*Public report created August 23, 2017*
*Original report reviewed by UMN Libraries Cabinet May 30, 2017*

# Electronic Records Task Force Phase 2 Final Report

## Documentation of the University of Minnesota Electronic Record Task Force's Second Year of Activities

August 23, 2017

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

*Task Force Members*
Lisa Calahan (co-chair), Carol Kussmann (co-chair), Valerie Collins, Kate Dietrick, Kevin Dyke, Lara Friedman-Shedlov, Betsy Friesen, Lisa Johnston, Mary Miller, Erik Moore, Arvid Nelsen, Jon Nichols, Michael Sutliff, and Amanda Wick.

Abstract:
The University of Minnesota Libraries sponsored the Electronic Records Task Force Phase 2 to monitor established workflows for electronic record ingest and to develop new workflows, policies, procedures and mechanisms for processing and providing access to users. This report documents the work of the Task Force during Phase 2 and provides recommendations for both short and long term sustainability.

Suggested Citation: Calahan, Lisa; Kussmann, Carol; University of Minnesota Libraries Task Force.. (2017). Electronic Records Task Force Phase 2 Final Report. Retrieved from the University of Minnesota Digital Conservancy. http://hdl.handle.net/11299/189543

# Table of Contents

# Executive Summary

Based on the Electronic Records Task Force (Phase 1) recommendations made in 2015[1], the Electronic Records Task Force (ERTF) Phase 2 spent the year of 2016 working with electronic records within the University of Minnesota's (UMN) department of Archives and Special Collections' (ASC) as well as the University Libraries' general collections. The ERTF was charged with establishing protocols for the processing of, and access to, electronic materials. To this end, twenty-nine ASC collections were processed and described in published finding aids, and 177 issues of a previously hidden electronic journal are now available through the library catalog.

To meet this year's deliverables, the ERTF Processing Sub-Group focused on the development of processing activities best practices and guidelines. The group updated existing processing levels to include electronic records, created guidelines for working with restricted materials, created descriptive guidelines, and continued to track the amount of time spent on ingest and processing tasks. One of the unexpected discoveries of tracking the processing efforts was that the assigned level of processing was not reflective of the level of effort spent on processing tasks.

Based on processing level or collection need, workflows were used to determine which processing activities would be carried out on a collection. The developed processing workflow lists eight actions that may be taken on a collection. The end result of the workflow is the creation of an organized collection that is described to users through an online resource. This online resource, a finding aid, provides preliminary access to users by describing where and how the records can be accessed.

Finding aids, published online through ArchivesSpace, are the first step in providing access to electronic records. Access methods explored by the Task Force include direct links to digital objects in existing repositories from a finding aid, providing on-site access in the reading room, or making records available through various services such as email, UMN Google Drive, or Box Secure Storage (a newly UMN approved enterprise system for storing, sharing and accessing sensitive and "private-highly restricted" files). When exploring and subsequently recommending these access method options the Task Force followed the University's security protocols for materials that contained "private-highly restricted" and "private-restricted" information and made sure to provide access to material only through appropriate means. After researching the access methods currently in use and existing user needs based on the current volume and limited number of e-records listed in finding aids, it was decided to continue to make collections available on an as-needed basis using the available access platforms, rather than implementing

[1] Kussmann, Carol; Nelsen, R. Arvid; University of Minnesota. Electronic Records Task Force. (2015). Electronic Records Task Force Final Report. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/174097.

a global access method at this time. Future needs will be evaluated as more collections are made accessible and user demands increase.

The Access Subgroup worked with staff outside of Archives and Special Collections (ASC) to uncover electronic materials held by the Libraries not currently available to the public. Many resources were discovered and a successful pilot project carried out by the E-Resource Management department made 177 issues of a previously hidden title available for checkout.

The long-term preservation of electronic records in ASC remains a concern and continues to be a work-in-progress. Currently, due to various security, technology, and processing requirements for different materials, there is *no single method* in place to securely and consistently back up and manage entire collections in the same location. Solutions supporting various aspects of the archiving and access process have been discussed but not implemented, as determining an appropriate digital asset management and preservation environment depends on many factors, including available technologies, collection readiness, level of security and access controls required, and cogent organizational policy. With greater awareness of requirements for electronic materials and their management, additional exploration of feasible architectures and methods for securing, backing up, monitoring, and providing appropriate levels of access to these materials must be done to ensure that an effective long-term solution is put in place.

Individual ERTF members spent roughly between one hour and 20 hours a month on ERTF tasks. Overall, 502 hours of tracked time was spent ingesting and processing electronic records, which does not include other tasks, such as group monthly meetings. To keep up with the influx of electronic records, the Electronic Records Task Force provides the following recommendations:

1. **Staffing**: Hire a permanent full time employee to work exclusively with electronic records within a centrally-positioned processing role. The past two years show that consistency leads to efficiency, and we anticipate the work in this dedicated capacity will grow as our accrual of electronic records continues to increase.
2. **Long-term Management**: Create an Electronic Records Management Group (ERMG) to address ongoing needs with ASC electronic records.
3. **Preservation**: Review current workflows and long-term management requirements to address immediate and long-term solutions for file backup, recovery, and preservation in compliance with the Libraries' Digital Preservation Policy Framework and University policies.
4. **Security of the Electronic Records Workstation**: Conduct a thorough review of security requirements for the workstation including physical security, device encryption, network firewall security, authentication, and authorization to the workstation and its contents. Ensure that when the workstation is upgraded it will be configured to comply with University Information Security Standards.
5. **Equipment**: Establish initial and ongoing financial support for hardware and software for the workstation as well as to preserve and provide access to the electronic materials in

the collections, including adding the workstation computer to the Libraries' Computer Replacement Cycle.

6. **Access to Archives and Special Collections Materials**: Explore options for providing access, via a Virtual Reading Room, to electronic records that require mediation.
7. **Access to Library Materials**: Create a committee separate from the proposed ASC Electronic Records Management group to further investigate ongoing needs for Libraries' purchased electronic materials. This committee would address both access and preservation of these materials. Noting the different needs and issues of published electronic materials and general ASC materials we suggest separate groups to allow for more focused work.

Advancement of these recommendations would, to the extent possible, leverage existing resources, but also call for new investments in staffing and equipment. The full articulation of these recommendations and preliminary resource requirements are provided in the recommendations section of this report.

## Summary of Task Force Activities

The Electronic Records Task Force was initially chartered and launched in May 2014 to address the immediate need to safely and securely acquire, transfer, and stabilize unique electronic records that existed in a multitude of file formats and on a host of contemporary and legacy media. In 2016, a second iteration of the Task Force[2] addressed the appraisal, arrangement and description, and access to electronic materials acquired by the University of Minnesota Archives and Special Collections (ASC) which included documentation of access needs of the Libraries' collections in electronic format that were not yet accessible to users.

ERTF was a collaborative effort comprised of staff from a variety of Libraries' departments and ASC units. The Archives and Special Collections department has 15 collecting units each with their own individual collecting scopes and staff, which lends itself to diverse donor and user expectations, as well as differing levels of expertise and experience among staff for managing electronic records of archival value. The Task Force members[3] during the second year (2015-2016) were selected for their technical, archival, preservation, and cataloging expertise from across the University Libraries:

- Lisa Calahan (co-chair), ASC, Central Processing
- Valerie Collins, ASC, University Archives/University Digital Conservancy
- Kate Dietrick, ASC, Central Processing Unit, Upper Midwest Jewish Archives
- Lara Friedman-Shedlov, ASC, YMCA Archives

[2] The full Project Charter is provided in Appendix A.
[3] Note that some of these members were only active for part of the year due to staff turnover and others stepping in to fulfill vacant roles.

- Betsy Friesen, Data Management and Access
- Carol Kussmann (co-chair), Digital Preservation and Repository Technologies
- Mary Miller, Collection Management and Preservation
- Erik Moore, ASC, University Archives/University Digital Conservancy
- Arvid Nelsen, ASC, Charles Babbage Institute
- Amanda Wick, ASC, Charles Babbage Institute

Resource personnel included Kevin Dyke (Spatial Data Analyst/Curator), Lisa Johnston (Research Data Management/Curation Lead), Jon Nichols (Digital Preservation and Repository Technologies), and Mike Sutliff (Technology Support and Consultation). John Butler (Associate University Librarian for Data and Technology) and Kris Kiesling (Director, Archives and Special Collections) served as the Task Force Sponsors.

As a whole, the Task Force met monthly. Two working groups were formed to address specific issues and met more frequently. Working group members consulted with colleagues outside of the Task Force; most often, but not exclusively with, 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 on an as needed basis.

The Processing Working Group addressed the tasks and deliverables that focused on the hands-on work with the electronic records. This included monitoring the existing ingest workflow, developing the processing workflow, and carrying out processing tasks. Additional tasks undertaken by the group were defining processing levels and defining critical elements to be included in an Archival Information Packet (collection items and information about them).

The Access Working Group addressed the access needs for Archives and Special Collections materials as well as for the Libraries as a whole. These activities produced a matrix documenting the existing access methods and additional brainstormed needs for the future. The group also observed a pilot project that was implemented by the E-Resource Management department to provide access to electronic library materials through the Libraries' catalog.

Pulled directly from the Task Force charge, specific tasks and related deliverables are listed below. Deliverables are described in more detail directly in the Detailed Description of Task Force Activities section of this report, and when appropriate provided as an appendix.

## Project Tasks and Deliverables

***Task 1: Develop Workflows for Processing Ingested Collections***

1. Develop workflows and role assignments for processing work, including appraisal, arrangement, and description.
2. Define minimal requirements for an Archival Information Package (AIP).
3. Determine short and long-term human resource requirements for completion of curatorial and technical work.

***Task 2: Define Processing Levels***

1. Define levels of processing (minimal, intermediate, full).
2. Understand the relationship between file formats, processing levels, preservation capabilities, and document requirements/concerns as needed within the broader context of the Libraries' digital preservation program.
3. Assign processing priorities based on processing level, preservation needs, and anticipated user needs.

***Task 3: Develop Access Methods that Address End-user Needs, Copyright, Data Privacy and other Information Security Requirements***

1. Document the spectrum of access requirements that meet user needs.
2. Identify and document the spectrum of distribution and access control requirements as related to copyright, data privacy, and other relevant information security policies and/or laws.
3. Document and expand protocol for using *existing* methods of access, including onsite capabilities and currently available online repositories. (e.g., UMedia, UDC, HathiTrust, etc.)
4. Identify unmet end-user needs and, where feasible, plan and implement new methods for access that address user needs.

***Task 4: Monitor Ingest Workflows***

1. Adjust ingest workflows as necessary.
2. Edit related policies and procedures as necessary.

***Task 5: Monitor Preservation Repository Development (when available)***

1. Work iteratively with the Digital Preservation and Repository Technologies staff on repository implementation and related development.
2. Monitor and assist with digital preservation policy development as needed.

All of these tasks, with the exception of Task 5, were completed and are addressed in the recommendations below. Task 5 was not completed because the preservation repository is not yet available. Digital Preservation and Repository Technology (DPRT) staff participated in the work of the Electronic Records Task Force; going forward DRPT and ASC will continute to discuss preservation and access options of the archival material being collected.

# Detailed Description of Task Force Activities

The work of the Task Force is part of an ongoing activity that the Libraries has committed to support and the work completed by the Task Force during 2016 adds to the growing knowledge base of staff working with electronic records. The following sections detail activities specific to this year's goals with explanations on how the Libraries can further support the efforts of long-term access and preservation of electronic records.

## Defining Archival Information Package (AIP) Requirements

One of the first tasks the group completed was to define the minimal requirements for an Archival Information Package (AIP). In the simplest terms, an Archival Information Package is what needs to to be kept for the long-term preservation of and access to the electronic files in collections. ERTF created working definitions for Submission Information Packets, Archival Information Packets, and Dissemination Information Packets (Appendix B). The AIP definition provided here is a summary of our working draft based on current resources and available storage options.

> ***AIP: Archival Information Packet***
> *Archival Information Package (AIP): An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within a system. The AIP often consists of the original files deposited, processed versions of data files and documentation, normalized files, and associated metadata.*[4]
>
> - Ingested files after they have been accessioned/processed into a new or existing collection.
> - Information about the files from tools used during accessioning/processing including:
>   - Reports generated from tools that document checksums, file structure, file name, file size, file extension, date created, date modified, file format, and the identification of any personally identifiable information.
>   - Current required reports are from HashMyFiles, Data Accessioner OR DROID, and Identity Finder if applicable.

Task Force members anticipate that in the future this definition might change as different methods or systems that may be used to manage archival collections may require different information or formatting.

## Processing Activities

The processing of electronic records requires in-depth knowledge of the context of the records, the historical importance of the material, a sense of how they were organized by the creator, and how they may be used by researchers in the future. Processing activities for electronic records may include information gathering and discovery about the context of the collection as well as the delivery media and file type, careful investigation of PII or other sensitive data, the reorganization of files, performing tasks to ensure the preservation of the files, and providing a description of the files in order to make them accessible to users. To complete the deliverables relating to workflows and processing, the ERTF defined processing levels, addressed the level of effort required for processing, and investigated how to set consistent processing priorities.

[4] Glossary. University of Minnesota. 2014. https://www.lib.umn.edu/dp/glossary#A

## Defining Processing Levels

In conjunction with defining the Archival Information Packet, the Task Force defined levels of archival processing for born digital material to better inform processing workflow steps. Processing levels are assigned to all accessioned collections based on the complexities and anticipated research value for an individual collection. Assigning a distinct level of processing provides a clear pathway for the expected and necessary processing tasks to make a collection available to researchers. The Central Processing unit had already established processing levels for analog collections, and it was the task of the ERTF to define how electronic records could fit into these broad categories, based on the perceived complexities of processing born digital records.

The ERTF defined how electronic records fit into three categories based on professional best practices articulated in *Describing Archives: A Content Standard* (DACS).[5] In summary the levels used by Central Processing are:

> **Minimal:** There will be no file arrangement or renaming done *for the purpose of description/discovery* enhancement. File formats will not be normalized. Action will generally not be taken to address duplicate files or Personally Identifiable Information (PII) information identified during ingest. Description will meet the requirements for DACS single level description.
>
> **Intermediate:** Top level folder arrangement and top-level folder renaming *for the purpose of description/discovery* enhancement will be done as needed. File formats will not be normalized. Some duplicates may be weeded and redaction of PII done. Description will meet DACS multi-level elements: described to the series level with high research value series complemented with scope and content notes.
>
> **Full:** Top level folder arrangement and renaming will be done as needed, but where appropriate renaming and arrangement may also be done down to the item level. File normalization may be conducted as necessary or appropriate. Identified duplicates will be removed as appropriate and PII will be redacted as needed. Description will meet DACS multi-level elements: described to series, subseries, or item level where appropriate with high research value components complemented with additional scope and content notes.

Full descriptions of these levels can be found in Appendix C.

After accessioning an analog collection staff have better intellectual understanding of what the collection contains and are able to assign a level of processing based on the collection

[5] Describing Archives: A Content Standard, Second Edition (DACS). Society of American Archivists. 2013, revised 2015. http://www2.archivists.org/standards/DACS

assessment. With analog materials this level usually corresponds to the level of effort being put forth. What ERTF found, however, is that the assigned level of processing did not necessarily correlate to the level of effort when processing electronic records. Based on a number of factors, a collection assigned a minimal level of processing could take days to address, while a collection assigned a full level of processing might only take hours. Therefore the level of processing, while helpful in setting expectations for final description, does not provide accurate insight into the amount of work that needs to be done to make electronic records accessible.

## Level of Effort

To address the lack of correlation between the processing level assigned to an accession and the actual level of effort being given to process the accession, a Levels of Effort document was drafted to help categorize the amount of staff time and resources expended when working with electronic materials. The anticipated level of effort may be more useful for setting priorities than assigned level of processing, as there is a closer one-to-one relationship with the amount of time required to complete the processing.

Results of this work can be reviewed in the Levels of Effort document in Appendix D. This document provides a suggested level of effort based on activities such as file renaming, removing duplicates, addressing PII, time spent in meetings, and time spent reviewing related accessions. Levels are based on the number of files the action is taken on; assuming more time or a higher level of effort is required to work with a larger number of files. With further study, levels of effort could be used to better understand how much work might need to be done with a collection to make it accessible.

## Processing Priorities

When determining the processing priority of electronic records, there are exponentially more factors to consider when ingesting and processing than with analog records. This is in part due to the inherent fragility of born digital records, but in large part due to the additional time requirements needed for collections that include electronic records. Some of the factors that influence priority decisions to ingest and/or process electronic records include:

- **Donor Priorities**: Does the donor want/need the original media back by a certain date? If so, the materials may be ingested sooner but not necessarily processed at a high priority.
- **Collecting Unit Expectations**: Is there a time commitment for the unit to make these records available? Is there anticipated user need for the materials? Collections with an expected higher use value have a higher priority than collections that don't.
- **Collecting Unit Staff Time Commitment**: Does the unit staff have time to work with ERTF? Processing requires input from unit staff. If unit staff do not have time to work with ERTF, the priority to work on their records decreases.
- **Collecting Unit Staff Knowledge Base**: Does the unit staff have the technical knowledge to ingest and process collections for themselves? If unit staff have the skills

to conduct the ingest and/or processing procedures, their own availability (time) is the only restriction to getting materials ingested and/or processed.

