which tasks.

### **Summary of Ingest Work Completed**

Between September 2014 and May 2015, 13 different accessions were transferred from external media to the Data drive and/or the Q drive by ERTF members. A total of 1200+ GB currently sits on the Q drive and 2300+ GB on the Data drive. For these 13 accessions, materials have been ingested, a SIP created, and recommendations made for what the AIP should look like. In some instances the AIP has been completed, but in most cases unit staff have yet to determine what to keep. This has been a bottleneck in the current workflow that should be addressed in the next iteration of the Task Force. (See above.)

As of May 2015, there are 37 additional accessions totaling approximately 5450 GB waiting to be addressed, the majority of which are already in-house. It is imperative to keep moving forward and not only ingest but further process this material to preserve and protect the Libraries' electronic assets. In addition, selection criteria must be defined, and description and levels of access to materials determined.

### *Documents Supporting Workflows and Tasks*

#### **Development of Donor Guides and Transfer Information Sheet**

Another aspect the Task Force addressed was providing information to ASC staff and to potential donors about the specific issues surrounding electronic records. Two guides were created: the guide for donors is intended to make them aware of content, format, and rights issues, while the guide for unit staff encourages them to ask specific questions of donors and work with the Task Force when bringing in electronic records. ([Appendix 8](#))

A record transfer sheet was also created for ASC staff to use for incoming collections. This transfer sheet documents basic information about the records transfer, including contact information, type of materials, provenance of materials, and information on restrictions or concerns about materials. This sheet could be completed by the donor or by curatorial staff based on conversations with the donor. The goal of this sheet is to capture basic information to help electronic record processors and unit staff better understand the accession prior to ingesting and exploring the files. This document is available as a Google document and as a Google form. ASC staff is encouraged to use the method that best fits into current practices. ([Appendix 9](#))

#### **Development of Policies/Procedures/Guides**

Throughout the year Task Force members documented almost everything they did when working with electronic records. If they had a question or problem or knew of a question or problem others were experiencing, the issue was researched and the result documented. For example,

- It was important to make it easy for staff to download files transferred via email or cloud services to a specific location when using the workstation, but also to empower them to change this location on their own computers if desired. The Task Force documented how to change the download location in three different browsers.
- The Task Force documented how dates on files change when moving or copying them and researched various programs that would assist with file integrity and not change the dates. The program that best fit our needs was written into the policies and procedures documents.

Information that could be applied to another situation or that others might find useful, such as in the examples above, has been added to the Electronic Records Help Guide, an ever changing Google Document that has been shared with ASC staff. Information not made available in the help guide was shared internally as a learning experience and can be found throughout project documentation.

Documentation on how to use individual software tools was also created to assist with completing the workflows in a consistent manner. Step-by-step guides for over 20 tools or processes associated with the workflow were created. New guides are created as new tools or procedures are identified. Current guides can be found in the Task Force's Processing Instructions and Documentation Google folder.

In addition to documenting their experiences, Task Force members also reviewed existing literature and best practices to better understand issues related to working with electronic materials. Some of these best practices were incorporated into the workflow, some were documented as stand alone best practices, and others became policy. Policies were created specifically for the ERTF to follow when working with electronic materials from ASC and can also be found in the Electronic Records Help Guide. Policies related to ASC electronic records may or may not also apply to other Libraries resources. The next Task Force should continue to add to the policies and explore how they may apply or differ from the needs in other areas of the Libraries.

### **Drafting of Addendum to Deed of Gift**

Members of the Task Force reviewed the current Deed of Gift for Archives and Special Collections and determined additional aspects needed to be addressed in respect to an electronic records donation, specifically, providing access to and preserving electronic records.

The Task Force drafted an addendum to use with the Deed of Gift for those donations that include electronic records. The Addendum covers three major issues: digital file management; privacy; and disposition of physical transfer media. ([Appendix 10](#))

The first issue included permissions for accessing materials with passwords or encrypted data; accessing file fragments, deleted files, or system files that were otherwise unknown to the donor; and providing preservation services related to these types of materials. In the case of encrypted files, obtaining passwords and or encryption keys is vital for the Libraries to legally access and make copies of digital materials.

Next, the Deed of Gift offered an option to disclose copyright status of the donated materials, but did not provide an option to disclose personally identifiable information (PII) and/or private information. The Task Force determined that including this option in the Addendum allows the donor an opportunity to disclose the presence of potential private information and alerts the

processing archivists of any known precautions that should be afforded the donation (e.g., identifying and removing Social Security Numbers).

Finally, the addendum addresses the disposition of the physical media on which electronic records may be transferred. Typically, disposition of archival materials is at the Libraries' discretion, however, in the case of electronic records, the media is often just a container with no artifactual value. While it may not hold value to the Libraries, it may have a monetary value for the donor (e.g., a 1TB external hard drive) and providing an option to return the media is important to offer and document.

The Addendum was reviewed by the University of Minnesota's Copyright Program Librarian, Nancy Sims for copyright and other legal concerns as well as by the sponsors. Task Force members addressed the questions raised and completed a draft for final approval by the University's Office of General Counsel.

### ***Determine Short and Long-Term Human Resource Requirements for Performance of Technical Work***

The work of the Task Force makes it clear that permanent, dedicated staff will be necessary to carry forward the Libraries' efforts around electronic records. New materials are being acquired on almost a weekly basis and a large backlog, consisting of floppy disks and hard drives, is present in existing collections.

As of May 30, 2015, the ERTF had ingested 13 collections and the Accession Log listed 37 collections awaiting ingest and processing. Eighteen of these represent accessions added to the Log over the past six months and most others were identified by unit staff as having processed physical components but unprocessed electronic materials on legacy media (disks, disk drives, etc.) still requiring proper ingest and processing. Additional Libraries holdings contain unique electronic materials, but, due to the scope of the Task Force, a broad survey of previously accessioned collections has not yet been undertaken and are not represented on the log.

In order to understand human resource requirements for ingesting electronic records the ERTF tracked time spent on ingest. Due to the nature of electronic records it is difficult to employ a standard to determine how long it takes to ingest a collection based on the number of items in a collection, the total file size of a collection, the media on which the collection resides, or the file types contained in the collections; all of these affect the ingest time. For example, a collection may consist of two average sized PDF files or two very large video files for which the ingest time would vary considerably. Total size of a collection might be a better estimate; however many variables also come into play. For example, a 2TB hard drive is very different than 2TB worth of information on a stack of DVDs which need to be ingested individually.

Industry standards for paper collections estimate processing time by cubic foot. Currently there are no similar standards for estimating processing time for electronic records. Until such a time as these are determined, the Task Force made a best effort to record a variety of factors that impact ingest time in real situations, such as number of files, total bytes, file types, media, etc. For this report the estimated number of hours to ingest an average collection was calculated based on the time it took the ERTF to ingest 13 collections, as well as the number of hours it took to ingest one GB of data in the same sample set. Given the lack of industry standards, demonstrated need is based on both values<sup>6</sup> but it is important to keep in mind that these calculations are estimates for the amount of time necessary to ingest, not appraise, arrange, or describe the materials, all of which are necessary prior to making collections accessible to the public.

---

<sup>6</sup> A Ingest Time Tracking file was created to calculate these estimates. Tab 1 addresses the collections we did ingest; Tab 2 addresses the collections we still need to ingest. As we work with more collections, gain experience, and accumulate more data, ingest and additional processing time estimates will be refined. A template was created to track the time taken to complete specific tasks for individual collections which can continue to be used.



|  | # of Collections | Amt of GB               | Hours to Process  | Time Spent  |
|--|------------------|-------------------------|---|---|
| <b>Initial Ingest</b>                                | 13 collections   | 1122 GB                 | 92.75 hours to ingest   | 10 hours/month spent ingesting content                        |
| <b>Averages for Initial Ingests</b>                  |                  | 12 GB/hour <sup>7</sup> | 7 Hours/Collection <sup>8</sup>   |   |
| <b>Future Estimates (remaining on accession log)</b> | 37 collections   | 5450 GB                 | 259 Hours to ingest based on # of hours/collection<br><br>454 Hours to ingest based on # of GB/hour | ~ 2 years to process <sup>9</sup><br><br>~ 4 years to process |

Table 1: Summary of Ingest Work and Estimates for Ingest of Remaining Current Accessions

Based on the figures in Table 1, if the average rate of 10 hours/month that ERTF members combined were able to spend ingesting electronic records were to continue, it would take an estimated two years to ingest just the 37 collections currently awaiting attention. This does not take into account efficiency that comes with practice, nor any new accessions.

ERTF members had varying amounts of time available to ingest collections; some members were only able to ingest a single collection while others were able to ingest many more, sometimes working together. Members learned that having larger blocks of time (2+ hours) available to ingest materials was more effective than smaller blocks. Finding these larger blocks of time proved difficult, given that most Task Force members had many other time commitments. It is clear that focused attention on electronic records ingest eliminates time spent becoming reacquainted with the software and workflow. Having a consistent understanding of the variables inherent to managing electronic records is necessary for efficient ingest processes.

To address the time constraints of existing staff and the expectation of additional electronic record donations, the ERTF recommends the Libraries dedicate, *at minimum*, a half-time staff person (20 hours/week) to ingest these materials. At current rates it would take a half-time person 13 weeks<sup>10</sup> to ingest the current workload, keeping in mind this does not address any unforeseeable issues that arise with individual collections, or any of the appraisal, arrangement,

<sup>7</sup> Calculation: total number of GB (1122) divided by “meeting time” and “ingest time” [listed on the Ingest Time Tracking file] (92.75) = 12.10 GB/Hour (12 GB/Hour)

<sup>8</sup> Calculation: Sum of “meeting time” and “ingest time” [listed on the Ingest Time Tracking file] (92.75) divided by 13 collections = 7.13 hours/collection. (7 hours)

<sup>9</sup> Calculations: Working about 10 hours a month it would take about 26 months to reach 259 hours. 26 months is over 2 years.

<sup>10</sup> Calculations: 37 collections x 7 hours/20 hours/week = 12.95 weeks (13 weeks)

or description of collections. Calculating how much time would be required for a half-time staff person to process the current backlog based on estimated GB results in a higher number, about 23 weeks<sup>11</sup>. This illustrates why there are no standards for estimating ingest time.

A half-time person whose primary responsibility would be working with electronic records should reduce the required time estimated for ingesting materials as their familiarity with working with electronic records grows. However it is important to keep in mind that the pace of new electronic record donations is increasing as individuals and organizations work daily in online environments. This is evidenced by greater percentages of electronic materials in recent collections, a trend that is expected to continue. In addition, and as a direct result of the Task Force's outreach efforts among ASC staff, staff feel more comfortable discussing electronic records with donors, which will in turn lead to donors being more confident that their materials will receive appropriate and timely attention and care. The Libraries cannot afford to continually add to the backlog of electronic records being donated, and must cope with incoming material in a sustainable manner.