- **ERTF Time Commitment**: What are the expectations for the amount of time ERTF members can commit to doing this work? ERTF staff must balance ingest and processing priorities with fulfilling their regular job responsibilities. Collections that are expected to be more complicated may get pushed to the backlog, as ERTF members may not be able to commit significant amounts of time for the work required.
- **Transfer Media Complications**: Is the ERTF familiar with the type of media the content arrived on? Will data extraction take time to research? If staff is unfamiliar with the media, more time and effort will be required to ingest the materials, which may lower the priority level.
- **File Transfer Complication**: Is access to the types of files being offered able to be provided? If not, will there be time to explore how to make files accessible? If the ability to provide access to certain file types does not exist, the priority for ingest/processing might be lowered.
- **File Preservation**: Are the types of files being offered by a donor well supported or are there preservation issues to consider? If the ability to support or preserve certain file types does not exist, the priority for ingest/processing might be lowered or increased based on preservation need.

Ultimately, these questions influence processing and time management decisions when assigning a level of processing and an anticipated level of effort to collections. The revised processing levels and analysis of the level of effort are our attempt to create internal guidelines in the absence of professionally accepted standards.

Due to the time requirements, prioritization strategies, and the scope of the Task Force, members have not addressed electronic records that were acquired prior to the existence of the Electronic Records Task Force - the backlog of disks, CDs, etc., that are sitting in boxes. At this time, ERTF suggests addressing legacy media only when a researcher requests the specific material or the collection is otherwise identified as a priority for additional processing.

## Workflows for Processing Ingested Collections

With definitions set for the Archival Information Package and the Levels of Processing, the Task Force began to refine the workflow previously created for ingesting collections and added specific steps for processing actions. Main tasks that needed to be addressed were appraisal activities, physical processing actions, and the arrangement and description of incoming accessions.

Due to the complexity of processing activities the Task Force also explored who might be best suited to complete each task. More often than not it was found that work might be done by various staff depending on an assortment of factors. As a result, work was often done as a collaboration between ERTF and ASC collecting unit staff. The following workflow (Figure 1)

shows the interactions between ERTF and unit staff throughout the process of accepting electronic records and processing them. Note the number of steps that are done jointly (the right column in Figure 1).

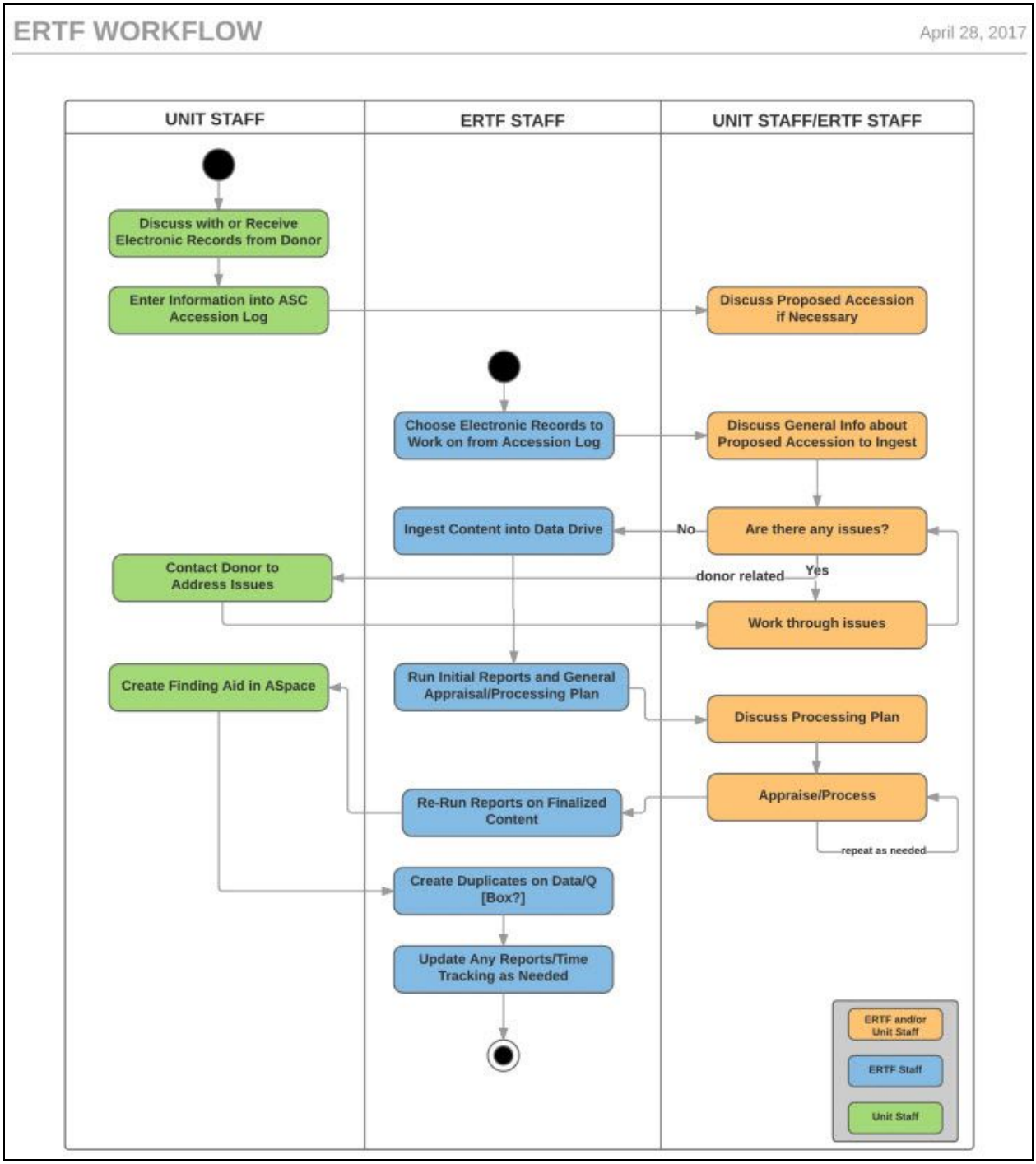



**Figure 1: Workflow Diagram Highlighting Roles and Responsibilities of Unit Staff and ERTF Staff**

Focusing specifically on the division of roles and responsibilities for processing activities (Figure 2), it is evident that the processing actions are completed collaboratively, while the AIP creation/addition and description, which is a bit more specialized in nature, is primarily completed by ERTF members.

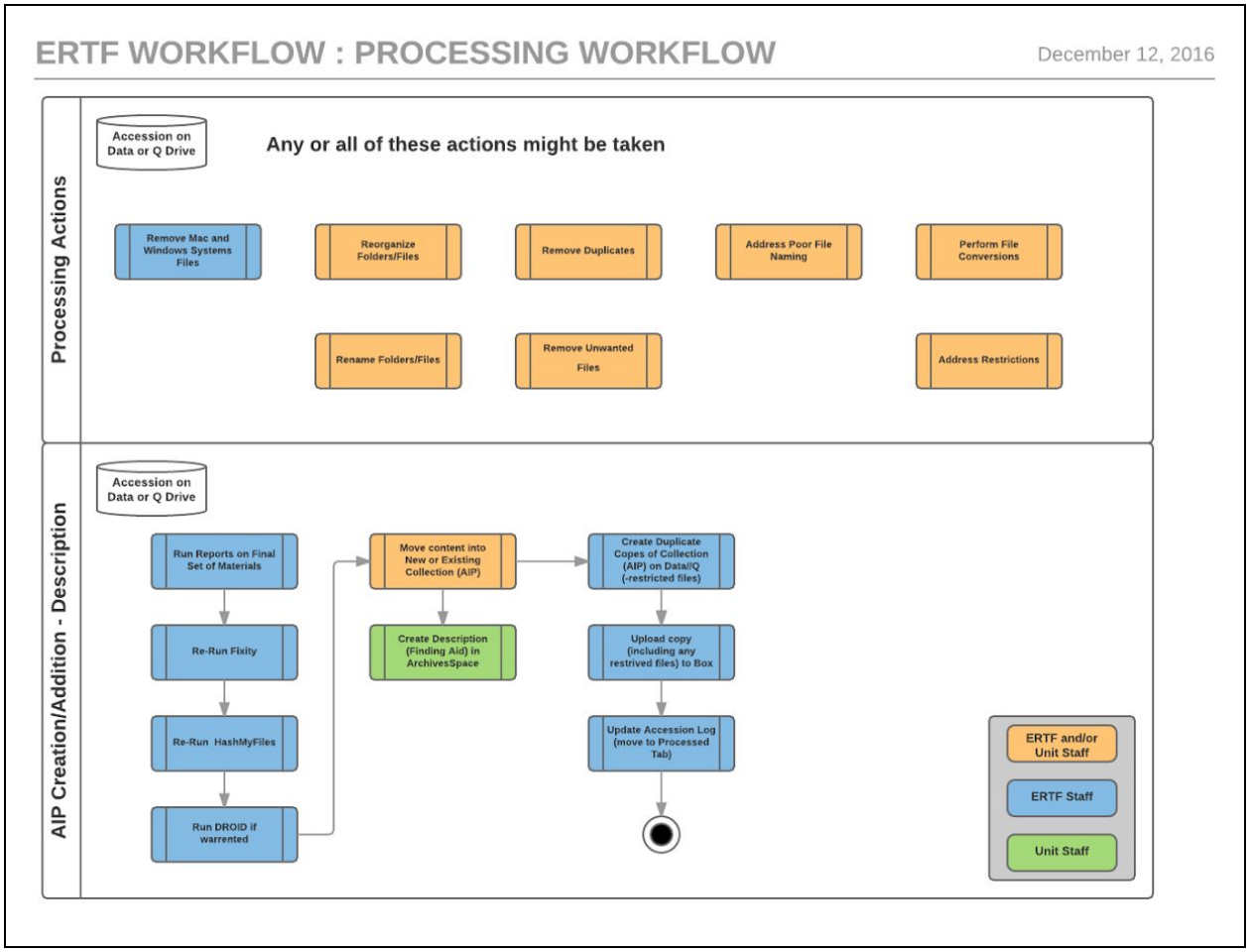



**Figure 2: Actions to be Taken During Processing by ERTF and Unit Staff**

The size, complexity, and existing organization of incoming accessions are the most substantial factors to consider when determining who is best suited to appraise and process the collection. ERTF members have knowledge about the long-term sustainability of file formats, issues related to file names (such as bad characters or file names being too long), and other factors that may affect long-term preservation and access to the collections. Unit staff have knowledge about the collection as a whole, the content and contextual importance, as well as an understanding of the archival value based on their interactions with the donor and the unit's collecting scope. Understanding both aspects is required in order to appraise and process the collection in the most appropriate manner. However, finding time to consult and work through any concerns with unit staff was a significant barrier to efficiently processing collections.

While adding steps to the workflow, the Task Force also monitored and adjusted the Ingest portion of the workflow, completing Task 4 of the project task and deliverables. Minor adjustments were made to the ingest portion of the workflow to account for the order in which to accomplish tasks. Main components and their current order are shown in the Ingest Workflow diagram in Appendix E. The Master Processing Workflow Document (Appendix F) was also edited to reflect these changes as well as all new processing activities. To assist in general task tracking a "general task checklist" for the main components of the workflow was created.

## Summary of Ingest and Processing Amounts and Time[6]

Utilizing the developed workflows, ERTF members successfully transferred 39 accessions from external media to the workstation computer between July 2015 and December 2016. Twenty-nine of these accessions were processed as 27 collections (two accessions were combined during processing) containing a total of 1438 GB; the remaining 10 accessions, an additional 467 GB, were ingested but not processed for various reasons, including unit priorities and available time. At the end of December 2016, the backlog included 35 accessions that were not ingested or processed.

Over the course of these 18 months, staff ingested and processed about 240[7] GB per person and spent 503[8] total hours ingesting and processing incoming material. To put this in perspective, 520 hours is 25% of a full time staff person. On average the Task Force estimated that for ingesting and processing collections, members were able to work at a rate of 3.5 GB per hour, and estimate that it took an average of 16 hours to ingest and process each collection.[9] These results are summarized in Table 1.

| # of Staff | # Collections | GB | Time |
|---|---|---|---|
| 6 people | 27 collections ingested and processed | 1438 GB | 421 hours |
| [5 people] | [10 collections ingested] | [468 GB] | [82 hours] |
| **Totals** | 37 collections | 1906 GB | 503 hours |
| **Averages** | For the 27 collections ingested and processed | ~ 3.5 GB per hour[10] | ~16 hours/collection[11] |

**Table 1: Summary of Collection Processing Work**

[6] Appendix G provides a summary of calculations used in this report.
[7] 1438 GB / 6 people = 239.6 or 240 GB per person.
[8] 421 + 82 = 503 hours total amount of time spent on electronic records
[9] These averages were calculated using the statistics for the 27 collections that were ingested and processed only. These do not include the 10 collections that were only ingested. This is because to be considered complete, collections must be ingested and processed.
[10] 1438 GB / 421 hours = 3.415 gb/hour or about 3.5 gb/hour
[11] 421 hours / 27 collections = 15.59 hours/collection or about 16 hours/collection

Due to the nature of the material and the tasks performed during ingest and processing, a collection may be ingested but not immediately processed. Circumstances that lead to a collection not being processed immediately after ingest could include anticipated low research demand, lack of available staff, funding, specialized knowledge, need for further direction from the collecting unit, or the prohibitive amount of time needed for a complicated collection.

Although Processing Subgroup members fastidiously tracked how time was spent on specific processing activities, not all time spent working with electronic records was recorded, and certain activities were just difficult to track. Known gaps in tracked time include time spent moving collections back and forth between drives after processing (as this is considered part of access), and time spent by unit staff on activities not otherwise easily tracked. In addition, the time recorded for meetings usually only reflects the time for recorded by ERTF staff and not the time for other participating unit staff. Therefore the total time spent by all staff listed under 'Meeting' could be doubled, or tripled in some cases, depending on the number of participants. While time tracking provides a good estimate of the time spent on each collection, it does not accurately reflect the *total* time spent working with electronic records. Nevertheless, documenting time for each activity is valuable because it provides a better understanding of what activities take the most time and can help make projections regarding future accruals.

## Analysis of Ingest and Processing Work[12]

During phase 2 of ERTF, we ingested 39 and processed 27 accessions, doubling the amount of ingested accessions from the previous year. This success can be contributed in part to staff's ability to use and build upon the existing ingest workflow, their gained knowledge and experience with the processes and workflows, and an additional Task Force member assisting with processing collections. Although ERTF doubled the ingest levels during the second phase of ERTF, the overall backlog only decreased by two collections. This is a good indication that the number of collections being acquired with an electronic records component will continue to increase.

The number of gigabytes (GB) that individuals worked with varied between 2 GB and over 1300 GB. The time spent by individuals ranged from 13 hours to almost 250 hours. In evaluating work completed during Phase 2 there was a general correlation between the number of GB in a collection and the time spent, however this is not always the case, as the charts below document.

Figure 3 shows that for accessions between 1GB and 50GB, there was a general correlation between the size of the accession and the amount of time it took to process the accession, and that the time generally increased as the size of the accession increased.

[12] Appendix G provides a summary of calculations used in this report.


**Figure 3: Time to Process Accessions Between 1-50 GB Displaying the General Upward Trend**

However, if we look at the largest and smallest accessions, processing time varied considerably, and may or may not coincide with the level of processing assigned. For example, a large collection that didn't take a lot of time may have been given a minimal level of processing, while a small accession that took more time may have received an intermediate or full level of processing. Figure 4 below demonstrates the four largest collections ERTF worked with this year; while Figures 5 and 6 represent the smallest collections, between 58 MB and 832 MB, and under 29 MB respectively.

In Figure 4, the collections of 180 GB and 230 GB, have ingest and processing times that increase as the size of the collection increases, but for the collections of 460 GB and 500 GB, the times to ingest and process decrease significantly. Figures 5 and 6 show that the time to ingest and process varies considerably and is not dependent on the overall size of the collection especially for the smallest collections.