In the long term, a full-time dedicated staff person is arguably the only way to address not only the ingest process, but also to assist with appraisal, arrangement, and description of electronic records in a consistent manner that parallels current practice with paper records. Without a dedicated person who has an in depth understanding of evolving workflows and protocols, and who can provide a consistent approach with curatorial staff, very little headway will be made in utilization of the workstation and addressing the records being collected. Given the fact that people and organizations today do most of their work in electronic formats, the nature of archival collecting is facing a fundamental shift. In the longer term it is likely that several electronic records specialists will be needed. Securing dedicated staff now will position the University Libraries at the forefront of organizations committed to providing exceptional access to unique and historic material.

## Communications and Outreach

### *Survey Working Group*

The Survey Working Group (Kevin Dyke, Mary Miller, and Erik Moore) was created around the same time as the Workstation Working Group to develop a survey for all Libraries staff regarding the existence of hidden collections in their respective units (that is, non-licensed, uncatalogued, digital content not in a repository or otherwise broadly accessible). The survey, as well as a follow-up brown bag discussion with ERTF members and respondents, was also used to identify concerns and needs for hidden collections across the Libraries. The survey provided data on the range of media types and file formats to be addressed in current and future workflow development.

---

<sup>11</sup> Calculation: If there are a total of roughly 5450 GB in the 37 collections remaining to ingest, and we use the 12 GB/hour calculation these additional 5440 GB of information would take roughly 454 hours which is about 23 weeks of a half-person's time.

Not surprisingly, CDs and DVDs are widely held media, and represent a significant concern amongst respondents of the survey. Email attachments were also increasingly a source for collections content. The majority of respondents indicated both concern about preserving hidden electronic content in the future, and indicated interest in further discussion about developing workflows for preserving their electronic collections; however, access to electronic records is often more in the forefront of people's minds. A survey about access or a one-on-one discussion addressing specific needs may produce additional information.

The survey results indicate that everyone feels they are in a unique situation, that their needs are different. This may be true in some instances, but there is usually more crossover than what can be seen on the surface. A possible model to consider when working with Libraries staff could be to have Task Force members or electronic records staff serve in a liaison role, to be available to address questions and concerns but also ask pertinent questions that may raise awareness of electronic records issues. Another model might include forming a group similar to the Strategic Digitization Review Group in which members would work together to address issues that are brought to them on an as needed basis. This latter model would require Library staff to be aware of such a group and feel comfortable contacting them as necessary.

As the workflow for ingest and processing electronic records continues to evolve and gel, the process becomes easier to share across the Libraries. CDs, for example, were ingested through the workstation during this first year and much time was spent deciding how to name folders and organize the collections. In the future, the established best practices could be extrapolated for other Libraries department needs. Other interaction models, such as discussed above, could also be pursued to increase the Libraries' overall knowledge of how to work with and provide access to electronic records.

### **Contact with Archives and Special Collections Staff**

The majority of the Task Force's work centered on materials from units within the Libraries Archives and Special Collections (ASC). Multiple members of the Task Force are also ASC staff themselves and attend the bi-weekly ASC meetings. This provided opportunities to keep in contact with the main audience and continually inform ASC staff about Task Force activities. The following statements document some of these ongoing interactions:

- ASC staff began seeking out Task Force members to ask for advice on working with donors of electronic records. Responses to questions being asked were later documented in the donor guides ([Appendix 8](#)) as resources for both ASC staff and the donors themselves.
- As a first step in understanding incoming accessions that include electronic records, Task Force members ask ASC staff to complete the ASC Records Interim Accession Log (a Google spreadsheet documented in [Appendix 6](#)) to better inform Task Force

team members about electronic materials recently acquired or soon to be acquired. This log has evolved throughout the project and is continuously being added to as staff become aware of new accessions that include electronic records. The log records answers to general questions about the electronic materials as well as information about the provenance of the donation that assist with electronic records processing.

- As the project moved forward and the Task Force was ready to test tools and initial workflows, the ASC Records Interim Accession Log was used to select accessions. Unit staff whose materials were selected met with Task Force members as needed to discuss individual acquisitions. Some acquisitions required one meeting, others required multiple meetings. Having a point person familiar with electronic records available to answer questions during this possibly extended process kept the process moving forward.
- ASC staff were also asked to review materials Task Force members were developing. For example, ASC staff provided feedback on both the donor guides and the E-records transfer sheet prior to implementation.

Ongoing discussion and open communication between ASC staff and electronic records processing personnel will be required as long as electronic records are part of the archival record.

### **Working with DMCI**

Throughout this project the Data Management and Curation Initiative (DMCI) has wrestled with similar issues, especially electronic records workflows and concerns around sensitive information. Task force members involved in both initiatives followed along as the workflows of the DMCI progressed. The main area where these two initiatives crossed was in working with and better understanding how to search for private and/or sensitive information. Staff on both groups explored programs and methods for identifying such data. The Task Force demonstrated Identity Finder and shared step-by-step documentation with DMCI members who began to use the tool in their own workflows. The Task Force and DMCI should continue to be aware of each other's work and explorations as there are many points at which information can be shared and appropriate common procedures established.

### **Presentations**

In addition to discussing Task Force activity with interested Library parties, the Task Force discussed/shared activities with others. Public presentations were given describing the Task Force activities which provided an opportunity for knowledge transfer and feedback between Task Force members and the University community as well as with other outside individuals and organizations.

- Fugitive Bits: Preservation and Access of Electronic Historical Records, Graduate Student Assembly; April 2014. [Carol Kussmann, Arvid Nelsen, Lara Friedman-Shedlov]
- BitCurator visit and presentations; July 9-10, 2014. [Porter Olsen]
  - An Introduction to Digital Forensics. This presentation was given to members of the Task Force as part of a two day hands-on training for BitCurator.
  - What Falls Out: Preserving our Digital Heritage with BitCurator. This presentation was a public presentation given by Porter Olsen as part of the BitCurator site visit to inform a larger group about digital forensics.
- Fugitive Bits: Taking Born Digital Records from Up in the Clouds Down to Earth, Upper Midwest Digital Conference; August 2014. [Arvid Nelsen, Lara Friedman-Shedlov]
- Electronic Records Task Force Update, UMN Libraries Brown Bag; September 2014. [Carol Kussmann]
- Demystifying Digital Records Processing... Step by Step, Byte by Byte, Library Technology; March 2015. [Carol Kussmann, Lara Friedman-Shedlov]
- Update on Task Force Activities [based on the Library Technology Presentation], Data and Technology Division Meeting; March 2015. [Carol Kussmann]
- Bits and Pieces: Preserving Born Digital Records, Minnesota Digital Library Annual Meeting; June 15, 2015. [Lisa Calahan, Lara Friedman-Shedlov]
- Electronic Records Task Force Update, Library Assembly; June 18, 2015. [Carol Kussmann, Arvid Nelsen]
- Fugitive Bits: Taking Born-Digital Records From Up in the Cloud Down to Earth, 2015 RBMS Conference; June 24, 2015. [Arvid Nelsen, Lara Friedman-Shedlov]

## Conclusion

In the past year Task Force explorations have resulted in the establishment of a state of the art workstation implementing an arsenal of professional tools including many developed for forensic analysis that permit safe and secure transfer and analysis of primary source materials in electronic formats.

The workflows created and implemented for secure transfer and stabilization of materials that adapt industry best practices and guidelines to the Libraries' specific needs and resources were hard won. Their development required intensive learning and experimentation, and in fact had to be reworked multiple times in light of new formats encountered and new tools implemented to meet emerging needs. In fact, the Libraries will face ever changing developments in the software and systems used to ingest, process, preserve, and serve collections. This is work that does not have a finish line, making it imperative that we keep moving forward following best practices or run the risk of not being able to care for the Libraries assets.

Workflows included not just specific tasks and tools, but also the identification of appropriate staff roles (both technical and curatorial/selection) for completing the range of steps in what is a complex process. Defining roles was also subject to change during the iterative process of workflow development. For example, participation of unit curatorial staff was incorporated into initial workflows to address appraisal and retention decisions, but this did not result in timely completion of tasks. Appraisal and retention in the context of electronic records are highly specialized and complex. The technical nature of this activity precluded unit staff, who are engaged daily in subject specialization and patron services responsibilities, to learn and retain knowledge of critical skills and policies. Careful consideration and discussions among ERTF members and sponsors led to reconfiguration of workflows to make the process more centrally managed and controlled. Decisions affecting processing beyond ingest in preparation for long-term storage and public access remains to be addressed.

Although access platforms have yet to be formally developed and were not within the scope of this charge, patron communities have requested access to electronic records. In order to make reasonable accommodations to the needs of ASC's patron base, we did prioritize the ingest of the requested electronic collections. Because current workflows have focused on ingest and not on more extensive processing, materials were often not processed further prior to being made available to researchers. While this was acceptable on a limited basis, the practice will not scale.

The Task Force met most of its goals and achieved the actual and practical ingest of a number of collections in the process of developing workflows. The scope of the Task Force charge included only the initial stages pertaining to the long-term stewardship of the Libraries' electronic assets. As the Libraries move forward to address long-term stewardship and access, knowledge gained from the work of the Electronic Records Task Force should be carried forward and built upon.

## Next Steps

Below the Task Force addresses what still needs to be done in relation to electronic records and makes suggestions on how it can be accomplished.

### Ongoing Development

Over the past year the Task Force created a workflow for securely ingesting electronic materials. Moving forward the Libraries needs to define and create post-ingest processing procedures that include appraisal, arrangement, and description of ingested materials. This process would be similar to the development of the ingest workflows, policies, and procedures and would include defining 1) the preservation object, which for electronic records is an Archival Information Packet or "AIP" and 2) discoverability mechanisms for electronic records.

Complete stewardship of electronic records must address both the preservation environment and the access environment. The ERTF should continue to work with the Libraries' digital preservation staff on incorporating these assets into a secure storage location, or when implemented, a digital preservation system. Until that day arrives, however, current available and developing resources for ensuring the integrity and security of ingested collections should continue to be evaluated.

Simultaneously, requirements for providing access to materials, including policies for determining and assigning levels of access and the creation of mechanisms for providing access--from secure, on-site, non-networked viewing stations, to mediated online access, to open access via the Web--need to be assessed and implemented.

### **Production Environment for Ingest**

The current Task Force developed a workflow that includes procedures, policies, and tools to ingest electronic records. To move forward, and to keep up with the current rate of new acquisitions, the Libraries must continue to utilize the established ingest workflow on an ongoing basis. It is not possible to wait until additional steps, including the forthcoming preservation system and yet-to-be-defined access system(s), are in place. The ultimate goal is to ingest all materials listed on the ASC log in a timely manner; acknowledging that new records may be donated at any time. To reach this goal, the Libraries must continue to evaluate, but assume the ingest workflow has moved past the development stage and is in production mode.

### **What We Need to Get This Done**

This report provides recommendations for short- and long-term staffing needs. However, given the new understanding of the time requirements for both development work and processing work, it is clear that ERTF members cannot continue to be responsible for ingesting electronic records for ASC, let alone address the standard archival description tasks that make access to the records possible.

ERTF recommends charging a new task force to develop post-ingest processing protocols, including the appraisal, arrangement, and description of electronic records. The new task force can monitor and tweak current ingest workflows, but would not have the capacity to develop post-ingest processing *and* continue ingesting incoming materials in a timely fashion. In addition, this new task force should establish the needed protocols for providing access to electronic records.

For the Libraries to build upon the work of the ERTF, and to keep from falling behind in addressing incoming electronic records, it is imperative that, at minimum, a half-time, dedicated staff person be hired. Ideally, to keep momentum going and to implement protocols that will be developed for description and access to these materials, a full-time staff person is needed. As

the experience of the ERTF has shown, this is not a function that can be added to the responsibilities of existing staff.

Equipment will be an ongoing need. The Windows workstation is currently on the replacement schedule, but the Macintosh computer is not. Over time this computer will need to be replaced and added to the replacement schedule. In addition, other hardware and software may be required to explore, ingest, process, and provide access to future accessions.



## Appendix 1: Charter for Electronic Records Task Force

Charter | 2014-Jun-09 | AN  
Approved by Libraries Cabinet March 18, 2014

### Electronic Records Task Force

#### Institutional Background

The Electronic Records Task Force will develop the capacity of the University of Minnesota Libraries to respond to emergent, critical needs in the preservation and accessibility of historically and culturally important documents found in the records and papers of persons and organizations central to the University's collections of distinction. Evolutions in technology have changed the way the University's donors create and document their work, from analog to digital media. Born-digital documents require new tools and workflows for acquisition, ingest, processing, preservation, mediation, and access. Building upon the work of the Digital Preservation Framework Task Force and the Digital Repositories Task Force and in collaboration, when appropriate, with the Data Management and Curation Initiative (DMCI), the Task Force will focus on the implementation of mechanisms for acquisition, ingest, and processing. Later work will be required to develop levels of mediation and modes of access.

The Digital Repository Task Force Findings, submitted by the Task Force and reviewed by Cabinet and LLC in July 2013, included recommendations (pages 16-17) for the implementation of a back-end system for ingest and storage of born-digital materials curated by the Department of Archives and Special Collections. The Electronic Records Task Force will aim to implement the Digital Repository Task Force's recommendations and then broaden the scope beyond ASC's immediate needs:

- define requirements for immediate stable, secure storage area that allows for curatorial access and control, and ensures the preservation of the digital assets through routine file back-ups;
- investigate existing tools (e.g. BitCurator, Curators Workbench) for born-digital file management and processing, including virus scanning, checksum creation, file format preservation, write-blockers for files and/or disk images, metadata recording, and opportunities for format migration or versioning;
- consider policies and procedures, best practices and standards around born-digital materials that require partnering with Libraries Data & Technology for secure storage and preservation environments;
- at the point of the digital preservation management system implementation, ASC archivists/curators should work with the Libraries digital preservation specialists to move digital assets into the DPMS and make full use of the tools that are available.

#### Purpose of the Task Force

The Electronic Records Task Force will establish and implement the ability of the Libraries to ingest and manage materials that exist natively or solely in electronic form. This includes identifying existing resources or acquiring new resources (technological and human), developing workflows, and defining staff roles.

Unlike the e-books and serials currently purchased or licensed by the Libraries, electronic records are defined here as unique or rare materials not created for the purpose of publication and public consumption. These materials, for the most part, are not finished products but the background or ancillary materials to finished works. They may be complex and variable aggregates of information and/or formats. Examples include, but are not limited to, born-digital archival materials (personal and organizational records that typically include a variety of file types), data sets, and geo-spatial data. Materials may be pertinent to both special and general collections areas. Selectors in those areas, working with creators and donors, will need to make decisions regarding selection and access.

These materials will require infrastructure and capacities currently available or forthcoming, such as: a long-term digital preservation environment; platforms for varying levels of access including (but not necessarily limited to) the open web, mediated access (through logins or formal request mechanisms and dedicated on-site workstations); and semi-dark repository and other access controls. They will also require special hardware and software capable of: ingesting content from multiple operating systems and storage devices including legacy technology; ensuring authenticity and integrity of data by minimizing transformations that occur during transmission (copying and moving files) and documenting the chain of custody and actions performed; and providing appropriate levels of access.

The products of this Task Force will support the work of Libraries' curators, selectors, and units in meeting their collection development strategies. The Task Force will define the roles of curatorial staff (identifying and communicating information to donors and technical staff) and technical staff (workflows for the creation of SIPs and AIPs). The Task Force will propose new positions as needed, and will identify hardware and software required for the Libraries to become a leader in this arena.

Electronic records present new concerns and considerations regarding the integrity and validity of data and appropriate levels of access to born digital materials while maintaining existing requirements of collection development policies and donor relations. The Task Force will address these so the University of Minnesota Libraries may provide the quality service rightfully expected by a top national research institution, by leveraging existing and forthcoming technical infrastructure, acquiring additional resources as needed, promoting best practices, and minimizing institutional redundancies.

#### **Sponsors**

John Butler  
Kris Kiesling

#### **Task Force Members**

Carol Kussmann, Co-Chair  
Arvid Nelsen, Co-Chair  
Lisa Calahan  
Kevin Dyke  
Lara Friedman-Shedlov  
Mary Miller  
Erik Moore

Justin Schell

### **Resource Personnel**

Lisa Johnston  
Jon Nichols  
Mike Sutliff

### **Duration**

12 months

### **Budget**

Ongoing costs are yet to be determined, but it is estimated that the initial workstation with appropriate processing software will cost \$10,000 for computer equipment across multiple platforms and for modifications to a workspace.

### **Process Tasks**

#### *Phase 1: Develop Initial Capacity for Electronic Records Ingest and Processing*

1. Identify, procure, and implement initial hardware and software needed for secure ingest and processing
2. Identify and secure access to initial file/storage space specified for working with incoming records and storage of processed records

#### *Phase 2: Define Tasks and Workflows for Staff*

1. Define tasks to be performed by staff engaged in technical functions (ingest and processing)
  - a. Assign role of ingest and processing to select members of the Task Force initially
2. Define tasks to be performed by staff engaged in curatorial/selection functions
  - a. Information to be gathered and conveyed to work group
  - b. Decisions to be made regarding selection, arrangement, levels of access, etc.
  - c. Information that can be conveyed to donors
3. Determine short and long-term human resource requirements for performance of technical work

#### *Phase 3: Develop Ingest and Processing Workflows*

Develop workflows to support secure ingest of electronic records and the creation of Submission Information Packets (SIPs) and Archive Information Packets (AIPs).

1. Survey electronic records sets in our holdings  
Note: Due to the timeframe of this Task Force, this work will focus on recently acquired and anticipated born-digital collections with a further focus on collections suitable for work developing and testing workflows. A full-scale survey of electronic records within all existing holdings exceeds the focus of this group, though recommendations may be made for units and departments that have such materials.
2. Select sample electronic records sets with which to test workflows exhibiting best practices as established in current professional literature and drawing on work already developed in-house by specific units, projects, or initiatives (e.g., ASC, DMCI)
3. Establish institutional workflows by modifying test workflows according to lessons learned in testing and according to institutional priorities and capabilities

### **Next Steps: Preservation and Access**

The immediate needs of born-digital electronic records and the timeframe of this Task Force require focus on the Libraries' capacity for ingest and processing of these materials, but as noted above they will also require long-term preservation and varying modes of access. The Libraries will need to define levels of access, identify mechanisms for providing those levels of access, and submit formal proposals for the acquisition or development of mechanisms needed. Upon completion of the goals stated in this charge, the Task Force may be redefined to contribute to this process. Tasks may include:

Working with the Digital Preservation and Repository Technologies department and other relevant departments or groups to:

1. Identify and secure access to existing or anticipated hardware/systems available to the University Libraries, including
  - a. Preservation repository, including long-term needs for data migration, versioning, emulation, etc.
  - b. Access repositories, including variable levels of access, restriction or mediation currently available to us
2. Develop a plan to secure the resources (technical and human) required to meet identified preservation and access needs not currently available to the University Libraries

### **Deliverables**

The outcomes of this group include:

- Workflows, procedures, and policies - as required - for accessioning and ingesting electronic records
  - Including specifying tasks performed by individuals with collection development responsibilities (decision-making, information gathering, and communication) and those performed by individuals with technical responsibilities
  - Note: Selection decisions will remain the purview of collection development policies in individual units and departments. Levels of access required for the electronic records will be the responsibility of curatorial staff
- Recommendations for technical infrastructure/hardware, including both existing or anticipated systems and others as identified, to support the policies and procedures.
- Recommendations for personnel to support the policies and procedures.

### **Scope**

The focus of this Task Force will be on Electronic Records coming into the Department of Archives and Special Collections under current collection development policies, with awareness of and sensitivity to materials coming in to general collections units that pose similar issues and require similar treatments. The goal is to provide an infrastructure capable of meeting the needs of any unit or department contending with acquisition and ingest of electronic records that are not met by other Libraries initiatives.

The purpose of this group is to develop the infrastructure and workflows to safely and securely ingest the flow of electronic records that have come to the Libraries already, that are coming in now, and that are anticipated. Selection decisions will remain the purview of collection development policies in individual units and departments. However, we will advise

units and departments on appraisal and selection policy as it relates to: 1) technical capabilities of the Libraries (e.g. – email management, website archiving) and 2) federal and state law and University policy (e.g. – private data).

The Task Force will work closely with concurrent initiatives, such as the DMCI;

- Currently there is cross-pollination of membership between these groups
- Communication and collaboration between groups is desirable, because
  - They have overlapping interests and needs
  - Interests do not have clearly definable borders and we will most likely encounter situations where we need to use each other's expertise (e.g. – If an ASC donation includes data sets)

Some distinguishing factors of materials to be addressed by this Task Force include:

1. Materials may be unique and or rare (i.e. – not widely available and/or published materials);
2. The University Libraries will be responsible for the physical custody and authenticity of authorized files
  - i.e. - these are not resources to which access is licensed and/or which are physically maintained by agencies outside of the University Libraries
3. Sources of materials (creators and donors) will include both University affiliates and non-University affiliates
4. Materials may require greater mediation/curation
  - i.e. - Library professionals are more likely to select collections and manage ingest processing, instead of relying on creators/donors to select and upload materials
5. Commitment to preservation and access is long-term and extends beyond the currently perceived informational value, which is often the focus of efforts around data sets

#### **Stakeholders and Reviewers**

- Libraries Cabinet
- Content Services Steering Committee
- Archives and Special Collections staff
- Data Management and Curation Initiative (DMCI)
- University Digital Conservancy co-Directors
- Digital Library Services
- UMedia Archive
- Data & Technology Division; Digital Preservation and Repository Technologies department
- Collection Management and Preservation Strategist
- Collection Development Officer
- CLA Digital Content Library

## Appendix 2: Hardware and Software

The first phase of the Electronic Record Task Force's charge (2014-2015) was to develop initial capacity for electronic records ingest and processing. The following documents the equipment acquired to meet this goal. Over time additional needs may occur.