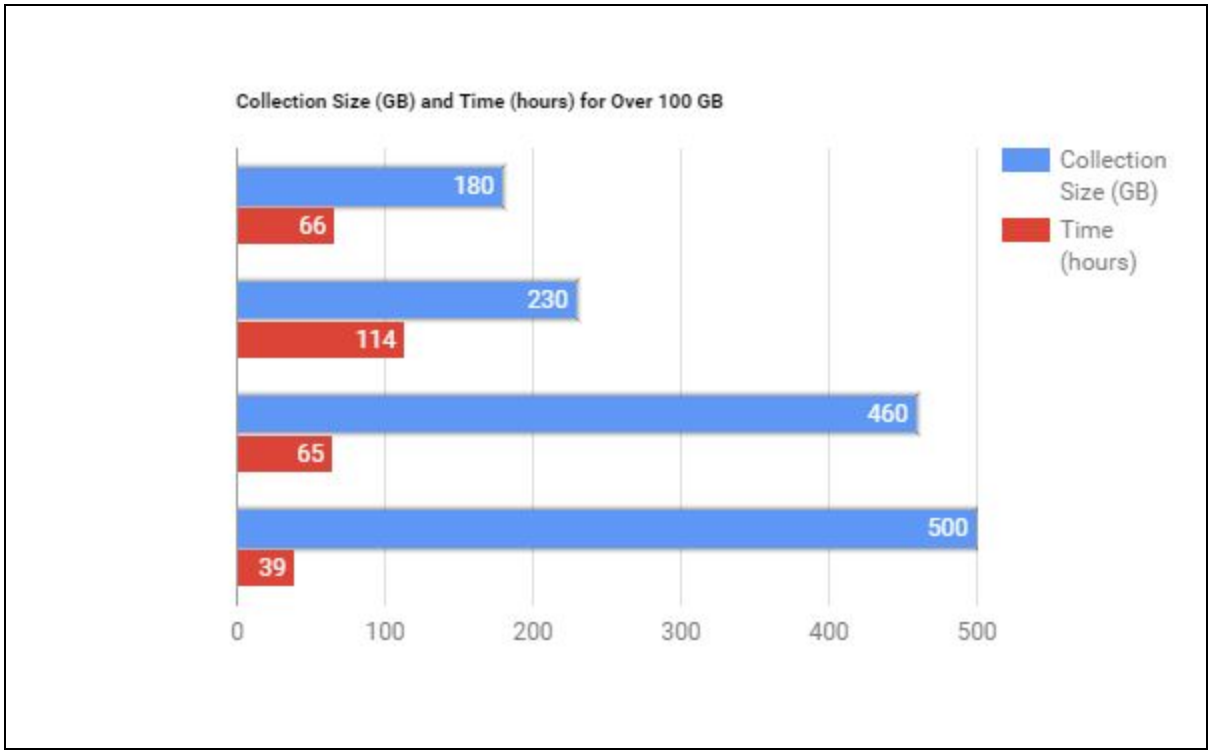



**Figure 4: Time to Process Accessions Over 100GB (Largest) Displaying Time Inconsistency**

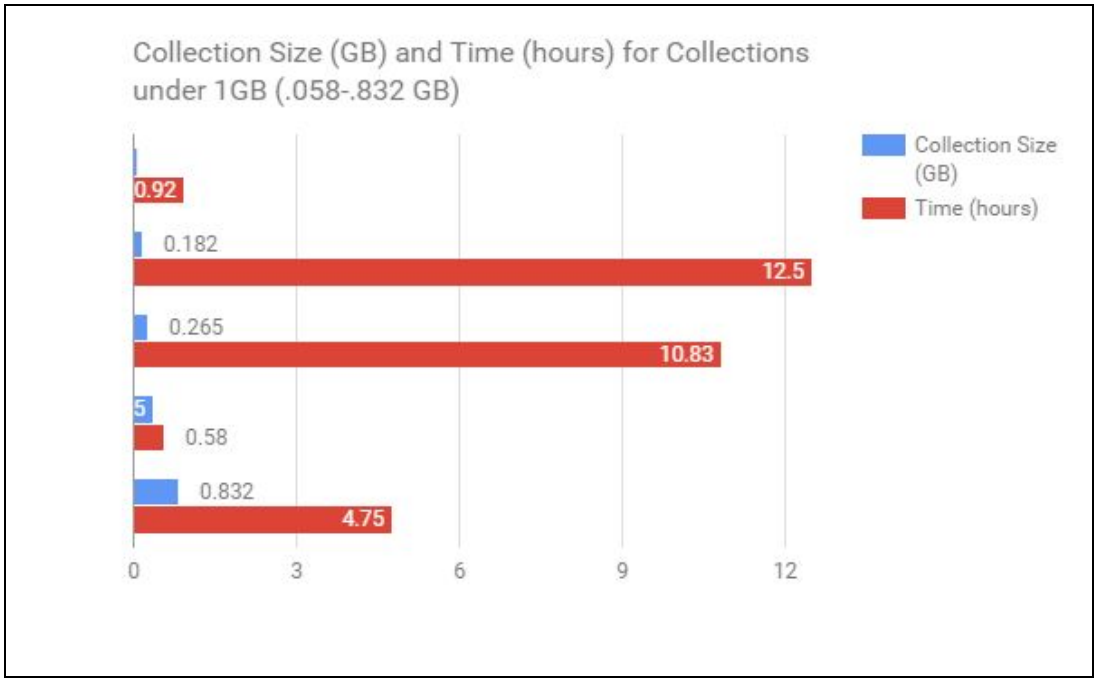



**Figure 5: Time to Process Collections Under 1GB (58 MB-832 MB) Displaying Time Inconsistency**


**Figure 6: Time to Process Collections Under 1GB (under 29 MB) Displaying Time Inconsistency**

These charts and our experiences show that while we might be able to make gross generalizations about how much time it takes to work with electronic records, the actual numbers are wide ranging. In comparison, the curators for the Data Repository for the University of Minnesota (DRUM)[13] have also experienced this wide range of time to curate a data set; with an average calculated to be about 3 hours per submission, the actual recorded times range from 30 minutes to 20 hours, as the time required is very dependent on the nature of the dataset.[14] Due to the lack of consistency in processing times, calculations used to address future work are based on averages across all of ERTF's work this past year.

## Projection about Future Ingest and Processing Time[15]

Looking forward, there are 35 accessions in the backlog waiting to be ingested and processed. Only 12 of these collections are of known size. The types of media represented in the group of unknown size range from 1Tb hard drives to hundreds of floppy disks. This report uses

[13] Data Repository for the University of Minnesota (DRUM) home page: https://conservancy.umn.edu/handle/11299/166578
[14] This information was taken from an email conversation with a DRUM curator 12-20-2016.
[15] Appendix G provides a summary of calculations used in this report.

averages from Table 2 to estimate the time to ingest and process all accessions listed on the current backlog, which are shown in Table 3 with calculations provided as footnotes.

| | # of collections | Amount of GB | Hours to Process | [estimated time it should take][16] | Time Spent (*ACTUAL*) |
|---|---|---|---|---|---|
| **Ingest and Processing (*Actual*)** | 27 | 1438 GB | 421 | | 23 hours/month or about 5.75 hours per week [17] |
| **Averages (*Actual*)** | | ~ 3.5 GB per hour[18] | ~16 hours/collection [19] | | |

**Table 2: Processing Time for ERTF Work on 27 Collections**

[16] Hours to process / 40 hours per week = number of weeks. Assuming 4 weeks in a month.
[17] 421 hours/18 months = 23 hours per month. 23 hours / 4 weeks = 5.75 hours per week.
[18] 1438 GB / 421 hours = 3.415 gb/hour or about 3.5 gb/hour
[19] 421 hours / 27 collections = 15.59 hours/collection or about 16 hours/collection

| | # of collections | Amount of GB | Hours to Process | [estimated time it should take][20] | Time Spent (*EXTRAPOLATED*) |
|---|---|---|---|---|---|
| **Future Accessions (*Known*)** | 12 | 300 GB | ~192 hours (based on # of hours per collection)[21] | 4.8 weeks or just over 1 month | ~ 8 months to process (based on time)[23] |
| | | | ~ 86 hours (based on # of GB/hour)[22] | 2.15 weeks or just over ½ month | ~ 4 months to process (based on time)[24] |
| **Future Accessions (*Unknown*)** | 23 | 897 GB[25] | ~368 hours (based on # of hours per collection)[26] | 9.2 weeks or just over 2 months | ~ 16 months to process (based on time)[28] |
| | | | ~ 256 hours (based on # of GB/hour)[27] | 6.4 weeks or just over 1 and a half months | ~ 11 months to process (based on time)[29] |
| **Total Future Accessions Estimated** | 35 | 1197[30] | ~560 hours (based on # of hours per collection)[31] | 14 weeks or 3 and a half months | ~ 24 months to process (based on time)[33] |
| | | | ~342 hours (based on # of GB/hour)[32] | 8.55 weeks or just over 2 months | ~ 15 months to process (based on time)[34] |
| | | | | | Average of 19.5 months to ingest and process the remaining accessions. |

**Table 3: Estimated Processing Time for Future ERTF Work**

[20] Hours to process / 40 hours per week = number of weeks. Assuming 4 weeks in a month.
[21] 16 hours per collection X 12 collections = 192 hours
[22] 300 GB/3.5 GB per hour = 85.7 hours (which was rounded to 86 hours)
[23] 192 hours/23 hours per month = 8.34 rounded to 8 months
[24] 86 hours/23 hours per month = 3.74 rounded to 4 months
[25] 897 GB was estimated by taking the average of the known averages (53GB/collection for the ingested and processed collections and 25Gb/accession for the waiting to be ingested and processed). The result is an average of 39 GB/accession. [(53+25)/2 = **39**]. If there are 23 accessions waiting to be processed at 39Gb each this is a total of 897 GB. (23 accessions* 39Gb/accessession)
[26] 16 hours per collection X 23 collections = 368 hours
[27] 897 GB/3.5 GB per hour = 256.3 hours (which was rounded to 256 hours)
[28] 368 hours/23 hours per month = 16 months
[29] 256 hours/23 hours per month = 11.13 months
[30] 300 GB +897 GB = 1197 GB total to ingest and process
[31] 16 hours per collection X 35 collections = 560 hours
[32] 1197 GB/3.5 GB per hour = 342 hours
[33] 560 hours/23 hours per month = 24.34months
[34] 342 hours/23 hours per month = 14.87 months rounded to 15 months

To summarize Table 2 and Table 3, over the past 18 months, ERTF staff spent about 23 hours a month, or 5.75 hours a week, to ingest and/or process electronic records. This averages out to 3.5 GB per hour or 16 hours per collection. Looking forward, there are 12 collections with a known total size of 300GB in the backlog. These 12 collections would take 86-192 hours to ingest and process based on calculations using GB/hour. Using an average[35] for the sizes of the unknown collections, we calculate the additional 23 collections in the backlog would take an additional 256-368 hours to ingest and process. Adding the known and estimated accessions in the backlog together, the time range is between 342 and 560 hours, or two to three months.

Ingesting and processing the 35 backlogged collections would take 2-3 months if someone was able to work on this task 40 hours/week. If, however, future ingest and processing is done at the rate ERTF members were able to devote to the task over the last year (23 hours/month), it would take 15-24 months just to clear the backlog. And it is unlikely that the average of 23 hours/month can be maintained going forward, given that the staff person who was able to commit the most amount of time to processing will need to turn their attention to other department level priority activities.

| Collections | Time Using Averages<br>(16 hours/collection and 3.5Gb/hour) | Time Using Actual<br>(23 hours/month) |
|---|---|---|
| 12 known | ½ - 1 month | 4-8 months |
| 23 unknown | 1 ½ - 2 months | 11-16 months |
| Total | 2-3 months | **15-24 months** |

**Table 4: Summary of Projected Time to Complete Future Processing**

## 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 to preserve and provide access to electronic records held by the Archives and Special Collections department. The work completed by the ERTF to ingest and process electronic records was a useful exercise, but continuing to manage the collections using temporary staff resources is not sustainable and falls short of the Libraries' mission as a good steward of archival material. As of December 31, 2016, the ERTF ingested 39 collections and the Accession Log listed 35 collections still awaiting ingest and processing with more being

[35] 897 GB was estimated by taking the average of the known averages (53GB/collection for the ingested and processed collections and 25Gb/accession for the waiting to be ingested and processed). The result is an average of 39 GB/accession. [(53+25)/2 = **39**]. If there are 23 accessions waiting to be processed at 39Gb each this is a total of 897 GB. (23 accessions* 39Gb/accessession)

added regularly. To put this in perspective, the accession log only reflects recent born digital acquisitions and does not address legacy electronic material (disks, disk drives, etc.) that continue to reside in boxes.

As previously stated, in order to better understand the human resource requirements for ingesting and processing electronic records the ERTF diligently tracked time spent on ingest and processing activities. Due to the nature of electronic records it is difficult to employ an all-encompassing standard to determine how long it takes to ingest and process a collection. The number of items in a collection, the total file size of a collection, the media on which the collection resides, its organization, and the types of files contained in the collection all affect the ingest and processing time. For example, a collection that is well organized and needs little attention to rearrangement could have many nested folders containing file names that are too long to transfer. This results in a considerable amount of time spent on troubleshooting activities to reduce the file names - in effect, increasing the time spent on a seemingly straightforward and uncomplicated collection. Total size of the collection may not be a good indication of time needed either, as a collection that includes multiple DVDs may contain fewer files/GB than an external hard drive but the time spent ingesting the content from the individual disks may take considerably more time. It is because of these complications that no standard for estimating ingest and processing time for electronic records exists.

Time tracking indicates that a total of 502 hours were spent on ingesting and processing activities over 18 months. The subset of ERTF members responsible for ingesting and processing the collections were restricted to conducting these activities when their schedules allowed. Regardless of the variable schedules, having larger blocks of time (2+ hours) available to ingest and process materials was more effective than smaller blocks of time. However, finding these larger blocks of time proved difficult, given that most Task Force members had many other time commitments to work around. It is clear that having focused and regular time to address electronic records ingest and processing tasks also eliminates time spent becoming reacquainted with the software and workflow. Staff that were able to work with electronic records more frequently maintained a higher level of familiarity and facility with the workflow. Having a consistent understanding of the variables inherent to managing electronic records is necessary to work efficiently.

Each of the 15 ASC units has electronic records in its collections, all requiring attention if they are to be preserved and made available. ERTF members worked primarily on records within their own unit, leaving non-unit specific members to address materials from unrepresented units. Continuing to rely on ERTF staff for this work will eventually cause forward progress for all units to stall, as is evident by the collections remaining on the accession backlog.

To address the discrepancy in processing activities between ASC units it has been suggested that the Central Processing[36] unit incorporate electronic records more fully into their processing activities. However, the 1.5 FTE Central Processing staff that was on ERTF currently does not have capacity to take on additional electronic records work themselves, as was demonstrated by the number of collections ingested or processed by Central Processing this past year (four of the 27 collections). In comparison, 3.5 FTE ERTF staff members who were associated with a collecting unit, focused on their own collections, worked on 13 of the 27, while a single Data and Technology staff member addressed 14 collections. In some cases multiple ERTF members worked in a collection to address troubleshooting and various activities. The percentages of this overall breakdown by staff affiliation is shown in Figure 7.

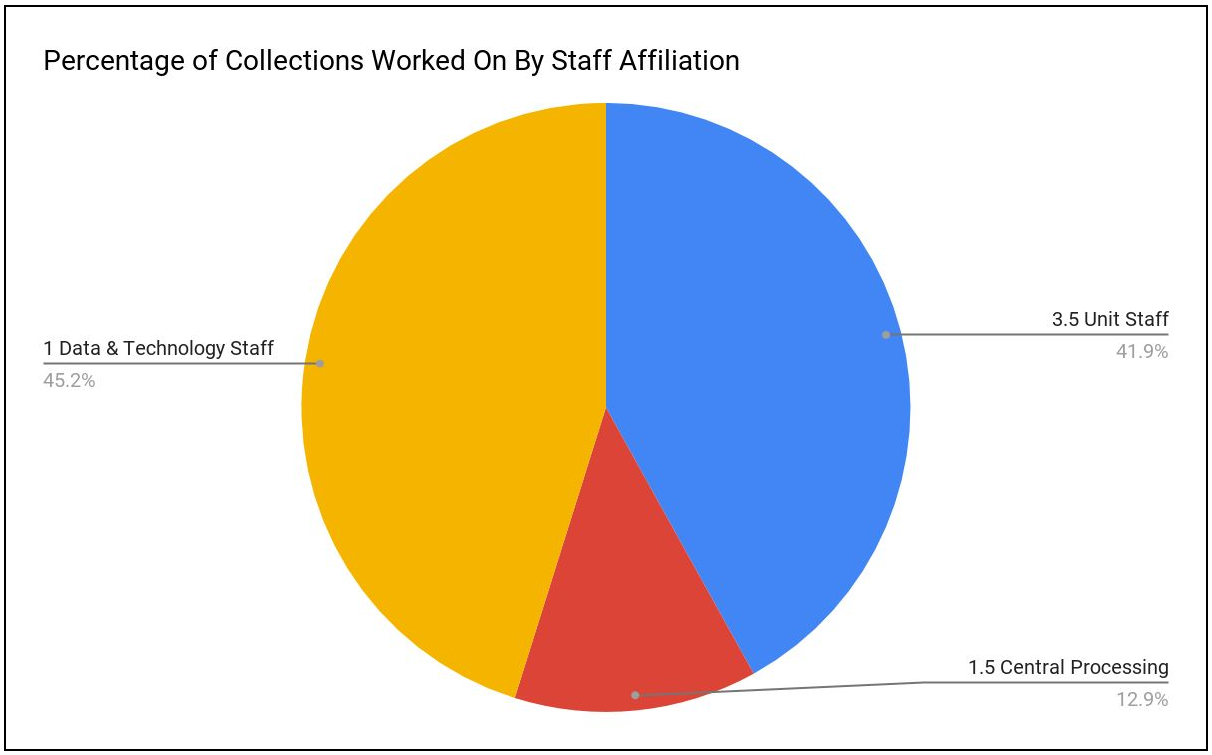



**Figure 7: Chart showing the Percentages of Work Based on Number of Collections By Staff Affiliation**

Based on unit and staff priorities, it is easier for those associated with a unit to find time to work on their own collections rather than those from another unit. Figure 8 below shows the *amount of time* spent on electronic records based on ERTF staff affiliation. Central Processing staff spent the least amount of time on collections, followed by Unit Staff who focused on their own collections, but both were superseded by one Data & Technology staff member who had the time to work on any and all collections as the need arose.