### Windows Workstation

Purchased and set up by Libraries IT for the specific use of ingesting and processing electronic records, this workstation was built based on recommendations from comparable institutions.

- **System:** Dell Optiplex 7010 minitower; Windows 7 64-bit
- **Processor:** Intel Core i7-3770 (Quad Core, 3.40GHz)
- **RAM/Memory:** 16GB
- **Storage**
  - Primary – 500GB hybrid drive (boots from SSD but stores on disks)
  - Secondary – 4TB drive for scratch space
  - Special Network Drive - 10 TB
- **Optical Drive:** Internal CD/DVD burner drive (16X DVD+/-RW SATA)
- **Write Blocker:** Internal forensic PC Tableau Write blocker (includes SATA, IDE, SAS, Firewire and USB 3.0 connections)
- **Other:** 4 front facing USB (2 USB 3.0 + 2 USB 2.0) 6 USB in the back
- **Network Connectivity:** Workstation is in Active Directory and SCCM (Configuration Manager)
- **Administrative Rights:** Admin rights to Install and Configure but not for daily use (UL-Install account)

### Macintosh Workstation

To better assist with ingesting and processing materials that were created on a Macintosh computer, we acquired a Macintosh computer that previously was used in the Libraries. Because of this, if the machine breaks down, it will not automatically be replaced. However, we felt that having an older machine was an advantage when working with older digital materials as many older drives would not connect directly to a newer machine.

- **System:** iMac, 27", Mid 2010; OS X 10.9.5
- **Processor:** 2.8 GHz Intel Core i5
- **RAM/Memory:** 8 GB, 1333 Mhz DDR3
- **Storage:** Internal 1TB
- **Optical Drive:** Internal CD/DVD burner drive
- **Other**
  - Built in SD card reader
  - 4 back facing USB ports
- **Network Connectivity:** Workstation is in Active Directory and SCCM (Configuration Manager)

- **Administrative Rights:** Admin rights to Install and Configure but not for daily use (UL-Install account)

### **External Drives and Equipment**

Adding to the functionality of the Windows workstation are the external drives both purchased or received free of charge.

- Memory Card Reader (“All-in-One” connects via USB, has 4 common slots)
- Zip disk drive
- 3.5” floppy disk drive
- 5.24” floppy disk drive
  - Needed to purchase external card/convertor for floppy drive connector to USB connector. Found a used case for a SCSI CD-Rom drive that fit perfect for the floppy drive assembly. Found external power supply on Ebay.
- Write blockers:
  - eSATA/IDE Bridge T35es-R2 write blocker
  - Tableau T8-R2 Forensic USB Bridge write blocker
- External Hard drives
  - 2 Tb Windows formatted (USB)
  - 2 Tb Macintosh formatted (USB)

### **Software and Tools**

Much of the software used in the ingest and processing of electronic records has been little to no cost. Programs that we use include but are not limited to:

- BitCurator
- Bulk Extractor
- Data Accessioner
- Data Discovery/Identity Finder (\$60)
- DROID
- Duplicate File Finder
- Eraser
- Fixity
- HashMyFiles
- TeraCopy (\$20)
- QuickView Plus (\$49)

### **Security**

As with all assets of the Libraries, security is a concern. To assist with security measures, the Workstation itself is located in the Charles Babbage Institute which is a locked suite with limited people having access to the room. Access to the preservation and working copies of the digital records on network drives are limited to members of the Task Force and unit specific staff.

## Appendix 3: Workflow Images

The following workflows document ideas and processes throughout the year.

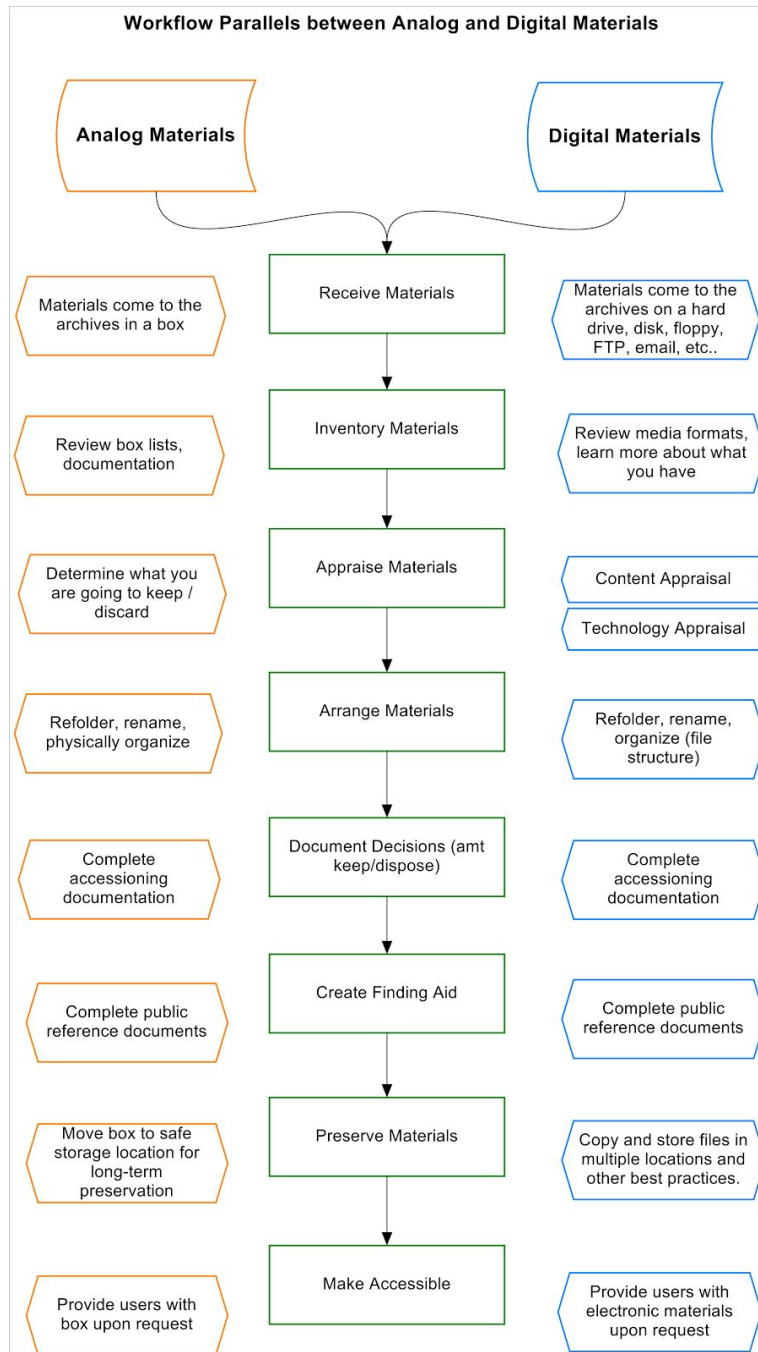


Figure A1: Capturing the similarities and differences between the analog and digital workflows.



## High Level Workflow for Processing Electronic Records Without Role Assignments

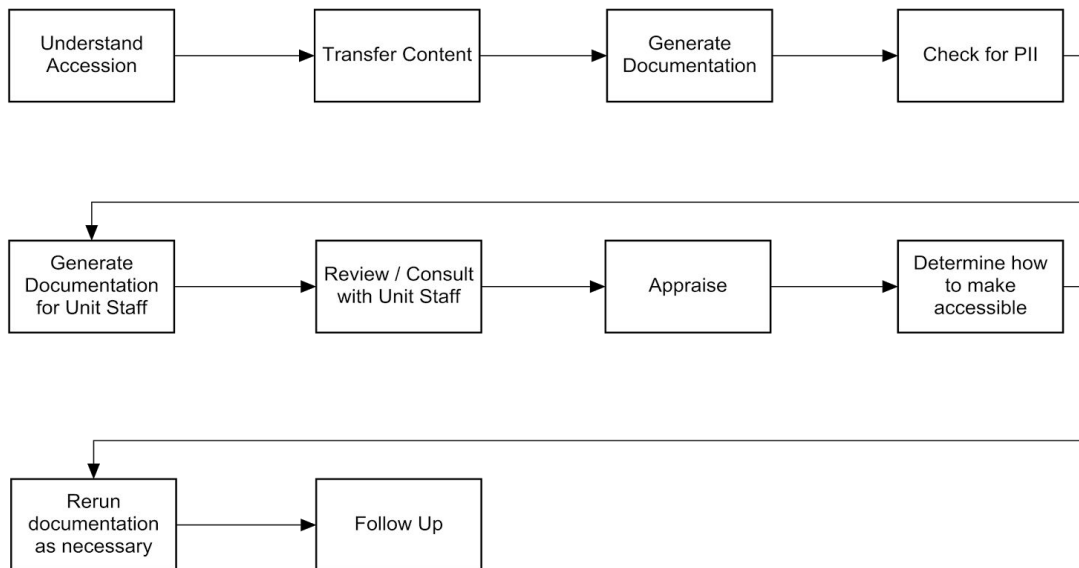


Figure A2: High level workflow documenting action steps and interactions between staff teams.