[36] The Central Processing unit works with all Archives and Special Collections units to inventory, process, describe, and create finding aids for archival collections.

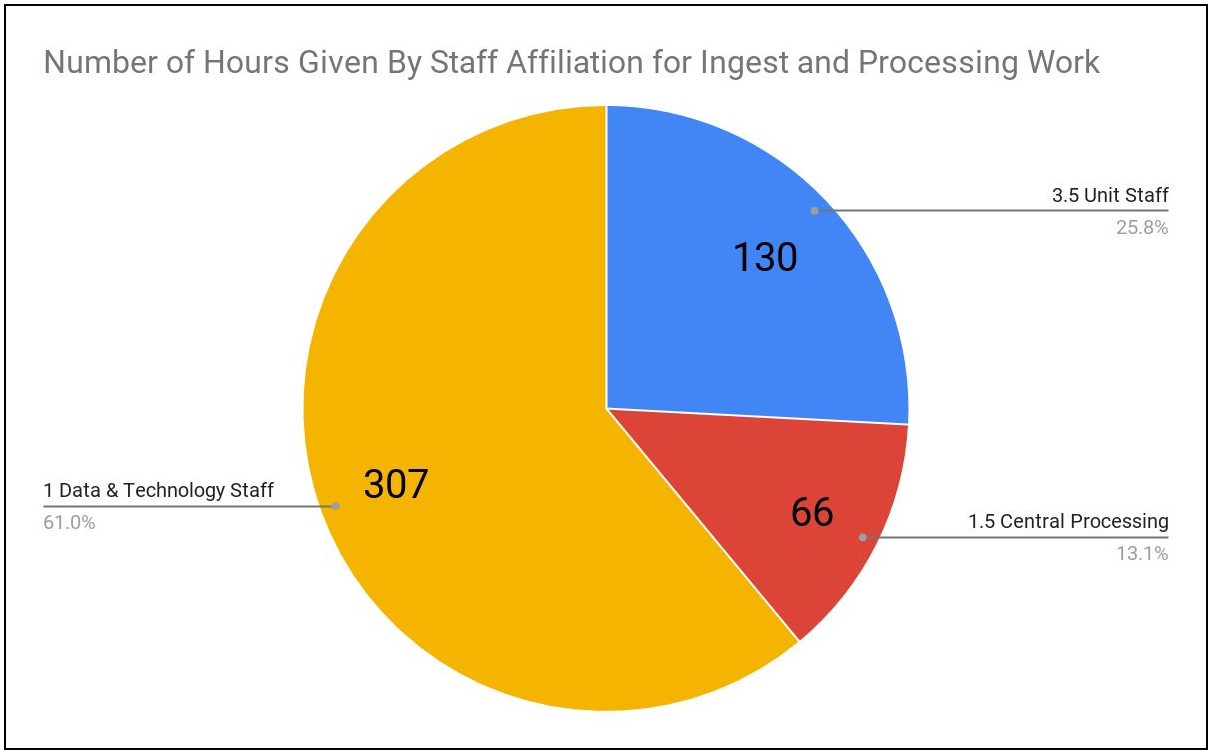



**Figure 8: Chart Showing the Hours and Percentages of Time Spent Working with Electronic Records by Staff Affiliation**

As Figures 7 and 8 show, the Data & Technology staff person spent the most time working on collections and was responsible for almost half of the work. It is expected that this person's available time and flexibility for processing with ASC electronic records will decrease significantly with responsibilities related to the implementation of a new system this next year. In addition, unit staff without electronic records processing experience cannot be expected to address born digital materials in their own collections - the work simply will not get done.

To immediately 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 work with ASC electronic materials. Using our calculations of current ingest and processing rates it would take a half-time person 28 weeks[37] to address the backlog, keeping in mind this does not include any unforeseeable issues that arise with each individual collection or time expected to be spent on consultation and collaboration activities with unit staff. Calculating how much time would be required for a half-time staff person to process the current backlog based on estimated GB/hour results is about 13 weeks.[38]

[37] Calculations: 35 collections x 16 hours /20 hours/week = 28 weeks
[38] Calculation: If there are a total of roughly 1197 GB in the 35 collections remaining to ingest and process, and we use the 3.5 GB/hour calculation these additional 1197 GB of information would take roughly 342 hours . 342 /20 hours a week = 13.1 weeks of a half-person's time.

A half-time person dedicated to electronic records will, over time, become more efficient, and be in a position to test new software and workflows. 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. As a result of the outreach efforts of the ERTF, ASC staff now feel more comfortable discussing electronic records with donors. For example, a discussion with the Minnesota Orchestra on best practices for organizing and managing their electronic records for future transfer to the Performing Arts Archives resulted in an immediate deposit. Proactively working with donors in this manner will enable them to feel 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 acquired, and must be able to work with incoming material in a sustainable manner.

In the long-term, a full-time dedicated staff person is the most responsible approach to working effectively and efficiently, to achieve quality work, and to maintain our leadership role in the field of electronic records management. This is arguably the only way to address the ingest and processing activities that assist with long-term access to and preservation of electronic materials. 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, any headway in addressing the records being collected will be made slowly.

Given the fact that individuals and organizations today do most of their work in electronic formats, the nature of archival collecting is facing a fundamental shift. This shift will require that the profession accepts the management of electronic records as a regular activity for all archivists. Securing dedicated staff now who can manage all ASC electronic record related activities will position the University Libraries at the forefront of organizations committed to providing exceptional access to unique and historic material. Suggested qualities of a successful candidate are provided in Appendix H.

## Providing Access (to Archives and Special Collections Materials)

Providing access to electronic records and promoting these unique assets is the main goal of ingesting and processing electronic records. As with analog collections, unit staff and ERTF members are cognisant of the varying factors that contribute to how and when a collection can be made accessible to users. Factors that influence the accessibility include:

- Absence or existence of files containing information defined by the University of Minnesota as Private-Highly Restricted or Private-Restricted
- Restrictions imposed by the donor or unit staff
- File types
- File size
- If the content was produced or sponsored by UMN

To gain an understanding of the types of access that are currently used and those that are needed, the Task Force developed the Access Criteria Matrix.[39] This Matrix defines three categories of access:

- *Unmediated*: Content that can be shared openly anytime and anywhere.
- *Mediated*: Content for which access must be moderated or restricted for some reason, which may limit use of material. This could include licensed content, content with a time- or event-based donor imposed restriction, a repository restriction, or a legal restriction.
- *Closed*: Content for which all access is denied (until a certain time- or event-bound date) after which content would become Unmediated or Mediated.

The Access Matrix also defines potential functional requirements and access methods for each category. For example, one requirement for access might be that materials with unmediated access can be made openly available online with or without the option/ability to download files. In our case, this could be accomplished by uploading appropriate materials to the University Digital Conservancy (UDC), UMedia repositories, or Google Drive with direct links to the material from the finding aid.

On the opposite end of the spectrum, a requirement for material that falls under the closed category would be to ensure that access would also be provided only to authorized (system/ERTF/unit) staff for management purposes. It would be useful to have a way to track expiration dates on closed status materials (similar to how the embargo feature works in the University Digital Conservancy), to monitor expiration dates and ensure appropriate access is provided. One option we may have for monitoring the long-term life cycle of collection restrictions is to the timebound tracking feature for restrictions available in ArchivesSpace.

Collection materials may fall into the mediated category for a variety of reasons, and the flexibility of access functionality is based on these variables. An option for providing mediated access includes restricting access to computers inside the campus IP network, or requiring authentication as a University of Minnesota user, either of which might be useful for a variety of materials that fall into the mediated category. Another option is to have users 'register' and be verified before gaining access. This is the online equivalent to the registration process for using materials in the reading room. Other useful functionalities desired include the ability to limit access to a specific period of time (e.g., one year), to provide limited ability to download or copy files, and to allow users to mark files from which they would like to request copies.

Various methods of access for many of the collections falling into the mediated category are currently in use by Libraries staff. When appropriate, staff have emailed files to users upon request. Staff have shared a link to a folder in Google Drive or Box with limited sharing settings and an expiration date for users. Content has been copied to a flash drive or CD/DVD. The content has also been made available on a locked down / restricted computer in the reading

[39] Informed by earlier work done by the ad hoc Electronic Records Policy and Procedures group in 2014, a new matrix was created. If interested, please contact ERTF for more information on the Access Matrix at lib-ertf@umn.edu.

room. To provide more instructions on specific access methods, ERTF created guides for UMedia (for e-records that can be made openly available) and Google Drive (for both open and restricted access). Additional instructions will be created for other methods on an as-needed basis for in-house use.

The Access Matrix also documents associated possible risks for and questions about each access method listed on the Solutions tab. Most of the questions are related to policies and procedures that have yet to be developed. Additional work will need to be done before implementing any of these solutions.

All access methods currently used are very hands-on and require significant staff interaction with individual patrons. While staff are accustomed to interacting with patrons to ensure their ability to access collection material, staff are eager to move away from such individualized practice to a solution that addresses patron expectations as well as automating access.

To this point, the ERTF sent out a survey to the ASC staff asking if they had electronic records in their collections and if they had provided access to any over the past year. 13 respondents from 12 individual repositories answered. All of them indicated that they had electronic materials in their collection but only four actively provided access to electronic material. Of the four who provided access, three of them simply pointed researchers to materials that were openly available online. The other request was fulfilled by the Digital Library Services department and not the archives. These results reinforce the idea that resources made available online receive higher use. It is much easier to point users to an online resource (or allow them to find it themselves) than needing to spend additional time working with an individual researcher on site who may want to view electronic records.

The desired ultimate solution is an Online Reading Room -- a single location that could provide access to electronic materials that fall into both the mediated and unmediated categories. This type of virtual environment would require users to register before being granted access to a specific collection. The system would track usage and allow unit staff to upload content and provide access as needed. ERTF began researching the feasibility and specifications for a virtual reading room and drafted a set of Virtual Reading Room Requirements. After drafting the requirements, ERTF decided that moving forward with the development of a virtual reading room was not feasible at this time. Members recognise that the importance of providing access to collection material through one succinct method is more sustainable and proves a firm commitment to stewardship than the current ad-hoc arrangement; however more research and development, which fall out of the scope of ERTF, would need to be undertaken to effectively implement such a system.

In addition to investigating the development of a virtual reading room, ERTF also considered how archival description impacts issues related to accessibility. In Fall 2016, ASC implemented ArchivesSpace as the management and discovery platform for ASC collections. Specifically, the ERTF updated the Finding Aid Descriptive Guidelines (Appendix I) to assist with creating

finding aids that include electronic records. Without accurately describing electronic records in finding aids (the main way researchers explore our holdings) there would be no way for people to know the records existed without personal intervention from unit staff. As we continue to process and describe electronic collections it is expected that more requests and use of electronic records will occur. With increased use, it will become increasingly important to find a way to manage access in a more consistent manner.

**Reading Room Access**

Access to electronic records should be provided on site at a computer in the ASC Reading Room, if other methods are not appropriate. In February 2016 it was determined that staff could provide access by transferring files to the desktop on the public access workstation in the reading room. However, around the same time, the status of the reading room computer was being evaluated by campus IT, Libraries IT, and ASC staff due to the higher security risk of public access workstations. At the end of the evaluation period, it was decided to upgrade the security level of the Reading Room computer to better meet University standards and categorize the computer as a 'kiosk'.

By design, kiosk computers are physically secure and have limited guest access to the internet. Staff were supposed to be able to add content to the desktop for researcher use, however it was noticed in October 2016 that the functionality to do so was not activated correctly and is not currently possible. That this was not noticed until October indicates that access was probably not provided to anyone in the Reading Room for most of the year.

Access to electronic materials via the kiosk computer can be done, however Libraries IT staff will need to do some work to make this possible. Due to the lack of immediate need and other access methods being used, this has not been a high priority. There is also some conversation about a second computer being added to the Reading Room which may provide additional options for providing access. Overall, permissions and controls to any and all computers in the Reading Room should be reviewed regularly.

## Preservation

The Libraries signed a contract for a digital preservation system in early 2016, however the system has not been implemented as of this writing. It is hoped that the developing preservation system will be able to perform many of the preservation activities required for electronic record collections held by ASC. Moving forward ERTF members and ASC staff should be aware of repository development and how the system might affect current and developing ingest, processing, and access procedures. Until the preservation repository is in full production, the Libraries need to ensure that the materials being held by the ASC units are properly managed and protected as assets of the Libraries.

Currently, the materials in ASC cannot be preserved using a single method. After being processed, most materials are copied to the network drive where backups are being made and irregular fixity checks are performed. However, files that contain Private-Highly Restricted (PHR) information are not allowed on network servers by the University. That means that files with PHR information remain only on the electronic records workstation hard drive and are not backed up or available anywhere else.

In order to follow through on our commitment of good stewardship of these unique materials, there must be a structured, sustainable way to preserve the materials that are ingested and processed for ASC. To preserve the context of the material as a whole, it would be best if all materials could be managed the same way instead of parsing out materials and treating them differently based on restriction status and subsequently tracking separate locations.

Suggestions have included using:

- *Box, a UMN enterprise secure cloud storage system*. All ASC materials, including PHR materials, could be uploaded into Box. However, Box has a file size limit of 15GB per file, which many ASC materials exceed. This would require that large files be stored elsewhere.
- *External Hard Drives*. Backups could be made on a schedule for all of the content on the workstation hard drive. The hard drives themselves would then need to be secured because they would contain PHR information.
- *Tape Drives*. Performing backups with local tape drives. The tapes themselves would need to be secured because they would contain PHR information.

None of these solutions are ideal or take preservation to the required level by actively protecting our content, which includes monitoring materials for unexpected changes over time. Material that is considered appropriate for long-term preservation should also be monitored for changes over time to make sure that the content has not been corrupted, and that we are able to provide access to authentic and reliable materials.

It is too early at this point in the adoption of the preservation system to make any specific statements about how its implementation could affect processing of archival material. Because of this, an interim preservation solution must be put in place; minimally, a solution for backing up all materials should be implemented immediately.

## Security

While the basic security concerns (and solutions) that were addressed during the first year of the Electronic Records Task Force still hold true, the focus this past year was on identifying files

that contained Private-Highly Restricted (PHR) files as defined by the University[40] and thinking about how to best preserve and provide access to those types of files. In establishing protocols, ERTF followed University best practices, consulted and referred to the University's documentation[41] that defines PHR and Private-Restricted, and other resources to determine appropriate storage locations based on privacy issues, including flowcharts to determine privacy categories.

In order to appropriately manage files that contain PHR information, such as social security numbers, health information, and some financial information, the first action is to identify the files and separate them into a "Restricted" folder within the collection, which allows staff to easily identify the material. This folder is used to separate materials with any type of restriction (legal, donor or repository imposed) from the rest of the collection. This folder is never transferred to the network with the rest of the collection; it remains only on the workstation hard drive.

Although the ERTF uses appropriate tools to scan all incoming data for PHR information, no known solution is 100% accurate. For example, a scanned image of a passport (unless it has undergone optical character recognition) is not going to be flagged as containing a passport number by a program that "reads" the text of a file. Currently, this would have to be identified by hand. This is one of the known complications and risks of working with electronic materials.

One potential method to address this possible security risk is to inform researchers and require them to notify staff if they come across sensitive materials, similar to the instructions in the Rights and Permissions section on the ASC Registration Form currently used for analog collections. More specifically, various Special Use Case Agreements have been drafted that release the University from being held responsible for data misuse. These forms place the responsibility on the researcher if they misuse private information they find in a collection. Sample text from these agreements can be found in Appendix J.

ASC staff are aware of the issues related to sensitive materials in general and take this into consideration when first accepting a collection, during processing, and when making it available. The internal ASC Data Privacy Policies and Procedures document assists ASC staff in making informed decisions on materials that contain all types of sensitive information.

Many discussions over this past year centered around following security protocols when providing access to electronic records. To this end, Box Secure Storage was explored as a possible access method for restricted information. Understanding the issues around

[40] Data Security Classification Policy. University of Minnesota. Last accessed: January 18, 2017. http://policy.umn.edu/it/dataclassification

[41] Many resources are produce by Liberal Arts Technologies & Innovation Services (LATIS) a few of which are listed here: Finding the Security Level for Research Data and Classifying Research Data. The University's Information Technology department also addresses these issues as found at the bottom of this resource about Box that provides guidance on available tools based on security level.

Private-Highly Restricted and Private-Restricted data, and how to address access needs, remains a main focus as we move forward.

The security of the workstation itself is controlled by limiting access to the room that the workstation is in, as well as physically securing it to the desk. Additional security measures should be evaluated based on a more thorough needs and risk assessment.

## Providing Access (to Libraries Electronic Materials)

In addition to the work completed for providing access to the Archives and Special Collections materials, the Access Subgroup looked at the issues surrounding providing access to previously purchased electronic Libraries' materials that were inaccessible to users.

In order to better understand the types of electronic records being collected and managed by the Libraries as a whole, the Access Subgroup met with most of the Libraries' Research and Learning departments to discern types of materials. Through these conversations it was discovered that the Social Science and Professional Programs department documented purchased digital materials on a spreadsheet during the 2015-2016 fiscal year, with a value of nearly $41,000.

Exploring ways to make these resources available, an Electronic Resources Librarian used Google Drive to pilot a low barrier solution to provide access to similar electronic materials. Initial testing of the pilot shows that Google Drive does provide low-barrier storage and access, however the sustainability of the process is questionable. A significant amount of staff time is required to manage permissions, upload files, and manage the organization of the drive. More information about the pilot was provided to the sponsors for review.

Through the work of the subgroup, numerous differences were found between archival and library materials, including the way in which they are acquired and managed. The processes of acquiring and providing access to Libraries electronic assets may be similar and parallel to ASC materials, but the long-term management, staff time and resources, and end user needs warrant a separate management group as defined under the recommendation for Libraries Electronic Materials Access.

## Other Activities

The following section documents other activities not specifically articulated in the Task Force's charter. These activities assist in sustaining operations and are ongoing.

### Upkeep on Workstation Computers

**Windows Workstation**

The Windows machine continues to be the main workstation. We try to ingest all of our materials directly onto its extra internal hard drive, a non-networked location referred to as the "data drive." The workstation computer was originally equipped with an extra internal 4TB hard drive to be used during the ingest and processing of electronic records. In June 2016, the 4TB drive was replaced with an 8TB drive to ensure a sufficient amount of space for incoming collections. To date this drive contains 3.06 TB of data. This drive is also currently the only location for complete copies of Archival Information Packets - as Private-Highly Restricted information is not allowed on the University network.

**Macintosh Workstation**

We continue to use a separate Macintosh workstation to ingest materials that come to us on Mac-formatted media. Of note, this hand-me-down machine, which we received last year, unexpectedly stopped working. Due to our workflow procedures we did not lose any data as no electronic records are stored directly on the Macintosh computer. (If this had happened on the Windows Workstation we would have lost a significant amount of information and work.) Initially IT was unable to immediately diagnose the problem and replaced the workstation with a newer hand-me-down computer of the same model. All user profiles were transferred to the 'new' computer, and programs were securely erased from the 'old' one. In the end, we can't rely on IT having 'leftover' computers for us to use for processing electronic records.

## Software Programs and Documentation

Software programs that may assist in the ingest or processing of files are added to a "Master List of Tools" as they are identified. Identifying new tools is an ongoing process as new needs or issues arise and new tools become available. Over the past year we tested and documented additional tools and updated some existing procedures.

New

- 5.25 Floppy Disk (use guide)
- 7-zip Info guide to address file names that are too long to move or delete
- Bulk Rename Utility User Guide
- Calibre: Converting HTML files to PDF
- Eraser
- Export Gmail to MBOX
- Fixity User Guide for Documenting Changes
- Handbrake: Converting Audio/Video TS files to MP4
- Managing Restricted Files
- Renaming Files in Bulk (using Bulk Rename Utility)
- ShowSize Info Guide

Updated

- Duplicate File Finder (update to pro; additional functionality)
- Data Accessioner (metadata transformer update)
- Electronic Records Transferred via Email and Cloud Services
- Fixty User Guide for Fixity Checking

- HashMyFiles (deleting files)
- XML Viewer

**Exploration of Box**

In addition to the tools listed above, ERTF members explored the use of Box as an option for both preservation and access to materials. Box is a secure cloud environment that allows storage and sharing of information with people inside and outside the University.[42] ERTF members began testing the functionality of Box in late November 2016, and followed up with a conversation with Box experts on campus in December 2016.

Box initially seemed like a good solution to provide a backup copy of the information on the workstation computer as well as a method for providing controlled access to a collection. However questions remain about known system-based limits and the lack of management flexibility. Considerations around the use of Box include:

- Managing permissions would need to be done by a central person to better monitor time limitations and other concerns.
- Management of backup copies would need to be done manually by a central person, as the 'automatic' method would increase the size of our holdings stored on the computer by at least twofold. The time it takes to upload collections could also be prohibitive.
- We have no central person at this point who would have the time to upload and manage the backups.
- Not all of the files could be uploaded to Box due to the 15GB per file size limitation.

Additional details on the testing done by ERTF and the remaining questions were provided to the project sponsors.

## Addendum to Deed of Gift

The Electronic Records Task Force drafted a Deed of Gift for electronic records in 2015. The Deed of Gift Addendum addressing the inclusion of electronic records in a donation was approved by the University's Office of General Counsel in February 2016. The Addendum covers three major issues: digital file management; privacy; and disposition of physical transfer media. Any collection that includes electronic records acquired after February 2016 must have this form completed prior to being ingested or processed. A copy of the text is found in Appendix K.

## Communications and Outreach

This year the ERTF's communication and outreach focused on determining the best way to process collections and was primarily with specific or interested parties. This enabled more individualized consultations with ASC staff based on need. However some public outreach was also undertaken.

[42] Box Secure Storage Service Page, UMN. https://it.umn.edu/technology/box-secure-storage

**Contact with Archives and Special Collections Staff**
Processing electronic records collections requires input from both ERTF members and unit staff who are familiar with the collections being ingested and processed. These individual consultations ranged from single 15-minute conversations to multiple hour-long meetings set up on an as-needed basis. In addition, all ASC staff were asked to participate in a survey about providing access to electronic records between November 2015 - November 2016, and staff continue to record incoming accessions on the ASC Records Interim Accession Log informing ERTF staff of future work. Ongoing discussions and open communication between ASC staff and electronic records processing personnel is a requirement for the success of processing electronic records.

**Public Presentations**
At least 19 formal presentations at a variety of conferences and blog posts were given or created by members of the Electronic Records Task Force over the past two years. Public presentations provided an opportunity for knowledge transfer and feedback between Task Force members and the University community, as well as with other individuals and organizations outside the University community. These presentations and reports have become resources for others, as most recently demonstrated by the citing of last year's Electronic Records Task Force Final Report in a D-Lib article.[43] Publishing reports, sharing documentation and procedures, and presenting at conferences are excellent ways to share our experiences with the community, and we intend to continue to do so to keep leading the way. The most recent presentations includes a workshop, presentation, and poster at three different conferences.

# Conclusion

The ERTF was successful in meeting the goals set out at the beginning of Phase 2 by taking steps to advance the electronic records program within Archives and Special Collections units as well as the general collections of the University Libraries. The main focus of the Task Force was the creation of documentation regarding the management of, and tasks appropriate for, processing electronic records. This included testing tools that could assist members in bulk operations, the creation of workflows and processing plans, and user guides for processing electronic records. Task Force members also addressed processing priorities and levels with the conclusion that the level of processing and/or description does not correlate to the level of effort to accomplish the work.

The goal of processing unique electronic archival material is to make it available to end users, whether they be skilled researchers or a high school student working on a project. Task Force

[43] DeRidder, Jody L. and Alissa Matheny Helms. Intake of Digital Content: Survey Results from the Field. D-Lib Volume 22, Number 11/12. November/December 2016.
http://www.dlib.org/dlib/november16/deridder/11deridder.html

members addressed current access points available for electronic records and researched potential opportunities to improve access points. Although current research use of electronic records is low, due to the newness of our ability to provide access to these materials, unit staff foresee a need for and understand the benefits of a virtual reading room to collocate the Libraries' rich and diverse electronic assets. However, given the outstanding variables regarding potential access support through a preservation system, it was decided to not recommend moving forward with a virtual reading room at this time. Although not sustainable in the long-term, the ERTF concluded that continuing to use the available access methods discussed in this report will provide researchers adequate access to available collections for now.

In addition, the Task Force monitored and fine-tuned ingest workflows and activities in accordance with their understanding of how processing workflows integrate with the ingest process. After implementing the processing workflows members were able to better define expectations for the completion of a Submission Information Package and an Archival Information Package in order to meet professional standards.

While significant developments were made to document protocols for the ingest and processing activities of electronic records within ASC, the long-term management of these assets have not yet received similar attention. Activities around both the preservation of and access to these materials must be developed side-by-side with supporting policies.

The long-term management of ASC electronic records requires a different set of policies that need to be approved and enforced at the department level. For example, all other electronic records that are acquired by the Libraries' repositories are doing so under strict policy or review. The University Digital Conservancy[44] accepts materials only if they have a direct tie to the University; and currently every file that is ingested into DRUM is individually curated. UMedia Archive[45] is used for digitized materials selected by staff. Minnesota Reflections[46] is for digitized materials approved by a committee guided by the Minnesota Digital Library's collection development policy. All of these repositories use item-level metadata, but ASC does not generate item-level metadata at the minimum or intermediate level of processing, and only sometimes in the full level of processing. There needs to be a policy-level decision that allows for this difference and provides flexibility to manage and preserve unique archival materials at an aggregate or collection level.

Based on the recommendations made in the report, the co-chairs are confident that the Electronic Records Task Force has met the stated goals and do not recommend continuing the Task Force with a third iteration. The Task Force strongly advises that the only sustainable way to move forward with stewardship of electronic records is to hire a dedicated staff person. The Task Force also recommends that in order to support future development of policies based on

[44] University of Minnesota Digital Conservancy: https://conservancy.umn.edu/
[45] UMedia Archive, University of Minnesota: http://umedia.lib.umn.edu/
[46] Minnesota Reflections, Minnesota Digital Library: http://reflections.mndigital.org/

anticipated changes to available tools, access, and preservation, an ASC Electronic Records Management Group be instituted to support development, research, and continuing staff education regarding the management of electronic acquisitions and collections.

# Recommendations

The following describes outstanding issues and recommendations that must be addressed to develop the electronic records program for the Archives and Special Collections and Libraries' electronic materials. Advancing these recommendations would leverage existing infrastructure and staff capacities, but would also require additional staffing and financial support. Preliminary resource requirements are offered beneath each recommendation.

**Staffing**

Consistency is important in being able to work effectively and efficiently. Without a dedicated person who has an in-depth understanding of evolving workflows and protocols, who can keep current with emerging standards and new software, and who can provide a consistent approach with curatorial staff, headway in addressing the records being collected will be made slowly. Securing dedicated staff who can manage all ASC electronic record-related activities will align the University Libraries with other organizations committed to providing exceptional access to unique and historic material.

1. **Recommended Action/s:** Hire a permanent full time employee to work specifically with electronic records within ASC. This person would lead and manage all activities related to the ingest, processing, access, and preservation of ASC electronic records. If resources are not available for the full recommendations it is essential that we move forward immediately with a phased implementation focused on ingesting and processing activities. Preliminary resource requirements:
    - New staff: 1 FTE Assistant Librarian
    - Existing staff: From Technology Support and Consultation, Digital Preservation and Repositories Technology, and Archives and Special Collections will be expected to collaborate with appropriate staff to address some of the recommendations made in this report.

**Long-term Development and Management**

Over the past year, ERFT members have worked with ASC and Libraries staff to better understand environments and suggest working solutions for processing and providing access to materials. The workflow that was adjusted to accommodate the processing of collections should continue to be used and modified as necessary.

2. **Recommended Action/s:** Sunset the ERTF and create an Electronic Records Management Group to oversee ongoing activities implemented by past ERTF members and address outstanding concerns specifically about access, preservation, and security.

**Preservation**
The long-term preservation of electronic records materials within ASC's stewardship remains a concern and work-in-progress. Currently, due to various technology, processing, and security requirements for different materials, there is, unfortunately, *no single method* available for securely and consistently backing up, managing, and preserving entire collections in the same location. Solutions supporting specific aspects of ingest, archiving, and access processes have been explored but are not as yet fully understood, let alone implemented. Establishing an appropriate digital asset management and preservation environment depends on many factors including available technologies, level of security and access controls required, and cogent policy around these matters. Given divergent requirements, a singular asset management, backup, and preservation solution may not be a feasible goal in either the near or long term. However, efforts can be made to establish a limited number of processes to manage the vast majority of preservation use cases.

3. **Recommended Action/s:** Request that the Digital Preservation and Repository Technologies department review all currently known workflow and long-term management requirements of electronic records as determined by the Task Force and recommend 1) immediate near-term solution(s) for file backup and recovery, and 2) a longer-term strategy for backup, recovery, and preservation that leverages the Libraries' Digital Preservation Policy Framework and preservation management system. All solutions and strategies must assume the need for University compliance when handling private data.

**Security of the Electronic Records Workstation**
The workstation used to ingest and process electronic records may, at any time and likely without intention, hold some private-restricted and possibly private-highly restricted data. This workstation resides in a workspace available to staff, outside of a secure data center, and as such requires a high level of security as its default setup and configuration.

4. **Recommended Action/s:** Request that the Libraries' Technology Support and Consultation department, working with University Information Security where necessary, conduct a thorough review of security requirements for the workstation including, but not limited to physical security, device encryption, network firewall security, and authentication and authorization for the workstation and its contents. On the basis of this review, upgrade the workstation using configurations that ensure compliance with University Information Security Standards.

**Equipment**
As part of the natural life cycle of technology, the computer equipment used for ingesting and processing electronic records will eventually need to be replaced or upgraded. We were able to replace the Macintosh computer this year with another that had been decommissioned, but we

can't depend on the availability of second-hand hardware. Additional equipment may also be necessary to preserve and provide access to materials.

5. **Recommended Action/s:** Ongoing financial support for hardware and software for the workstation, as well additional equipment to help preserve and provide access to the electronic materials in the collections. Pending review of functional and security requirements, specific recommendations may include additional local secure storage capacities for onsite backup, a second workstation to increase productivity, and additional hardware and software as needs change. The workstation will also need to be maintained and managed as part of the Libraries' Computer Replacement Cycle. Preliminary resource requirements:
    - Initial purchase and life-cycle replacement funding for the ingest and processing workstation, data backup technologies, and any other computing equipment to support the long-term management and access to these electronic records.

**Access**

Access to collections is currently in a state of flux. As more collections are described in finding aids and made available, we will gain a better understanding of user needs and expectations on how users can access electronic materials. Until access is requested more frequently, it will be difficult to determine the best access solution(s).

6. **Recommended Action/s:** Continue to describe electronic records in finding aids and solicit recommendations for access methods. Work to determine if there is a single access method solution for ASC materials as requested by ASC staff, and explore options for providing access to electronic records that require mediation via a Virtual Reading Room.

**Libraries Electronic Materials Access**

The Libraries is purchasing electronic materials for patron use, however the availability of these resources is not always publicly known as they were often purchased for a specific reason. All materials purchased by the Libraries should be cataloged and made available to all.

7. **Recommended Action/s**: A group or entity other than the proposed ASC Electronic Records Management Group should be charged to further investigate ongoing needs for the Libraries purchased electronic materials. The concern is how to provide access to electronic materials that were purchased by different library departments for which there is no access platform or a central way to manage them.

# Appendix A: Project Charter

## Electronic Records Task Force Phase 2

Proposal for continuing work in 2015-2016. Approved by Libraries Cabinet October 13, 2015.

### Background and Context

The Libraries are deeply immersed in the collection, development, and use of digital materials as witnessed by activities and services across the Libraries: University Digital Conservancy (UDC), Data Repository for the University of Minnesota (DRUM), Open Scholarship and Publishing Services, Digital Arts Science and Humanities (DASH), and the department of Archives and Special Collection (ASC) to name a few. To best determine how to work with incoming electronic records that are unique and not controlled by license agreements, the Libraries initiated the Electronic Records Task Force (ERTF) in 2014.

Over the past year, the ERTF built a physical environment to securely ingest electronic materials donated to Archives and Special Collections units. In addition, the Task Force developed workflows for ingesting those materials along with supporting policies and procedures. These steps, however, were only the first of many when addressing the long-term needs for access to and preservation of electronic materials.

The Electronic Records Task Force Report, submitted to Cabinet on June 29, 2015 and reviewed on July 7, 2015, included recommendations for moving forward to address the needs of born-digital materials ingested by the Department of Archives and Special Collections. Next steps include establishing workflows for processing ingested materials (appraising, arranging, and describing) and creating policies and procedures and mechanisms for making them available to researchers and other users. Specific aspects of the next phase of work include:

- **Processing**: Appraisal, arrangement, and description are all part of the act of "processing." Electronic records may also need to have technical actions performed on them prior to making them accessible to users. This may include removing duplicate files, renaming files, moving files to new locations, addressing actions for restricted materials, or converting into more accessible or preservation friendly file formats.
- **Appraisal**: As with paper records, electronic records need to be reviewed for overall value. We can't afford to save everything that comes to us in digital format. We must take the time to review the incoming materials to make sure they fit our collecting criteria and that we have the appropriate technologies to preserve and provide access to them.
- **Arrangement and Description**: Electronic records may or may not come to us in an orderly fashion. It is imperative that time be spent arranging and describing electronic records in a uniform fashion in order to make them accessible in a way that users can understand them. Without this, the value of collections may be lost.

The uniqueness of born digital archival records requires a focus on preservation that reflects the Libraries' commitment to the long-term retention and accessibility of valuable cultural material. This forward thinking mission requires long-term goal planning and needs assessment in order

to appropriately address how best to provide access to electronic records. Access methods, which need to be explored, will vary based on content and material type. For example, some records in the Libraries' collections require supervision during use (or other controls around access and use) due to the sensitive nature of the content, while other records may be made fully available via the web. Long-term preservation is currently being addressed with best practices. The planned procurement and implementation of a digital preservation repository and management system will significantly improve upon the current short-term methods of preservation and access.