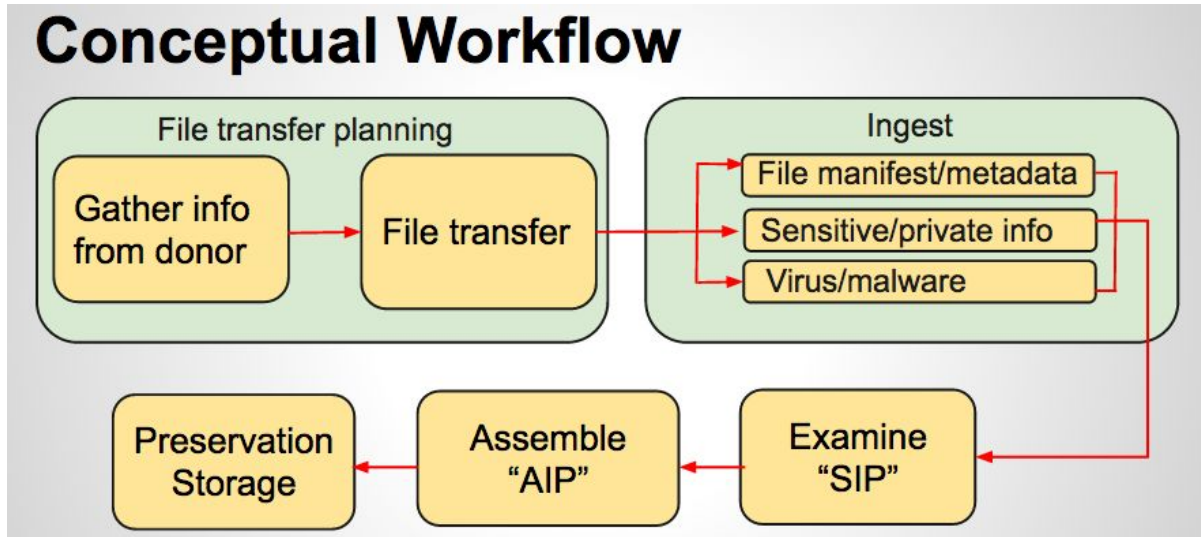


Figure A3: Conceptual workflow steps

## Storage/Preservation Locations

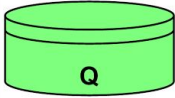
Electronic records for ASC can be found in the following locations.



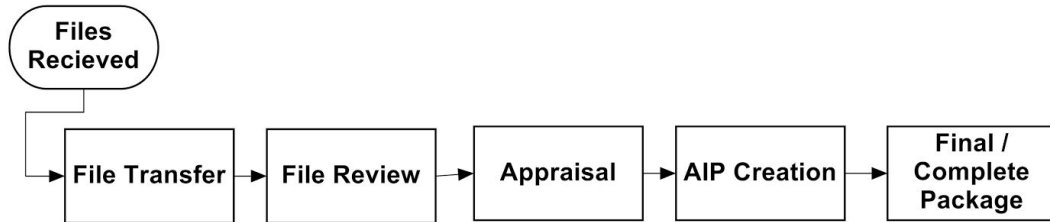
- ERTF Access (workstation only)
- Initial Transfer Location
- Final Full Copy (including PII info)
- Location not backed up.



**Do Not Use for Digital Archive Files**



- Unit Staff [selected] and ERTF Access
- Final Full Copy (minus PII info)
- Location "backed up" with Shadow Copy (versioning) by OIT.

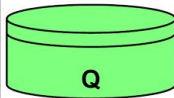


Initial Storage for files transferred for review

Review completed and documentation created

Electronic Record Workspace (if is PII involved\*)

Completed Accession 'storage' for all collections



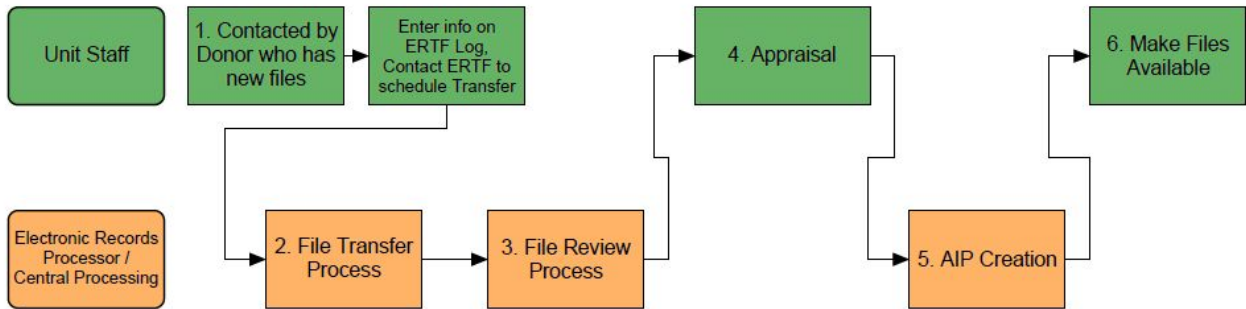
Electronic Record Workspace (not including restricted content)

Completed Accession 'storage' (for collections without PII)

\* if PII or other restricted information is included in the materials, work must be done on the Data drive. PII information can't be stored on Q or L.

Figure A4: Documenting workflows in relation to the storage/preservation environment

### Feb-March 2015 Initial Role Assignment Ideas



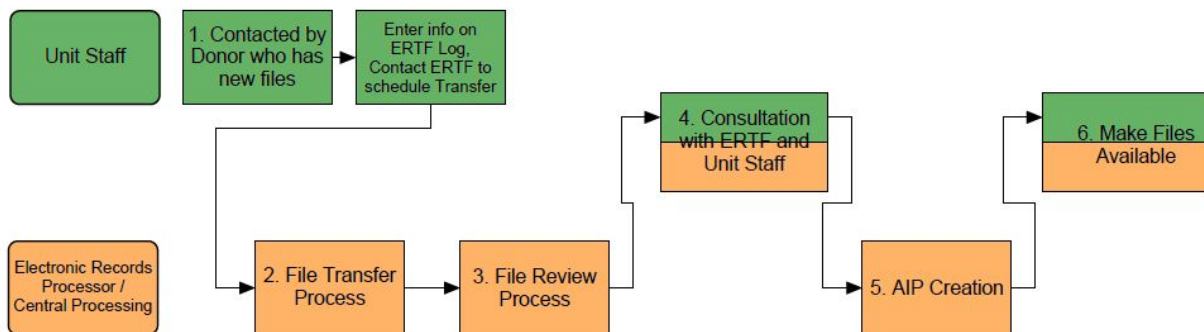
Numbers in the steps above correlate to other workflow files listed below.

1. Donor Contact - Pre Transfer
2. TransferFilesToDateDrive
3. (2 Workflow steps) CreateCollectDocumentation and Review with Unit Staff
4. [In Development]
5. [In Development]
6. [In Development]

Note: As of April 2015, this file will not be updated.

Figure A5: Initial role assignments for workflow steps.

### April-May 2015 Role Assignment Ideas



### Associated Workflow Documents

1. Donor Contact - Pre Transfer
2. TransferFilesToDateDrive
3. (2 Workflow steps) CreateCollectDocumentation and Review with Unit Staff
4. [In Development]
5. [In Development]
6. [In Development]

Figure A6: Reevaluated role assignments for workflow steps based on experiences.

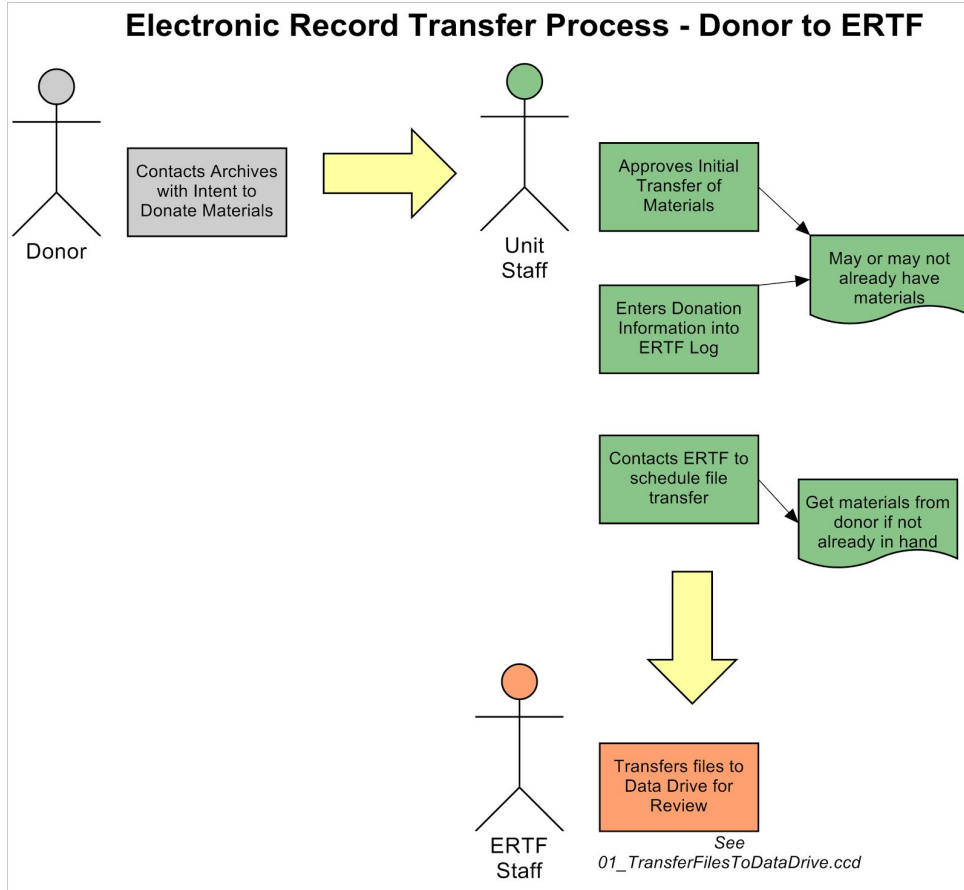


Figure A7: Details for first step of workflow. Getting files to ERTF staff.

### Transfer Files to Data Drive

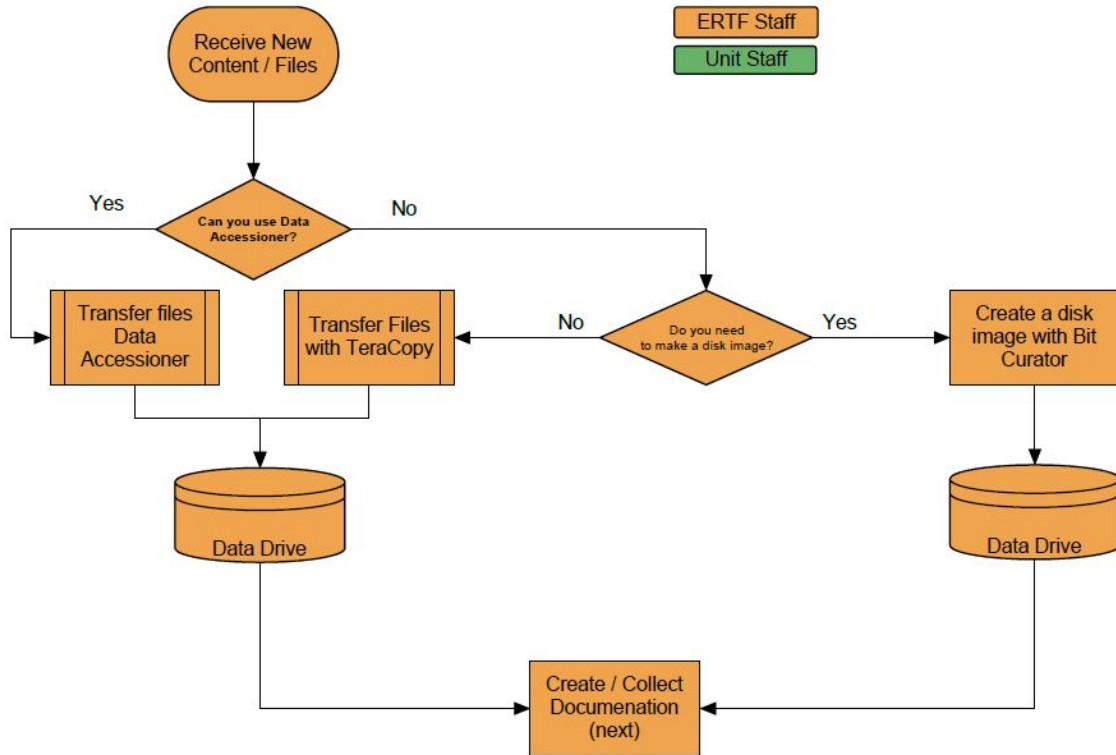


Figure A8: Second workflow step. ERTF transferring files to secure location.