## Scope and Purpose of the Task Force

Extending the work of the Electronic Records Task Force into its next phase will enable the Libraries to keep moving forward to address the unique needs of the electronic records in its holdings. Electronic records must be processed before we can make them accessible to our researchers; without this step we are not responsibly able to provide access to any of our electronic records.

If the Libraries are unable to address the processing phase of electronic records management, although we will remain able to ingest records we will create a volatile backlog whose management will be more difficult. Lack of personnel may also lead to our inability even to ingest electronic records. If we are unable to ingest or process records we will not be fulfilling our responsibility to our donors and researchers who expect high quality services from us.

This next iteration of the ERTF will monitor the workflows established for ingest and develop new workflows for processing both current and anticipated holdings of electronic records. Critical steps include:

- defining processing and descriptive levels and best practices
- defining staff responsibilities for processing actions
- developing processing workflows
- developing access policies and procedures
- identifying existing resources or acquiring new resources (technological and human) to fulfill responsibilities to make these electronic records accessible.

The Digital Preservation and Repository Technologies department is leading the Libraries' efforts to develop, implement, and operationalize a digital preservation program. This includes developing policies and procedures supporting the Libraries' Digital Preservation Framework that affect the Libraries as a whole, as well as implementing a digital preservation management system that will support the long-term preservation of archival electronic materials. Workflows, policies, and procedures developed by the ERTF designed for archival materials may also be applicable to the Data Repository of the University of Minnesota (DRUM), the University Digital Conservancy (UDC), and other areas under the technical purview of the Digital Preservation and Repository Technologies department (DPRT). The ERTF will work with and consult with staff focused on these projects to ensure efforts aren't duplicated and to share valuable information.

## Process Tasks and Deliverables

Tasks listed below may be addressed simultaneously as they relate to one another.

### *Task 1: Develop Workflows for Processing Ingested Collections*

1. Develop workflows and role assignments for processing work, including appraisal, arrangement, and description.
2. Define minimal requirements for an Archival Information Package (AIP)
3. Determine short and long-term human resource requirements for completion of curatorial and technical work.

Note: Explore and, if applicable, implement the functionality of ArchivesSpace to track decisions made and work completed.

### *Task 2: Define Processing Levels*

1. Define levels of processing (minimal (MPLP), intermediate, full)
2. Understand the relationship between file formats, processing levels, and preservation capabilities, and document requirements/concerns as needed within the broader context of the Libraries' digital preservation program.
3. Assign processing priorities based on processing level, preservation needs, and anticipated user needs.

### *Task 3: Develop Access Methods that Address End-user Needs, Copyright, Data Privacy and other Information Security Requirements*

1. Document the spectrum of access requirements that meet user needs.
2. Identify and document the spectrum of distribution and access control requirements as related to copyright, data privacy, and other relevant information security policies and/or laws.
3. Document and expand protocol for using *existing* methods of access, including onsite capabilities and currently available online repositories. (e.g., UMedia, UDC, HathiTrust, etc.)
4. Identify unmet end-user needs, and, where feasible, plan and implement new methods for access that address user needs.

### *Task 4: Monitor Ingest Workflows*

1. Adjust ingest workflows as necessary
2. Edit related policies and procedures as necessary

Note: Ingest workflows and procedures may change with the implementation of ArchivesSpace

### *Task 5: Monitor Preservation Repository Development (when available)*

1. Work iteratively with the Digital Preservation and Repository Technologies staff on repository implementation and related development.
2. Monitor and assist with digital preservation policy development as needed.

Note: In the long-term, it is hoped that the preservation repository will be able to perform many of the preservation activities required for collections with electronic records. ERTF members should be aware of repository development and how the system might affect current and developing ingest, processing, and access procedures.

## Sponsors

John Butler
Kris Kiesling

## Task Force Members

Lisa Calahan, Co-Chair
Carol Kussmann, Co-Chair
Kate Dietrick
Lara Friedman-Shedlov
Betsy Friesen
Erik Moore
Arvid Nelsen
*(Digital Repositories and Records Archivist position when filled)*

## Resource Personnel

Kevin Dyke
Lisa Johnston
Mary Miller
Jon Nichols
Mike Sutliff

## Duration

12 months

## Budget

For this phase, the majority of our costs are expected to be in staff time, however a pool of funds should be available to address any hardware or software needs that may arise, including a reading room access workstation. We are asking for $3000.00, to supplement the $5000.00 balance carried forward from the previous iteration of the Task Force.

## 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 B: Information Packet Definitions

Defines Submission Information Packets, Archival Information Packets, and Dissemination Information Packets which assist in understanding requirements for ingesting, processing, preserving, and providing access to ASC electronic records.

***SIP: Submission Information Packet***

*Submission Information Package (SIP): An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.*[47]

- Includes the files *as provided* by a donor and ingested onto the workstation computer.
- May include some additional information about the collection as provided by the donor or curator

***AIP: Archival Information Packet***

*Archival Information Package (AIP): An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within a system. The AIP often consists of the original files deposited, processed versions of data files and documentation, normalized files, and associated metadata.*[48]

- Ingested files after they have been accessioned/processed into a new or existing collection
- Information about the files from tools used during accessioning/processing
    - HashMyFiles report (documenting: checksums, file structure, file name, file size, file extension, date created, date modified)
    - PII results (documenting which files, if any, contain Personally Identifiable Information)
    - Data Accessioner XML file (documenting: checksum, file format identified, size, file structure) ***OR*** DROID report (documenting: checksum, file format identified, size, dates, file structure)

***DIP: Dissemination Information Packet***

*The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS."*[49]

- Full or partial copies of or versions of the AIP that are requested and provided to a user
- Information about the files may also provided if necessary or requested

[47] https://www.lib.umn.edu/dp/glossary#S
[48] https://www.lib.umn.edu/dp/glossary#A
[49] https://www.lib.umn.edu/dp/glossary#D

# Appendix C: Processing and Descriptive Levels for Archival Materials

Originally for use only with analog materials, the definitions in this document were updated to include electronic records.

## Processing and Descriptive Levels for Archives and Special Collections Archival Materials

All archival collections would be divided into the following two categories: new acquisitions and backlog. Once the collection is in physical possession of ASC a level of processing and description will be assigned, taking into consideration perceived research need, monetary value, and associated gift or grant funding. For more information about how to describe electronic archival materials based on the level of processing review the *Description Guidelines* document.

**New Acquisitions**
All new, incoming collections will be minimally processed (as described below). This step will ensure no additional collections are added to the unprocessed backlog. Using minimal processing at the time of acquisition does not preclude the collection from having further, more detailed, processing at a later date.

**Backlog Collections**
Collections currently in the unprocessed backlog will have one of the following 3 levels of processing assigned, depending on such criteria as research value, monetary value, outside funding, and current condition. It will be possible to apply more than one processing level to a single collection. For example, intermediate processing may be generally appropriate for a collection, but a single series of meeting minutes filed chronologically may be arranged and described minimally.

**Note about Duplicates/Removing Items from Collections**
When implementing minimal processing techniques, 'best practice' processes vary about how to identify duplicates and other items for removal from archival collections. These processes also currently vary between ASC units. To try and standardize processing so students are not required to learn specific processes for each unit, the following procedures should be followed:

Duplicates should always be removed if easily found. No more than 2 duplicate copies should be kept in a collection, unless the copies are rare or unique. An item which includes significant changes or notes is not considered duplicative.

The processing coordinator will alert home unit if items fall under FERPA, HIPAA, and other federal privacy rules to determine disposition based on the *ASC Data Privacy Policies and Procedures* and appropriate restrictions.

**Processing Level Definitions for Central Processing**

**Minimal**

- Analog collections (or analog materials in hybrid collections) – processed and described to the box level; series may be assigned depending on size and/or original order; no physical preservation activities carried out on the collection; no physical rearrangement of the materials. Description will meet requirements for DACS single-level (collection level) description. Processing plan created if further processing and description needed.
- Born Digital collections (or born digital material in hybrid collections) -- No file arrangement or renaming for description/discovery enhancement. A SIP will be created (includes transfer from original media, scan for viruses and PII, checksum validation, file format identification, identification of duplicate files) on which further actions will be taken to create the AIP. Description will meet requirements for DACS single-level (collection level) descriptions. File formats will not be normalized. Action will not generally be taken on duplicate and PII information identified as part of ingest. Processing plan created if further processing and description needed.

**Intermediate**

- Analog/hybrid collections – Processed and described to the series level with high research value series complemented with scope and content notes. Some collections may be processed and described to a combination of the folder level, series level, and/or box level inventory; routine preservation activities may be carried out, depending on need.
- Born Digital/hybrid collections -- Creation of SIP (includes transfer from original media, scan for viruses and PII, checksum validation, file format identification, identification of duplicate files) on which further actions will be taken to create the AIP. Top level folder arrangement and top-level folder renaming as needed. File formats will not be normalized. Some duplicates weeded and redaction of PII. Description will meet DACS multi-level elements: described to the series level with high research value series complemented with scope and content notes.

**Full**

- Analog/hybrid collections -- Processed and described to the folder level; series and subseries assigned, may or may not include physical arrangement. May include item level description where warranted. Routine preservation activities carried out as necessary; includes folder level inventory.
- Born Digital/hybrid collections -- Creation of SIP (includes transfer from original media, scan for viruses and PII, checksum validation, file format identification, identification of duplicate files on which further actions will be taken to create the AIP. Top level folder arrangement and top-level folder renaming as needed; renaming and arrangement at item level where appropriate. File normalization conducted as necessary or as appropriate. Duplicates identified and removed, redaction of PII as needed. Description will meet DACS multi-level elements: described to series, subseries, or item level where appropriate; with high research value components complemented with additional scope and content notes.

# Appendix D: Levels of Effort

*Created to help explain the level of effort needed to process electronic records. The more files that need direct attention, the higher level of effort is required. The level of effort is a better indicator than the level of processing for how much time and energy was or will be required to address the processing needs of a collection.*

The Electronic Records Task Force was asked to define levels of processing for electronic records in 2016. As they should be, these processing levels were defined based on the existing processing levels for analog materials. However, during the actual processing of some sets of electronic records we realized that the processing level did not always correlate to the level of effort that was being put towards a collection. We didn't want people to view statistics on the level of processing and assume that something assigned with a minimal level of processing was 'easy' or 'quick' to complete. To address this concern, we created a table to document a level of effort for each accession being processed.

**Level of Effort**

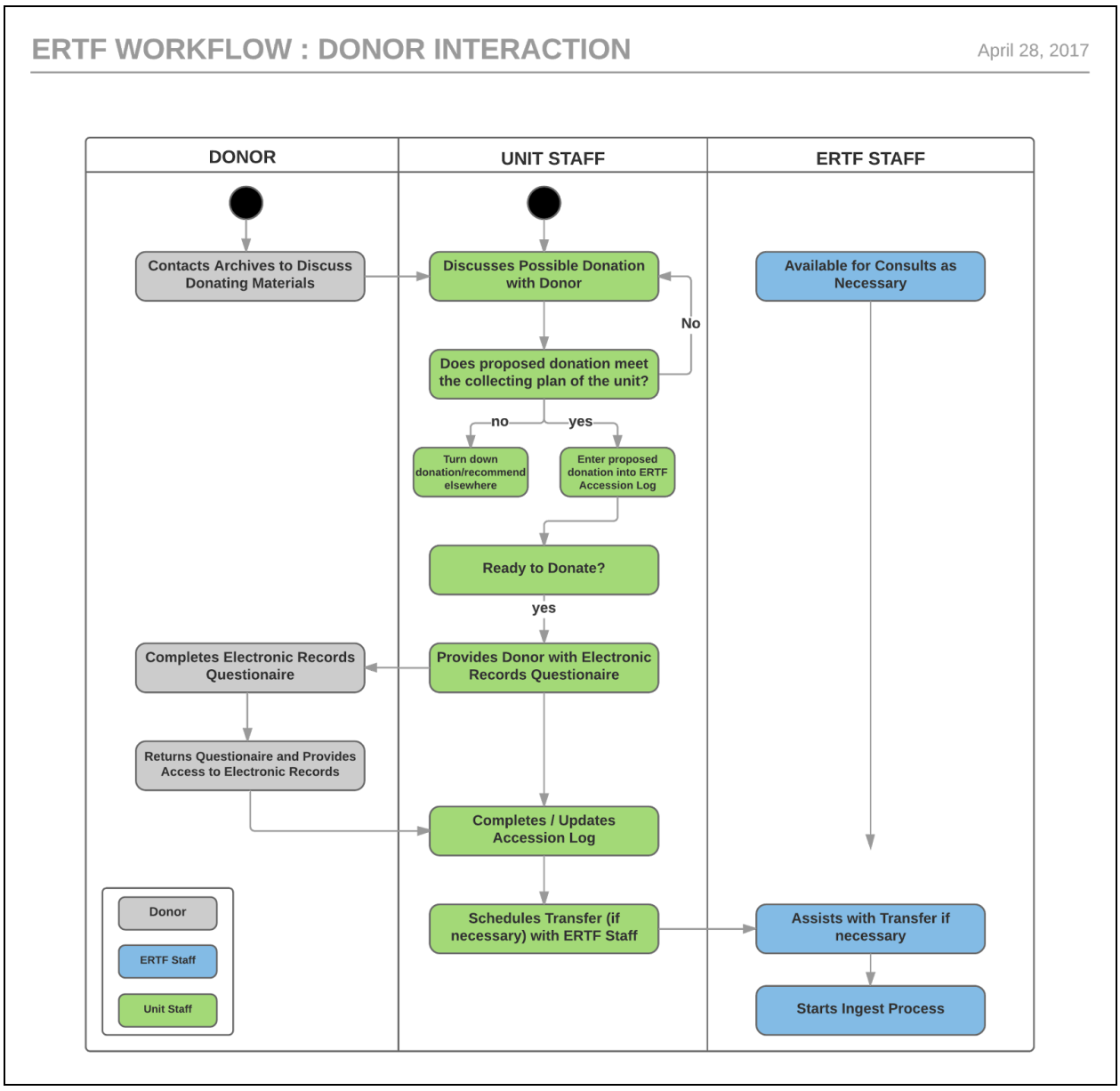
| | **Minimal** | **Intermediate** | **Full** |
|---|---|---|---|
| **Folder/File Renaming (bulk rename utility)** | 0-25 files | 26-100 files | 100+ files |
| **Folder/File Renaming (by hand)** | 0-10 files | 11-25 files | 26+ files |
| **Removing of / Reviewing Duplicates (tool)** | 0-25 files | 26-100 files | 100+ files |
| **Removing of / Reviewing Duplicates (by hand)** | 0-10 files | 11-25 files | 26+ files |
| **Addressing PII** | 0-10 files | 11-25 files | 26+ files |
| **Meetings (curator/donor)** | 0-1 meeting | 2-3 meetings | 4+ meetings |
| **Comparisons across accessions/collections** | NA | Comparisons across accessions/collections (small collection) | Comparisons across accessions/collections (large collection) |
| | | | |
| **Arrangement** *(? done before deciding on description levels - to make things manageable in the first place)* | Moving 0-10 Folders/Files | Moving 11-25 Folders/Files | Moving 26+ Folders/Files |
| | | | |

We don't want determining the level of effort itself to add a lot of work to the process. We could assign the level of effort based on:

- Gut feeling
- Overall time it took (or some percentage based on the size of the collection)
- Or documenting where each of these activities fell during processing such as can be seen below.

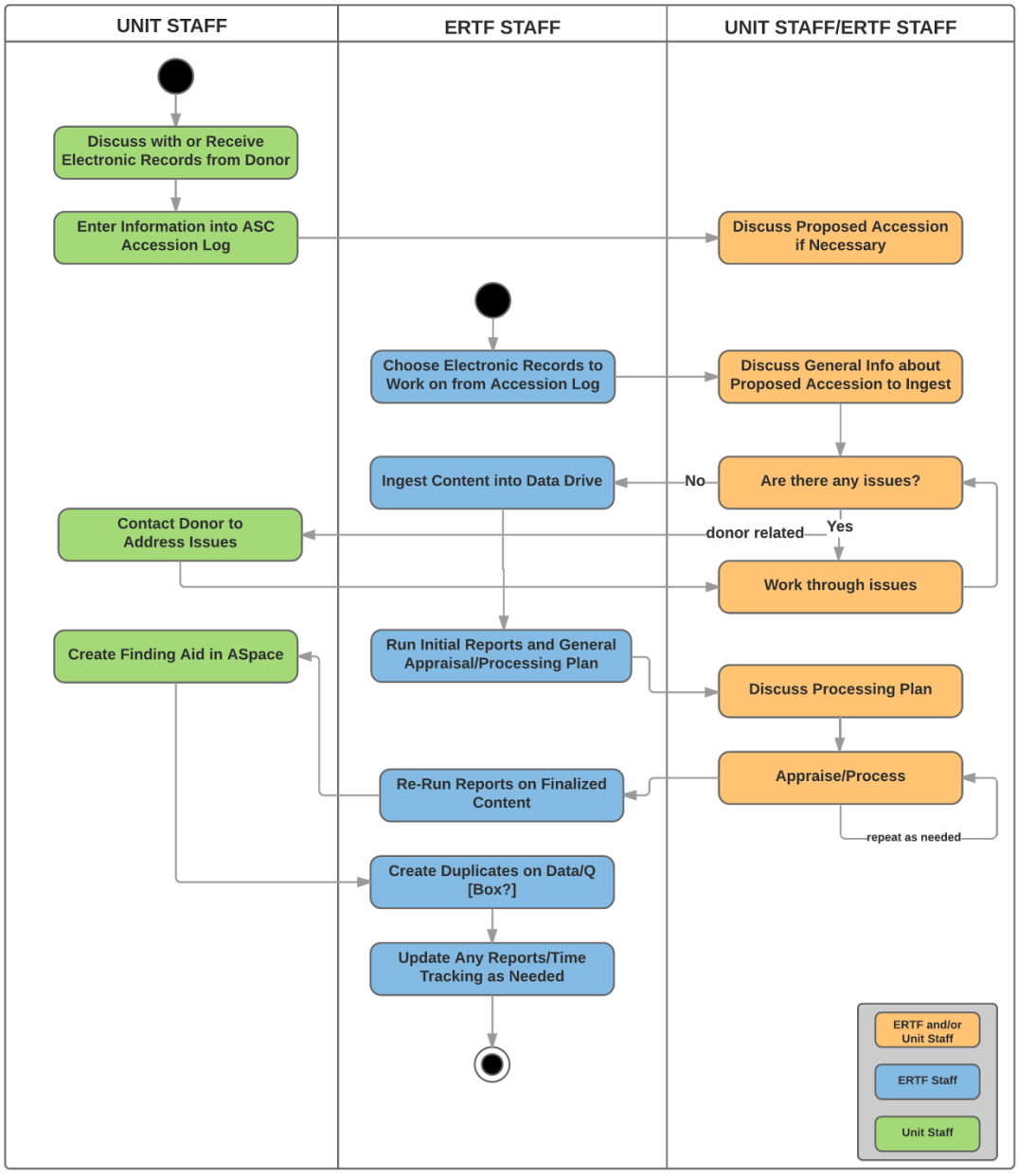
This information can be recorded on the existing Accession Log.

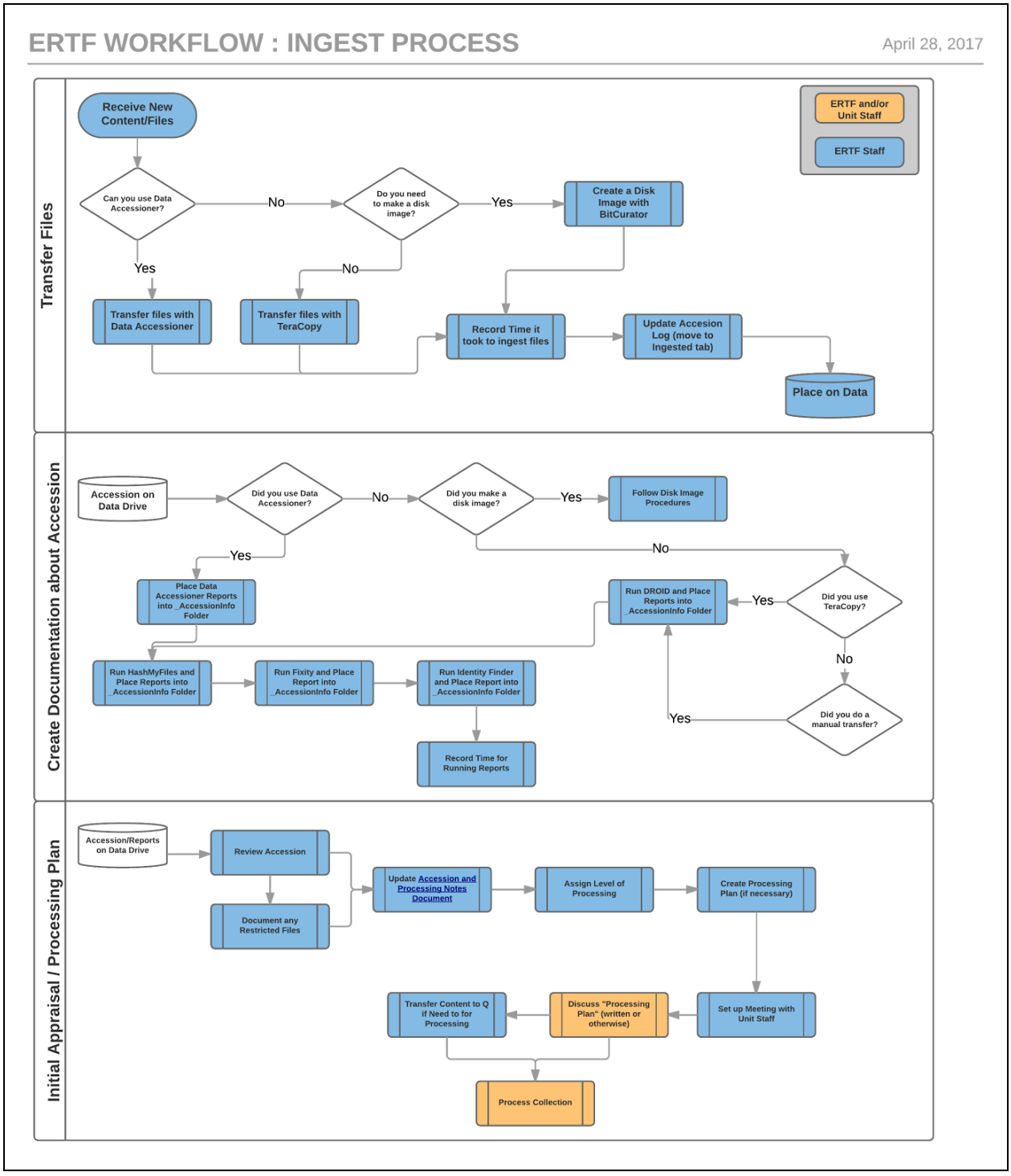
# Appendix E: ERTF Workflow Diagrams

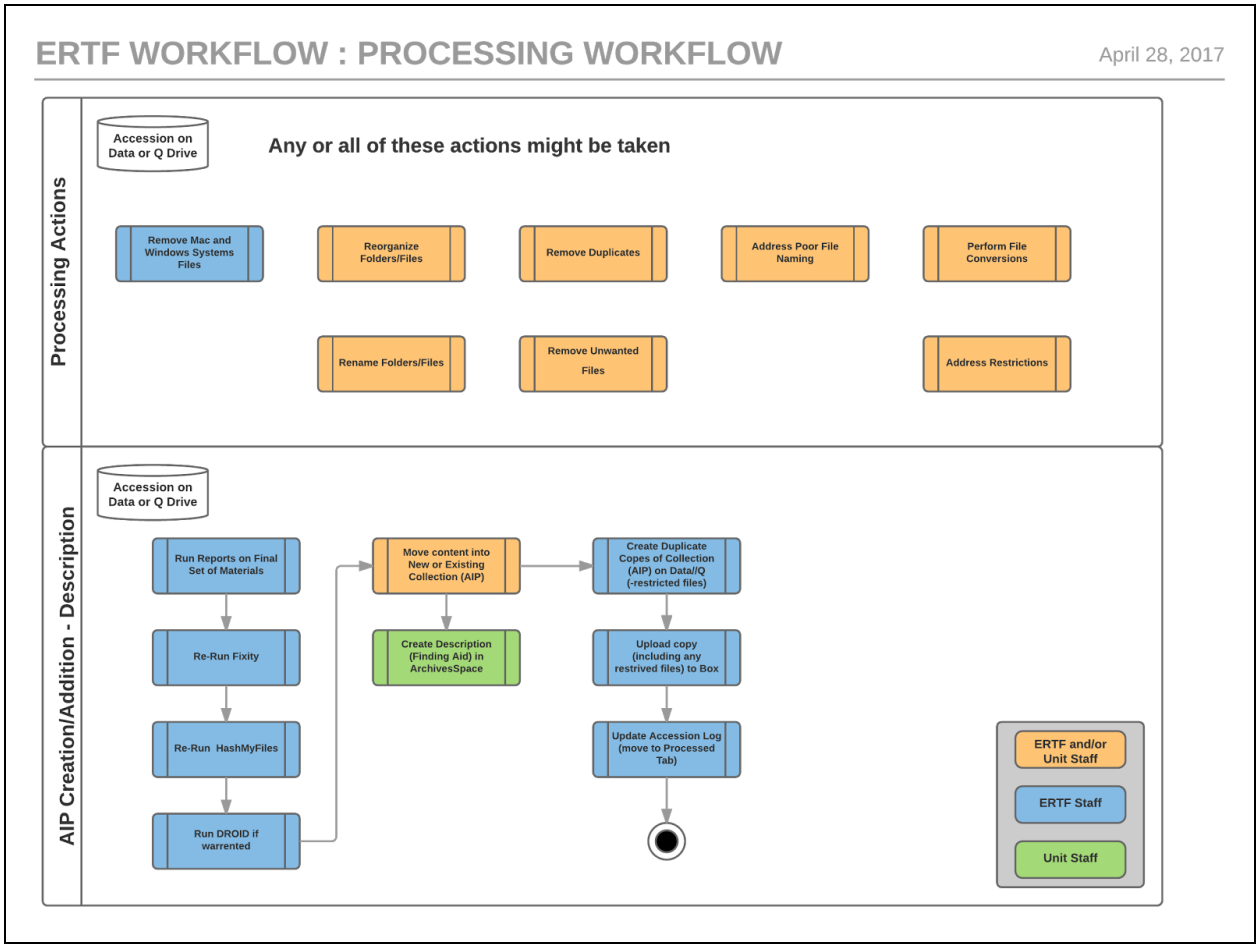


## ERTF WORKFLOW

April 28, 2017




ERTF WORKFLOW : INGEST PROCESS
April 28, 2017
Transfer Files
Receive New Content/Files
ERTF and/or Unit Staff
ERTF Staff
Can you use Data Accessioner?
No
Do you need to make a disk image?
Yes
Create a Disk Image with BitCurator
Yes
No
Transfer files with Data Accessioner
Transfer files with TeraCopy
Record Time it took to ingest files
Update Accesion Log (move to Ingested tab)
Place on Data
Create Documentation about Accession
Accession on Data Drive
Did you use Data Accessioner?
No
Did you make a disk image?
Yes
Follow Disk Image Procedures
Yes
No
Place Data Accessioner Reports into _AccessionInfo Folder
Run DROID and Place Reports into _AccessionInfo Folder
Yes
Did you use TeraCopy?
No
Run HashMyFiles and Place Reports into _AccessionInfo Folder
Run Fixity and Place Report into _AccessionInfo Folder
Run Identity Finder and Place Report into _AccessionInfo Folder
Yes
Did you do a manual transfer?
Record Time for Running Reports
Initial Appraisal / Processing Plan
Accession/Reports on Data Drive
Review Accession
Document any Restricted Files
Update Accession and Processing Notes Document
Assign Level of Processing
Create Processing Plan (if necessary)
Set up Meeting with Unit Staff
Discuss "Processing Plan" (written or otherwise)
Transfer Content to Q if Need to for Processing
Process Collection


ERTF WORKFLOW : PROCESSING WORKFLOW
April 28, 2017
Processing Actions
Accession on Data or Q Drive
Any or all of these actions might be taken
Remove Mac and Windows Systems Files
Reorganize Folders/Files
Remove Duplicates
Address Poor File Naming
Perform File Conversions
Rename Folders/Files
Remove Unwanted Files
Address Restrictions
AIP Creation/Addition - Description
Accession on Data or Q Drive
Run Reports on Final Set of Materials
Re-Run Fixity
Re-Run HashMyFiles
Run DROID if warrented
Move content into New or Existing Collection (AIP)
Create Description (Finding Aid) in ArchivesSpace
Create Duplicate Copes of Collection (AIP) on Data//Q (-restricted files)
Upload copy (including any restrived files) to Box
Update Accession Log (move to Processed Tab)
ERTF and/or Unit Staff
ERTF Staff
Unit Staff

# Appendix F: Processing Workflow Instructions

This document provides basic steps for the ingest and processing of electronic records. The internal version links out to more detailed instructions for many of the steps and software to be used.

## General Process for Ingest and Processing of Digital Content

*Note: This workflow reflects current practice as of June 2017. The original format of this document is a Google Doc that contains numerous links to other documents that may or may not continue to persist. Because of this we have made only the main body of the workflow available below. The text in bold after or within a step indicates where additional documentation existed as of June 2017.*

### A. Ingest Materials

1. Understand Accession
   Review information gathered by unit staff based on the questions listed in the **donor guide** and the **electronic records transfer sheet**. e.g.
   - Is this a new collection or part of an existing collection?
   - What are the expectations for processing and future access for the materials?
   - File transfer options (donor to curator/archives)
     - External hard drive or flash drives
     - Network transfers
     - Cloud services (Google drive, Dropbox, Email attachments)
     - Gmail (to Mbox file) (**Google Takeout Instructions**)

2. Review Accession Record Information
   2.1. Check for Accession Number in ArchivesSpace (as listed in the Accession Log)
      2.1.1. If no number in ArchivesSpace, contact unit staff. Ingest and processing cannot continue without an Accession Number in ArchivesSpace.
   2.2. Review Deed of Gift
      2.2.1. If acquisitioned after February 2016, a Deed of Gift Addendum is required before work can continue.
      2.2.2. If prior to February 2016, this is not a requirement, but may be requested.

3. Create File Structure and Paperwork Trail

3.1. Create folder in _ASC Electronic Acquisitions Folder for each accession in appropriate location (in a Units folder) in Google Drive [ex: for YMCA, Y20140321_Conard]
3.2. Create tab on the Unit's Time Tracking sheet for the Accession (if a Time Tracking Sheet does not exist follow **these instructions**).
3.3. Make a copy of the **_Accession and Processing Notes Template**, move it to the folder and start to fill it out as the collection is being transferred in the next step...
3.4. Set up new location on Data Drive using proper naming conventions (**instructions 2017**)

4. Begin to transfer content to Data Drive
4.1. Prepare to Transfer files to Data Drive
4.1.1. Connect media to workstation (**instructions**), noting differing instructions for PC vs. Macintosh formatted media.
4.1.2. Locate and/or Download content from email / cloud services
4.1.2.1. Dropbox/Google Drive/Email Attachments (**instructions**)
4.1.2.2. Gmail (to Mbox file) (**Google Takeout Instructions**)

4.2. 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. (**instructions**)
4.2.1. Use Data Accessioner (DA) if possible (**instructions**)
4.2.2. Use TeraCopy if DA is not possible (**instructions**)
4.2.3. Perform a logical transfer if unable to use Data Accessioner or TeraCopy.
4.2.4. Depending on the media, creating a disk image is another option (**instructions**)

*Note*: If there are multiple pieces of storage media in the accession, complete all transfers prior to creating any additional documentation. This allows for the creation of a single report from DROID that covers the entire accession folder. Also, if you use Data Accessioner and only change the Source Name/Identifier you will only get one XML file instead of one for each.

5. Update **Accession Log** (move to Ingested Tab and edit fields as necessary)

6. Check for ._Ghost files, .DS_Store files, and Thumbs.db files
6.1. If found, remove them (**instructions**)

7. Generate documentation (**information/instructions**) and collect in _AccessionInfo folder for the current accession.
Note: If you notice ._Ghost files, .DS_Store files, or Thumbs.db files immediately after transfer, remove these prior to generating additional documentation. (**instructions**)

7.1. Data Accessioner reports (created at the point of running the program; xml file)
7.2. If TeraCopy or logical transfer was used for file transfer, run DROID (**instructions**) and collect documentation.
7.3. Run HashMyFiles (**instructions**) and collect documentation on the entire accession.
7.4. Run Fixity on the accession for the purpose of documenting changes from the initial deposit to the final package that we are saving. (**instructions**)

8. Check for PII/SEI
   8.1. Using Identity Finder ( **instructions**)
      8.1.1. Save reports before and after clean up with Identity Finder
      8.1.2. If files with PII are deleted, document in the **_Accession and Processing Notes**
      8.1.3. If files with PII are being kept (whether permanently or temporarily), we will need to document this. Files with PII will need to be listed on the Restricted File Log_**Template** (google sheet). See Processing section below for further instructions.

9. Record time it took to transfer files/run reports. (**Instructions**)
   9.1. DROID and DataAccessioner have log files to use.
   9.2. Teracopy lists how long the transfer took when complete.
   9.3. Other actions require estimates. Watch your time.

10. Review the Deed of Gift Addendum to determine what to do with Original Media (Media that was taken from a box is NOT to be returned, per our **working policy**.)
   10.1. If the media is to be returned, give it back to the curator to get it back to the donor.
   10.2. If the media is not to be returned, we will need to securely dispose of it *after the collection is fully processed*.

## B. Information Sharing and Review

11. 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.] Please note in this documentation if this accession will be added to an existing collection or will be creating a new collection.
   11.1. Accession and Processing Notes document (located in the **_ASC Electronic Acquisition Folder**) [Use this to record information to document the steps processors took and problems they might have had. ]

11.2. Create a Processing Plan *if 1) the collection will not be processed immediately 2) someone other than yourself will be processing the collection.* (**Processing Plan Template** in this **folder** for easier copying)
11.3. Document Restricted Files if necessary on the **Restricted Files Log** Use this to record information about files with any restrictions, whether imposed by the donor or by the repository. (**Instructions**)

12. Set up Meetings
    12.1. Meet with the unit staff and/or central processing staff to discuss reports.
        12.1.1. Share reports with the unit staff and/or central processing staff to review prior to meeting. (Email the location of the reports on Google Drive.)