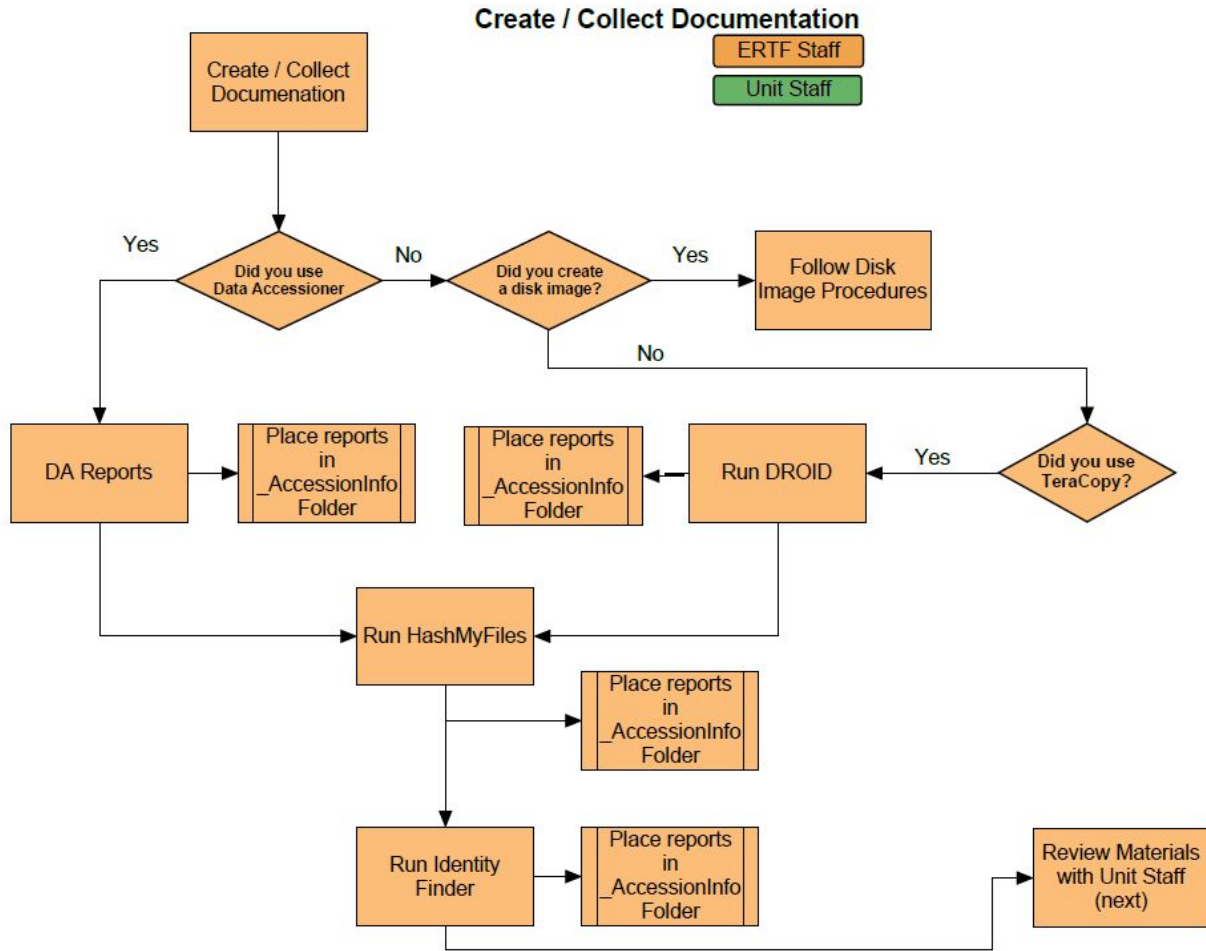


Figure A9: Third workflow step. Reviewing files, documenting collection, collecting metadata.

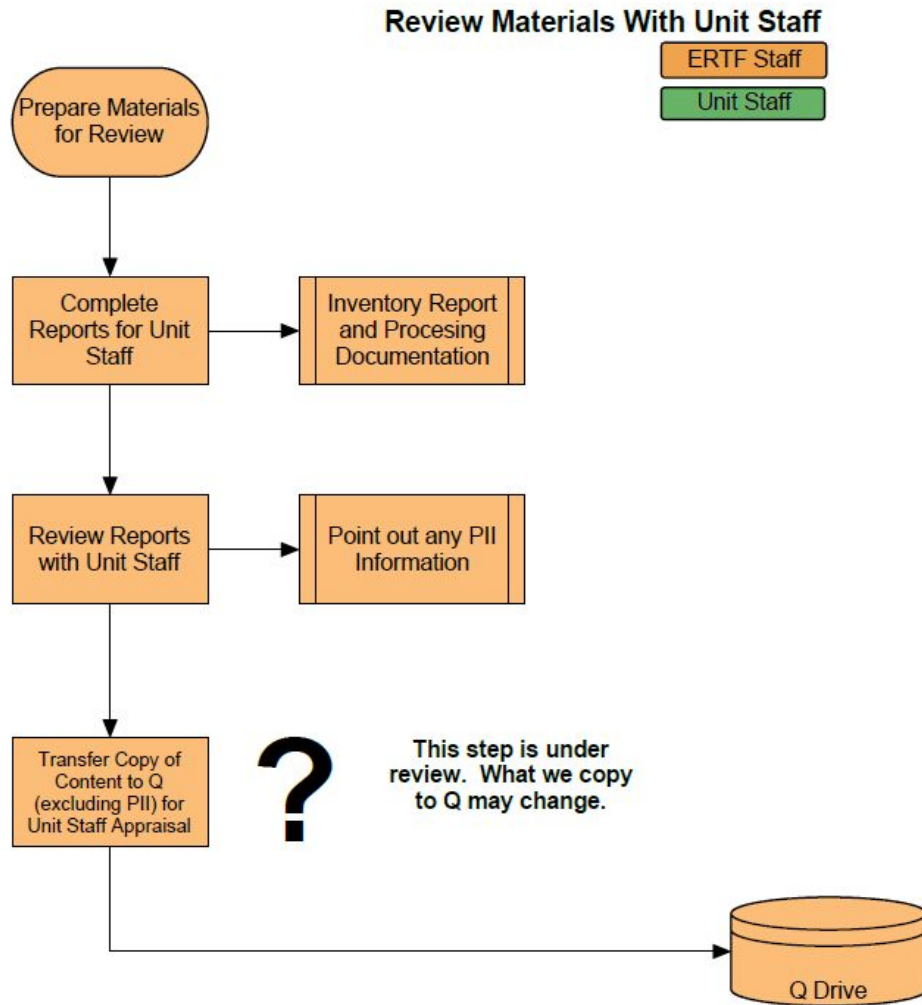


Figure A10: Fourth workflow step. Reviewing materials and asking questions with Unit staff.

## Appendix 4: Master Processing/Ingest Workflow Instructions

### General Process for Handling Digital Content Workflows/Procedures

*Note: The process for handling digital content described below was drafted 2014-2015 and reflects practice current as of May 2015. The original format of this document is a Google Doc. Because this Google Doc contains numerous links to documents that may or may not continue to persist, we have made only the main body of the workflow available below. The text in bold after or within a step indicates where as of May 2015 additional documentation existed.*

#### A. Ingest

1. Understand accession (**Survey / Donor Guide**)
  - 1.1. Is this a new collection?
  - 1.2. Is this part of an existing collection?
  - 1.3. What are the expectations for the materials?
  - 1.4. File transfer options (donor to curator/archives)
  
2. Review Deed of Gift
  
3. Begin to transfer content to Data Drive
  - 3.1. Set up new location using proper naming conventions (**instructions**)
  - 3.2. Prepare to Transfer files to Data Drive
    - 3.2.1. Connect media to workstation (**instructions**)  
*Note: Determine if you are working with Macintosh formatted media and remove unnecessary files. (Link to Draft)*
    - 3.2.2. Locate and/or Download content from email / cloud services (**instructions**)
  
  - 3.3. Transfer files to Data Drive  
*Note: Determine the appropriate tool to use for transfer based on the format of the media that contains the materials. (Link to Draft)*
    - 3.3.1. Use Data Accessioner (DA) if possible (**instructions**)
    - 3.3.2. Use TeraCopy if DA is not possible (**instructions**)
    - 3.3.3. Perform a logical transfer if unable to use Data Accessioner or TeraCopy. (instructions?? - something to document what you started with and what was moved??)
    - 3.3.4. Depending on the media, creating a disk image is another option (**instructions** - however this has not been fully tested)



*Note:* If there are multiple pieces of storage media in the accession, you may want to complete transfer of all before moving on to step 3, especially if TeraCopy is being used for any of the files. This allows creation of a report from DROID that covers the entire accession folder and have that information in one place.

4. Generate documentation (**information/instructions**) and Collect in \_AccessionInfo folder  
Note: If you notice .\_Ghost files, .DS\_Store files, or Thumbs.db files immediately after transfer, remove these prior to generating additional documentation. (**instructions**)
  - 4.1. Data Accessioner reports (created at the point of running the program; xml file)
  - 4.2. If TeraCopy or logical transfer was used for file transfer, run DROID (**instructions**) and collect documentation.
  - 4.3. Run HashMyFiles (**instructions**) and collect documentation on the entire accession.
5. Check for .\_Ghost files, .DS\_Store files, and Thumbs.db files
  - 5.1. If found, remove them (**instructions**)
  - 5.2. Rerun previously completed documentation - not including Data Accessioner.
6. Check for PII/SEI
  - 6.1. Using Identity Finder (**instructions**)
    - 6.1.1. Save reports before and after clean up with Identity Finder
    - 6.1.2. Removed File Log\_ **template** (google Doc) **\*\*Template** (google sheet - preference) (*DRAFT - procedure not set*)

## **B. Processing**

*Note: We have not spent much time addressing processing to this point. This is part of our focus for the next iteration of the Electronic Records Task Force. Below represents current thinking but has not been tested.*

7. MORE PROCESSING?
  - 7.1. What specifically is needed here? How do we move from the SIP to the AIP. We can't keep everything. Who has time to address issues and concerns? Task Force members? Curators and unit staff?
8. Complete reports for processor and curator to assist with appraisal and curation. [This can be started earlier in the process but needs to be completed before transferring content to Q and talking with the curators.]