13. Discuss Processing Plan with Unit Staff
    13.1. Determine who will be doing the processing (central processing / unit staff)
    13.2. Determine if content will be transferred content to Q Drive
        13.2.1. Transfer to Q is warranted if processing can and will be happening on a computer other than the ERTF workstation
        13.2.2. Some accessions may be too large to transfer to Q
        13.2.3. Some accessions may be additions to existing collections, you will need to make sure the full collection is in the processing location and a backup copy exists elsewhere. [ie. processing work could be done on Q with a backup on Data, or processing could be done on Data with a full backup on Q]
        13.2.4. Some collections may require processing tools that are only available on the ERTF workstation
        13.2.5. Some collections may contain large amounts of restricted data that cannot be stored on Q and cannot easily be separated
    13.3. Notify unit staff if any sensitive information was found, separated, and is still on the Data drive.

## C. Processing

Based on the collection at hand and the assigned processing level, the following actions may or may not be performed. Other tools listed on the **Workstation Tools and More** spreadsheet may be of use.

14. Initial processing on accession
    14.1. Initial processing on an accession may include any or all of the following actions.

- Remove any ghost files, DS_store files, Thumbs.db etc
    - Delete by hand

        - Use tools
            - Remove Unnecessary System Created Files (**instructions**)

- Remove empty directories (**instructions**)
    - c:Program Files(x86)/RemoveEmptyDirectories/RED2.exe

- Address poor file naming
    - Find additional files with bad file naming (**instructions**)

- Remove unwanted files
    - Securely delete with Eraser (**instructions**)

- Remove duplicates
    - Duplicate File Finder (**instructions**)
    - HashMyFiles (**instructions**)

- Folder/File Organization
    - TeraCopy (**instructions**)

- Folder/File Renaming
    - Bulk Rename Utility (**Instructions**)

- File Conversions
    - Video TS files to MP4 (**Instructions using Handbrake**)
    - HTML files to PDF (**Instructions using Calibre**)

- Restrictions
    - If the collection is restricted or contains highly restricted files, follow the procedures for identifying and logging (**instructions**)
    - To better understand the content that may or may not be restricted on the archival level review the **ASC Data Privacy Procedures and Policies** document.
    - Log restricted files in the **_Restricted Files Log** for that collection (template), being sure to take note of the types of restrictions outlined on the 2nd tab.

15. If appropriate, integrate accession into existing collection
    15.1. Make a copy (or work from an existing copy) of the existing collection.
    15.2. Move accession files to the same folder as the copy of the Collection files

15.3. Begin integrating the files from the new accession into the existing folders/series of the Collection using any/all of the processing actions above. Create new series as appropriate.

15.4. After all of the files have been integrated into the Collection's structure, do some final processing

15.4.1. Search for and remove duplicates as necessary across entire collection

15.4.2. Address restricted files

15.4.2.1. If the collection is restricted or contains highly restricted files, follow the procedures for identifying and logging (**instructions**)

15.4.2.2. To better understand the content that may or may not be restricted on the archival level review the **ASC Data Privacy Procedures and Policies** document.

15.4.2.3. Log restricted files in the **_Restricted Files Log** for that collection (template), being sure to take note of the types of restrictions outlined on the 2nd tab.

## D. AIP Finalization, Storage and Description

16. Set up Collection file structure

Choose one of the options below for how to organize the files.

16.1. When you have a new collection

16.1.1. Rename the top level folder for the collection with the collection number and name (ie. PAA120_Hardenberg)

16.1.2. Rename the _AccessionInfo folder as _AccessionInfo_accessionnumber_dateofaccession

16.1.3. Create a new folder called _AccessionInfo at the same level as the _AccessionInfo_collectionnumber_dateofaccession

16.1.4. Move the _AccessionInfo_collectionnumber_dateofaccession folder into the new _AccessionInfo folder

16.1.5. Create a new folder called _CollectionInfo

16.2. When you are adding to an existing collection [All of your files should now be combined within the existing collections file structure.]

16.2.1. Rename the _AccessionInfo folder from the accession you added to the collection to _AccessionInfo_collectionnumber_dateofaccession

16.2.2. Move the _AccessionInfo_collectionnumber_dateofaccession to the main _AccessionInfo folder for the collection.

17. Rerun reports on final set of materials[50] and save to _CollectionInfo folder

[50] The final set of materials may be a single accession - if a new collection is being created. Or it might be a collection made up of multiple accessions - one for which a new accession was just added to.

17.1. Rerun Fixity (**instructions**) to capture the changes made during processing.
17.2. Rerun HashMyFiles to generate a list of file names, checksums, relative file path
17.3. Potentially run DROID or other file format identification tool if warranted (e.g. to document unusual file formats).

18. Copy files to and from Q/Data so that there are duplicate copies of all processed files and reports in both places, taking care to not move any highly restricted files to Q.
    18.1. If copying over a merged collection, you will want to make sure you are copying over the ENTIRE set of files (in some manner) rather than just the new files.
    18.2. If collection contained restricted files that were left on Data while processing took place on Q, be sure not to delete the restricted files from Data by replacing the entire folder with the processed versions from Q. (See "Restrictions" instructions in section C)

19. Save the Acquisition and Processing Notes document as a PDF and add it to the collection _AccessionInfo folder for the accession just completed on the Q/Data drive.

20. Describe Files in ArchivesSpace
    20.1. Create a resource record in ASpace. It could be a basic record spawned from accession record or an archival object added to an existing resource record
    20.2. See the guidelines for born digital material in **Description Guidelines**

21. Update **Accession Log** (move to Processed tab and edit fields as necessary)

22. Record your time it took for processing on the time log.

## E. Preservation Actions

[These are things that we might want to consider doing to assist with long-term preservation. We are not actively doing these things at this time. Specific documentation will need to be created if we decided to do any of these things.]

23. Set up Fixity checking on final set of materials. Possible programs include:
    23.1. Fixity (**fixity checking instructions**)
    23.2. Exact File (**public documentation**)

24. How do we follow up? What is our responsibility to follow up on?

# Appendix G: Documentation for Calculations within Report

A spreadsheet was used to provide documentation for calculations listed in this report. Each tab on the spreadsheet was used for a different purpose as documented below.

- **TimeTracking by Collection**: Time was tracked for various ingest and processing activities at the collection level. This tab separates the collections that were both ingested and processed from the collections that were ingested only. The total time ERTF spent includes work from both sections, while all other calculations are based only on the ingested and processed collections. All future collections must be both ingested and processed so it was decided to use the combined number only when calculating future work estimates.
- **Times for Individual Time Calculations 1-3-2017**: This tab pulls the collection name, and total number of minutes from the TimeTracking by Collection tab. In addition, it records the person responsible for the work, the assigned Processing Level, and the Level of Effort that was given to the collection. The number of GB per collection was also recorded.
- **Final Individual Time Calculations 1-3-2017**: This tab pulls the final numbers per person from the Times for Individual Time Calculations 1-3-2017 and calculates the number of hours per person and the percentage of a full time position that was able to be given towards ERTF work. The staff affiliation (curator-unit staff, central processing, and outside ASC) were also listed which was used to determine how much work was being done by affiliation/group.
- **ASC Waiting 12-20-2016**: This is a list of the Accessions that were listed on the Incoming Accession Log that were waiting to be ingested and processed at the end of 2016, documenting the backlog of files at this point in time for report writing purposes.
- **Final Number Ingested Only**: Pulled from the Incoming Accession Log, this provides additional information about the collection as provided by unit staff when originally added to the Accession Log. Used to document the 10 Ingested Only collections.
- **Final Number Accessions Processed**: Pulled from the Incoming Accession Log, this provides additional information about the collection as provided by unit staff when originally added to the Accession Log. Used to document the Ingested and Processed collections.

For additional information about the calculations within the report please contact the Task Force at lib-ertf@umn.edu.

# Appendix H: Electronic Records Specialist Position Description

The ERTF recommends the hiring of a full time Electronic Records Specialist to manage the electronic records within ASC collections. If there are concerns about how a full time staff member's time would be filled, based on informal conversations with professional colleagues, one full-time Digital Collections Assistant at a large historical society estimates that currently about 50% of their time is spent ingesting and processing, another 25% on program administration (documentation and training), and the remaining 25% on collaborations and support on digital projects.[51] Below are basic recommendations for the recommended staff position.

Main duties for this position would include:

- Accession, process, and describe electronic records for all units
- Maintain, update, or improve existing digital archives workflows and policies
- Address preservation and management of ASC electronic records
- Create new workflows and policies as needed
- Be the main contact for all ASC staff with questions about electronic records
- Contribute as needed to internal and external projects with electronic records
- Assist donors with the transfer of born-digital collections

Proposed General Skill Set:

- Strong organizational skills
- High level of attention to detail
- Experience with electronic record archival workflows
- Experience with tools relating to electronic record processes
- Experience working with various types of digital media such as floppy disks, USB drives, and DVDs
- Working knowledge of digital archive standards such as Trusted Digital Repositories and OAIS
- Demonstrated ability to work independently and to ask questions as appropriate
- Demonstrated ability to maintain focus while working on several projects concurrently
- Demonstrated ability to acquire new skills and learn new technologies
- Ability to problem solve and troubleshoot technical issues
- Strong online research skills
- Strong verbal, written and interpersonal communication skills.

[51] These time estimates are from email conversations with staff at the Minnesota Historical Society. These times reflect current estimates of one position and do not reflect expected future needs. January 10, 2017.

# Appendix I: Finding Aid Descriptive Guidelines

Provides guidance for incorporating born digital material into archival finding aids.

## Finding Aid Descriptive Guidelines Terminology

Please use the suggested standardized language below to differentiate between describing born digital and born analog material in order to enhance access and context.
**"Electronic Records"** when describing born digital records.
**"Digital Surrogate"** for material born analog that has been scanned or otherwise digitized.
**"Digital Files"** to represent a mix of electronic records and digital surrogates, or where there is question regarding whether or not the material is born digital or surrogates.
**"Archived Website(s)"** when describing one or more websites archived and made accessible via a web-crawler for archival purposes.

## Finding Aid Description

The description of electronic records and digital surrogates should be described in a manner that reflects the overall intellectual arrangement and level of processing carried out on the collection. How a staff member decides to describe material will depend on the intricacies and inherent original order of the collection. Acceptable description examples include:

### Collection Level

- Representation in collection-level extent statement, including:
    - total file size (e.g. in KB, MB, GB, etc.)
    - total number of files
- Access/Use of Materials Note:
    - Description of how and where digital files can be accessed (e.g. by following link for DAOs; accessible only on Reading Room computer, etc.)
- Descriptive mention in a collection-level scope note

**Example**

> **Extent:** 3 cubic/linear feet (# of boxes); 2 GB (# of files)
> **Scope and Content:**
> The collection includes paper material and digital surrogates. The digital surrogates represent paper documents digitized by the donor; these include scans of documents that are included in the collection in paper form as well as material that was retained by the donor in its original form.

### Series/Subseries or Record Group Level

- Collection-level description as described above
- Appropriately named series/subseries title, including:
    - Descriptive series/subseries scope note (if relevant)
    - Series/Subseries extent statement with file size and number of files

**Example**

> **Series.** Chairwoman Doe Administrative Records, circa 1995-2000
> **Extent:** 1.5 MB (125 files): 20 jpg, 18 pdf, 87 .docx
> **Series Scope and Content Note:**
> Electronic Records records created by Chairwoman Doe in their role as Chair of the Board of Directors. The records represent Doe's working files and includes draft and edited variations of project reports, .jpeg images and text and layouts used for the production of printed brochures.

### Folder Level

- Collection-level description as described above
- Folder title(s), including:
    - Circa date span(s) if relevant. Date span can be estimated by using either the last modified date of the files or estimated dates based on knowledge about the collection.
    - Extent statement(s) with file size and number of files (and optionally, file formats or types).
    - Scope note, if relevant
- If the folders containing electronic records or digital surrogates are part of a series or subseries, a descriptive note for the series may include mention of the digital files and/or the series title may include an extent statement with file size and number of files.

**Example:**

> **Series.** Digital Files
> **Scope and Content:**
> Primarily digital scans in .jpg format of correspondence, notes and annotations, photographs, and other materials also available in their original paper as part of this collection. These digital files include scans of some material for which the original was retained by the family. Dates given below refer in most cases to the dates the original, scanned material was created, rather than the dates the digital scans were created.
>
> Crandon, undated and 1925-1979. 114 MB (75 files).
> Photographs, brochures, graduation program, 75th anniversary materials, and memorabilia.
> "Dear Mary Katherine and Francis" manuscript, 1982. 51.1 MB (34 files).

### Item Level

*NOTE: It is our recommendation that item level description only be carried out when warranted and supported by the collection processing plan.*

- Collection-level description as described above
- If relevant, Series/Subseries or folder level description as described above.
- Folder title, date span if relevant
- Item title, including:
    - Date, if known/relevant
    - Extent statement with file size and file format
    - Scope note if relevant

**Example 1**

Images

1 House, address unknown, St. Paul, circa 1960
Digital Image Link

2 Family photograph, circa 1960
Digital Image Link

**Example 2**

Blank County files. 1 ZIP file (201,100 KB):
Content from Disc 1. 431 TIFF files and 1 TXT file.
Content from Disc 2:
ReadMe_CD2.txt
geo_clip folder. 431 TFW files, 431 TIFF files and 1 DB file.
geo_ref folder. 431 TFW files, 431 TIFF files, 1 DB file, and 1 AUX file.
Content from Disc 3:
items.txt
page_trs.xls

## ArchivesSpace Fields*

*See ASpace Best Practices for more information

Collection level

- Extent Subrecord will need specific extent measurement for size of electronic records.

Collection level notes:

- Abstract note updated if providing extent information is unit practice.
- Arrangement note will need to reflect electronic records
- Use and Access Restriction notes will need to be updated to reflect electronic records where relevant.
- Scope and Content note will need to provide information about extent and format types of electronic records.

Component record level

- Provide series or file level extent if appropriate for electronic record description.
- Provide Scope and Content note for series or file to aid access.

# Appendix J: Sample Special Use Case Agreement Language

The following provides examples of wording in existing use agreements that show how the Archives and Special Collections address providing access to sensitive materials. Full text of these agreements are provided in internal ASC Data Privacy Policies and Procedures document.

About recording information about collection:

- I hereby agree to maintain the confidentiality of materials in the _____ collection. I agree to omit in my note-taking any of the following personal identifying information: names, address, or other identifying location information.

- Researcher shall record Sensitive Personal Information and other data contained in the: _____ collection only in such a manner that subjects cannot be identified, either directly or through identifiers linked to the subjects.

Future use:

- I agree not to include names or other personal identifying information in any written or spoken communication in any format, including publication, teaching, conversation, social media or any other form not listed here.

- Researcher acknowledges that the intent of this agreement is to assure that any report, paper, article, or other item produced by Researcher in any medium will be indistinguishable from one that could have been produced if the _____collection contained no information that could be characterized as "individual health information" or "personally identifiable information" under any potentially applicable set of rules in effect at any time, and agrees not to use any method, approach, or technique for recording or memorization that could result in the disclosure of any Sensitive Personal Information in any medium, anywhere.

- I agree not to include names or other identifying information in any written or spoken communication in any format, including publication, teaching, conversation, social media or any other form not listed here.

Additional Review:

- I agree that, if requested, I will allow archives staff members to inspect my notes and computer/tablet/phone before leaving the archives, to confirm that proper names and identifying personal information have not been included in notes and no images of the documents have been taken.

Indemnification:

- Researcher agrees to indemnify the University, its regents, officers, agents, and employees, and defend them against and hold them harmless from any claims, suits, loss, and damage, including reasonable attorneys' fees, arising out of any action or inaction by Researcher related to Researcher's access to the _____ collection.

Other:

- I agree to not have any contact with any of the individuals identified in the records now, or in the future, either in writing, in person, by phone or in another format not listed here.

- I agree not to scan, photograph or otherwise reproduce restricted documents from the _____ collection.

# Appendix K: Deed of Gift Addendum for Electronic Records

**University of Minnesota Libraries**
**Deed of Gift Addendum for Electronic Records**

This addendum addresses issues related to donations in the form of electronic records.

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 also agree that if in the future I wish to deposit substantially altered versions of these materials to an entity or institution, the University of Minnesota has the right of first refusal before they may be offered to any other entity or institution.

**I. Discretionary practices specific to electronic records**
I agree that the Libraries, at its discretion, in accordance with University policy and with applicable law, may do 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**
I have indicated by my initials below whether or not these materials 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:
  _____
  _____
  _____
  _____

I acknowledge that the Libraries will review the materials in an attempt to identify items that contain personally identifiable information and/or private information (as defined by applicable state and federal laws and regulations) and redact such information, and agree to notify the Libraries of any need for additional redaction.

**III. Disposition of Storage Media**
I acknowledge that 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 I have indicated by my initials that I prefer the media to be returned to me.

___ 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