- 8.1. Create Folder for Collection at the top level of this Google Folder  
[Electronic Records Task Force 2014 / ASC Electronic Acquisitions]
    - 8.1.1. Name with Collection Number and/or Name
  - 8.2. Processing Documentation (**link to template**)
    - 8.2.1. Information for the processor to document steps they took and problems they might have had.
  - 8.3. Inventory Report (**link to template**)
    - 8.3.1. Information for the Curator to review that provides a summary of the materials, questions, and thoughts for next steps.
  - 8.4. Email Curator location of reports
9. Review materials in consultation with curator
    - 9.1. Transfer content to Q Drive
    - 9.2. Consult with curator
    - 9.3. Notify if any sensitive information was found and still on the Data drive
  10. Make decisions on what to keep or not keep
    - 10.1. File clean up
      - 10.1.1. ghost files, DS store files, etc
        - 10.1.1.1. Delete by hand
        - 10.1.1.2. Need to write instructions for command line clean up
      - 10.1.2. duplicates
        - 10.1.2.1. Duplicate File Finder
    - 10.2. Folder/File Organization  
It is ok to move files around.
    - 10.3. Folder/File Renaming  
It might be beneficial to rename folders and/or files to assist researchers in understanding the collection being presented to them.
      - 10.3.1. By hand
      - 10.3.2. Using Tools
  11. Determine goals on how to make accessible
  12. Rerun reports on final set of materials
  13. Set up Fixity checking on final set of materials
    - 13.1. (**Fixity** Instructions)
    - 13.2. Exact File Instructions
  14. How do we follow up? What is our responsibility to follow up on?

## Appendix 5: Policies

### **Moving/Copying Files between Drives/Folders**

[1/23/2015] When transferring files from one location to another after initial transfer to ERTF workspace it is recommended to use TeraCopy. Using TeraCopy allows us to verify the file transfer process of moving or copying files as well as preserve 'original' file dates. [*Instructions on using TeraCopy were created.*]

### **Original Hardware**

[1/12/2015] In general, we are not interested in collecting original hardware. In some instances, it may be necessary or we may prefer to work with original hardware to access older digital media. In this case, we would dispose of or return in accordance to the procedure for 'Original Media'. It will be up to the collecting unit if original hardware will be retained as part of the collection [as an object in itself].

### **Original Media**

[1/12/2015] Digital material that comes to ASC on physical media will be copied off the media during the file transfer/appraisal/accessioning process. After this process is complete, the media will be securely disposed of. Alternatively, if requested and agreed upon in writing, the original media will be returned to the donor after the initial processing is complete.

### **Separation Sheets**

[2/19/2015] When working with media that is already part of an accessioned collection (the disk in box scenario), a separation sheet should be used. Place a separation sheet in the box indicating what has been removed. If after appraisal, the digital content is transferred to the 'digital archive', the finding aid should be updated to indicate digital files (rather than media) are part of the collection, and the media removed from the box following the Original Media Policy.

## Appendix 6: Accession Log and Instructions

The Accessioning Log and Instructions were developed as a starting point for surveying and tracking incoming born digital material and used by the Task Force to identify material on which to test workflows and procedures.

### ASC Electronic Records Survey Instructions

ASC Staff is asked to complete a survey about incoming electronic records to assist in initial records processing. The following describes the requested information for each set of records as listed on the ASC Electronic Records Interim Accession Log. [Note: This log can also be used to address collections in which digital materials already exist and are ready to be addressed as such.]

There are two main tabs in the Interim Accession Log. ASC Staff should fill out the first tab - ASC To Review. ERTF staff transfer information from ASC To Review tab to the ASC Reviewed tab after records have initially been processed. Information is the same on both of these tabs.

### Field Descriptions

The following fields should be completed by unit staff prior to or immediately after receiving electronic records.

**Original Date Entered into Log (yyyy-mm-dd):** Enter in the date you fill out the Accession Log. This information will be useful for ERTF staff to better understand how much lead time ASC staff receives prior to a donation of electronic materials as well as prioritizing collections to work on.

**Unit Accepting Material:** Select your unit from the dropdown list.

**Unit Contact:** Enter the name of the person ERTF should contact to discuss the processing of the entry.

**Accession Number:** Enter the accession number of the collection the entry relates to if known. If unknown, leave blank. Add this information in when known.

**Title:** Enter in the title of the collection.

**Creator:** Enter the name of the creator of the records.

**Donor/Who transferred From (if different from creator):** Enter the name of the person/entity who donated / transferred the records if different from the Creator.

**Brief Description:** Enter in any information you have about the collection that might be useful for processing electronic records specifically. This could include information about the number of files/folder, the media you expect (CDs, Hard drives), operating system (mac or windows), or they types of files or what the records document and how it relates to the rest of the collection (if other materials are available.) [Some of this information is asked for in other fields, but could be expanded on here.]

**Size:** Enter in the total size (or approximate size) of the accession if known in GB, MB, TB (whichever is most appropriate for the donation).

**File Formats, if known:** Enter in the expected/known file formats. Word, Excel, PDF, Tiff... etc...

**Proposed Date of Transfer:** Enter in the proposed date of transfer if the transfer has not yet happened. Ex: Spring 2015, September 2015.

**Date of Transfer:** Enter the date that files were transferred to your possession.

**Method of Transfer:** Enter the media type or method of file transfer. EX: Hard Drive, Flash Drive, Email, Dropbox...

**Initial Storage Location (path):** If you transferred the files yourself and placed them somewhere other than the original media list that location here. This practice should not be common place as ERTF staff should be consulted for the initial transfer. This location might also refer to the location in which media (from an already accessioned collection) exist such as a box location.

**Notes, Concerns:** Enter in any information you would like to pass on to ERTF staff about concerns you have about the files/accession/transfer process. ERTF staff will address these concerns with you prior to working on the collection.

**Link(s) to more info/inventory, if applicable:** If other documentation exists, list it here. ERTF staff will discuss this information with you prior to working on the collection.

These next set of fields are used by ERTF staff during/after initial transfer is complete

**Comments:** ERTF staff will document when the collection was reviewed/transferred and by whom. Other information may also get recorded in this field.

**Working Storage Location (path):** ERTF staff will record the file location where Unit Staff can find the records to further review and process.

**Completed Storage Location (path):** ERTF staff will record the final file location of the materials after Unit Staff and ERTF staff have completed processing. This location is what will be considered the 'complete' package that needs to be cared for long-term.

**Initial Discussion with Staff (?):** ERTF can document when the initial conversation happened with staff. [This field is new and under review for usefulness.]

**Follow Up Discussion with Staff (?):** ERTF can document when the follow up conversation/s happened with staff. [This field is new and under review for usefulness.]

**Date Transferred to Reviewed (?):** ERTF can use this field to document when the accession line was moved from the first tab (ASC To Review) to the second tab (ASC Reviewed) to track progress and efficiency. [This field is new and under review for usefulness.]

**Access Plan / Link to Finding Aid (?):** ERTF can use this field to document how access will be provided to this content. If possible, a link to the associated finding aid should be recorded. [This field is new and under review for usefulness.]

## Appendix 7: Three Phase Work Plan

### Phases of Work for WorkGroup

#### Phase 1: Testing (July-September 2014)

- Reviewing tools and see what 'products' they produce and how they can work to assist us with processing
- Use the template to document research.
- Use the calendar (Digital Processing Workstation) to schedule time (can invite others to join you if you like - on the working group or a curator if you want to show them the process)
- Use the log to record what type of testing we did - just to see what others are working on. Might help determine if we need to address specific issues (cover gaps).
- *DELIVERABLE: High level standards / needed outcomes / what tools might help [complete]*
  - Folder full of testing information
  - Chose to work with Data Accessioner, DROID, and HashMyFiles (Spider/Identity Finder - decision pending)

#### Phase 2: Drafting Workflows (October-December 2014)

- [For use prior to this as well, the existing draft workflow/process documents]
- Work on process for using tools, when and where they are used, in what situations.
- Test these
- *DELIVERABLE: Draft workflows (multiple situations) (Processing Instructions Draft)*
  - CDs, DVDS, hard drives (mac/windows)
  - Disk Image (drafted)

#### Phase 3: Finalizing Workflows/Suggestions (January-March 2015)

[project wrap up May 2015]

- Edit and modify draft workflows. Continue to polish processes to make it as smooth as possible.
- *DELIVERABLE: "Finalized" workflows documented in User Guides [folder]*

## **Appendix 8: Donor Guides**

Donor guides were created to provide guidance and background information for discussing potential donations that include electronic records. One version was created for staff and another version that could be given to potential donors to review for themselves.



## Accepting Electronic Records: Staff Guide

Accepting electronic records into your collections brings with it new preservation and access opportunities and challenges. Having answers to and/or discussing the following with your donors will assist in processing collections that include electronic records.

Questions? Ask the Electronic Records Task Force! ([lib-ertf@umn.edu](mailto:lib-ertf@umn.edu))

### Discuss the following with potential donors to better understand the records:

- What can you tell me about the files? [*Looking for context.*]
  - When were they created? In what context? By whom?
  - In what computing environment? (*make, model, operating system*)
  - Have they been modified? How? Why?
  - Where are they currently located? Have they been moved from their original location?
  - What file formats are they in? (*see below*)
  - How are files related to each other or other materials in the collection? (*duplicates*)
- Do the records contain or possibly contain *Personal / Private information*?
  - Social security or credit card numbers
  - Medical / legal information
  - Personally identifiable information
- Are there any associated restrictions?
  - Copyrights (intellectual property concerns)
  - Legal restrictions (sensitive information)
  - Other requested restrictions
- What are the expectations for public access?
  - View online or available for download
  - Request access from ASC
  - View onsite only



<http://digitalbevaring.dk/om-sitet/>

### File Formats Considerations

- While we **can accept** any and all digital files we **will not be able to preserve and/or provide access** to all digital file formats. For this reason we may convert files to another format or opt not to accept some file types.
- Review current common preservation formats and **discuss options with the Electronic Records Task Force**.

(<http://conservancy.umn.edu/pages/policies/#preservation>)

### Transfer Options Considerations

- There are many ways to receive files from donors.
  - External media: hard drive, flash drives
  - Cloud Services: Google Drive, Dropbox
  - Email: Files sent as attachments to download
  - FTP: Connecting to remote servers
- Consider the amount of files, location of the donor, and donor comfort level of each method.
- **Discuss pros and cons of each based on donation characteristics with the Electronic Records Task Force**

### Appraisal of Electronic Records

*We can't keep everything.* Just as with more traditional analog records, electronic records need to be appraised. Electronic records go through both a *content* appraisal and a *technical* appraisal. The content appraisal is best conducted by unit staff; the technical appraisal by the Electronic Records Task Force in consultation with unit staff. Technical appraisal focuses on the file formats.

For content appraisal, ask yourself:

- Does this content fit within existing collecting policies?
- What is the research value of this content? How does it compare with other records in the existing/new collection? Is it available elsewhere? Why should we keep this?

### Final Steps / Follow Up with the Donor

Transfer of records will require a signed deed of gift / donor agreement that addresses electronic records. If applicable, removable media will be returned to donor or handled as agreed upon after the transfer is complete.

## Additional Context and Guidance for Staff Working with Donors of Electronic Records

**The Role of the Electronic Records Task Force (ERTF)** The Electronic Records Task Force was officially launched in May of 2014 to address the ingest and initial processing of electronic records. We have started working with ASC staff to develop ingest and initial processing workflows and are moving towards procedures / policy development. As we do this, and begin to work with you on your collections, we are also here as a resource for you. We need to know what your questions and concerns are so we can address them as we move forward. What are your needs? What are your questions / concerns? What can we help with?

### File Format Considerations

- Not all files formats can be preserved.
- Not all files formats can be accessed.
- General guidelines for formats are to accept 'preservation' ready file formats.
- Often different 'levels of preservation' actions are determined for different types of formats.  
<http://conservancy.umn.edu/pages/policies/#preservation>
- We will **convert** file formats to a more preservation friendly format if necessary; the original format may or **may not be saved**.

\*Specific preservation guidelines will depend on many factors.

### Transfer Media Considerations

How we copy the files to the processing location will depend on the media type:

- Most media will allow for a 'logical' transfer (copying only the files we want)
- Some media may require creating a *disk image* (see below). Note: This process will capture **deleted files**. (Often used for older media such as 5 ¼ floppy disks.)

Methods used to transfer files **may change the created date**. We can work to limit this, but may not always be able to.

- For example, an emailed file that is downloaded will often have a 'created' date of the date it was downloaded. Because of this we are more focused on the 'last modified date'

\* Work with the ERTF to determine the best method

### Resources and Tools

The ERTF has reviewed many tools for working with electronic records.

- What are you trying to do? We might have a suggestion on a tool to make your job easier.
  - Renaming files
  - Finding Duplicates
- Question about a format or media? Can't open a file?

\* **Contact the ERTF!**

### Disk Imaging

Disk Imaging is a method for acquiring access to content on media not accessible by other means. What you should know:

- This method will not be used in most cases
- Deleted files can be recovered
  - Be aware that if the media contains files the donor thought were deleted forever, this method might recover them.

\* We are currently deciding about what to do with disk images that we do create. Should we keep them, or discard them?

### Things to Think About / Be Aware Of

The following concepts should be on your mind as you work with electronic records.

- Enter current/future acquisitions that contain electronic records in the Interim Accession Log ([ASC Website](#))
- Discuss transfer methods and coordinate transfer with ERTF staff (*Files should not be on the L Drive.*)
- Don't over handle digital materials – *opening / moving files could change their metadata!*
- Understand how materials will be stored / preserved / available long-term (levels of preservation)
- Understand the policies and procedures related to acquisition, transfer, copying, embargo/restriction, user access, long-term preservation, and secure disposal of electronic records.

\* **Contact the ERTF to talk about your next electronic acquisition or if you have questions.** ([lib-ertf@umn.edu](mailto:lib-ertf@umn.edu))

# Guide for Donors of Electronic Records

The University of Minnesota Libraries Archives and Special Collections (ASC) considers electronic records an integral part of its collections. We accept electronic material of enduring value created or received by a person, family, or organization. Key issues that you can discuss with Libraries staff are outlined here.

## Content and Rights Information

Discuss and understand the following with Libraries staff if applicable to your donation:

- Copyright / intellectual property concerns
- Personal information (*social security/credit card numbers, etc.*)
- Legal and/or other restrictions
- Contextual information on the how, when, why these records were created and by whom
- Possible recovery of deleted information (*if disk imaging is used*)
- Expectations for access



Social Security #

## File Formats

Be prepared to discuss the file formats of your records. The Libraries can support some file formats better than others. For this reason we may recommend conversion to another format for long-term preservation purposes before or after transfer, or opt not to accept some file types.



The transfer and preservation of email poses special issues that will be addressed in conversations regarding your donation.

## Other Information (Metadata)

The more information you have about your donation the better. This information can be used to assist Libraries staff who process your records and make it accessible to future users. Useful information includes more about:

- Content (*what is it and why is it valuable*)
- Context (*of creation, use – ex: a draft or final copy*)
- Organization (*file structure / file naming conventions*)
- Technical information (*types of computer, software programs*)
- Creator/s name and main use

## File Transfer

There are various ways to transfer files to the University Libraries. Discuss with staff which method is best for you based on the type and size of your donation. Options may include:

- External media such as:  
CDs/DVDs, hard drives, flash drive etc.
- Cloud service (Google Drive, Dropbox)
- File Transfer Protocol (FTP)
- Email attachments



If you have files on older media such as floppy disks or older computer hard drives please let us know.

## Access

Our goal is to ensure that materials transferred to the University Libraries can be made available to the general public. Libraries staff can discuss with you practical expectations for access to your materials based on the donation's content, overall size, file formats, and other concerns.

## Final Steps / Follow Up

The transfer of records requires a signed Deed of Gift and/or a Donor Agreement that addresses electronic records. Depending on the collection, open and ongoing conversations with staff regarding preservation and arrangement of materials may be necessary. If applicable, removable media will be returned to the donor or handled as agreed upon after the transfer is complete.

## Appendix 9: E-Records Transfer Sheet

The Electronic Records Transfer Form was created to assist in the collection of data about electronic records at the time of transfer. A document format and a Google Form format were created based on unit staff/curator feedback.

### Electronic Archival Records Transfer Information Sheet

*Complete and return to Archives & Special Collections (ASC) Staff.*

#### Contact Information for Person Transferring These Records:

Name:

Organization (if applicable):

Phone:

Email:

#### Person to Contact for Questions About This Material (if other than above):

Name (and Title):

Organization (if applicable):

Phone:

Email:

1. Date material is/will be transferred to ASC: TBD
2. What media or method is being used to transfer the material?  
 external hard drive                       Upload to server (e.g. Google Drive or DropBox)  
 flash/USB/thumb drive                       Removable disks (e.g. floppy disks, CDs, DVDs)  
 Other -- please describe:
3. Total size of transferred files (e.g. in MB or GB) (Can be an estimate):
4. Who created these records? Please list names and/or positions or titles of primary individuals.
5. If created as part of a larger organization, what part of the organization (e.g. department, office, program, project) created these records?
6. Time span during which the records were created, or best estimate (e.g. 1995-2008):
7. Summary of file types included, to the best of your knowledge:  
 MS Word (.doc, .docx)                       MS Excel (.xls, .xlsx)                       PDF

- Other word processing  
(.html)  images  
 Sound  
 Other -- please describe:
- Other spreadsheets  
 Databases  
 ZIP or other compressed files
- web pages  
 Video

8. Types of records

- reports  
 surveys  
 newsletter s  
 photographs  
 speeches  
 Other -- please list:
- memos/correspondence/letters  
 grant applications  
 brochures  
 scrapbooks  
 publications
- financial records  
 press releases  
 training materials  
 diaries/journals  
 minutes

9. To the best of your knowledge, do these records contain any information that is sensitive or private?

- social security numbers  
 financial data  
 health data
- current or recent home addresses  
 proprietary info  
 other -- please describe:

Please indicate where in the collection any such data may be located:

10. Are there any restrictions or concerns about who might have access to this material? If so, describe.

11. Any other notes, context, or helpful information on these files.

## Appendix 10: Deed of Gift Addendum for Electronic Records

Addendum to the standard Archives and Special Collections Deed of Gift for use with collections that include electronic records.

### **DRAFT - University of Minnesota Libraries Deed of Gift Addendum for Electronic Records**

This addendum addresses the particular needs of electronic records being donated to the University of Minnesota Libraries.

I also understand that upon transfer, the University of Minnesota will become the custodian of the donated materials. I agree not to donate, distribute, or sell these materials, or substantially similar versions of them, to another entity or institution. I understand that if in the future I wish to deposit substantially altered versions of these materials to another entity or institution, the University of Minnesota has the right of first refusal.

#### **I. Discretionary practices specific to electronic records**

The Libraries, at its discretion, in accordance with University policy and with applicable law, may conduct the following with the Donated Materials:

- Disable or bypass passwords or encryption systems, if any, to gain access to the Donated Materials.
- Recover deleted files or file fragments, if any, and provide access to these materials.
- Provide access to log files, system files, and other similar data that document use of computers or systems, if any are received with the materials.

Subject to the terms and conditions, if any, stated below:

---

---

---

---

#### **II. Privacy**

The Libraries will review the materials in the collection in an attempt to identify items that contain personally identifiable information and/or private information (i.e., as defined by all applicable state and federal laws and regulations). Please indicate (initial) whether or not these materials that contain private electronic information.

\_\_\_\_\_ To the best of my knowledge, these materials *do not* contain private information.

OR

\_\_\_\_\_ To the best of my knowledge the materials *are likely* to contain private information. Please check all that may apply:

- Social Security numbers
- Passwords or PINs
- Credit Card numbers
- Financial records
- Medical records
- Other materials that have specific privacy concerns -- please specify:

---

---

---

---

### III. Disposition of Storage Media

Computer hardware and/or removable media that is not being retained as part of the collection will be securely removed and/or destroyed after the transfer is complete, unless you prefer the media to be returned to you. Please indicate (initial) your preference.

\_\_\_ I DO NOT want computer hardware and/or removable media returned.

OR

\_\_\_ I DO want computer hardware and/or removable media returned to the following:

Name: \_\_\_\_\_

Address: \_\_\_\_\_

Phone: \_\_\_\_\_

Email: \_\_\_\_\_

SIGNATURE OF DONOR:

\_\_\_\_\_ Date:

\_\_\_\_\_  
Name:

Title: Donor

\_\_\_\_\_ Date:

\_\_\_\_\_  
Name:

Title: Elmer L. Andersen Director of Archives and Special Collections