Developing Policies, Procedures and Workflows for Stewarding Born Digital Collections
in the University Library and University Archives
at the University Of Illinois Urbana-Champaign

BY

TRACY POPP

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Certificate of Advanced Study in Library and Information Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

CAS Project Committee:

        Associate Professor Jerome McDonough
        Professor Chris Prom, Chair
        Assistant Professor Kyle Rimkus

Abstract

Broadly, electronic records are presenting libraries and archives with a new set of challenges. The goal of this project is to develop the first iteration of a set of policies, procedures, guidelines and workflows to address issues specific to stewardship and preservation of born digital materials held in at the University of Illinois Urbana-Champaign University Archives collections and establish a base of formal policies, procedures and workflows for the Preservation Unit's Born Digital Content Preservation service staff to consult when assisting others with born digital Library collections materials. This paper outlines the project development, methodology and first iteration of documentation produced to meet this goal, and to begin coordinated efforts to steward born digital collections at the University of Illinois Urbana-Champaign's University Library and University Archives.

Table of Contents:

Introduction

Broadly, electronic records are presenting libraries and archives with a new set of challenges. Traditional library functions and archival actions of collecting and accessioning, arranging and describing and providing access are applicable regardless of the content carrier; electronic records are subject to the same general processing as paper-based materials – albeit the tools and methods used to do so are necessarily distinct from those of paper records due to distinctive material and structural differences. As stated in the paper *An Approach to the Preservation of Digital Records* (Heslop, Davis & Wilson, 2002), "Digital records, while fulfilling the same general business purpose as paper records (reports, letters, memos), are inherently different from their paper counterparts. The most obvious difference is that digital records are mediated by technology, which means that to experience digital records a person must have the right combination of hardware and software." (p. 8).

Underscoring this difference between physical, paper-based materials and electronic records is that paper-based materials are eye-legible; they are not dependent upon technological infrastructure or layers of interpreters such as computer hardware and software in order to read, interact or observe them. Information loss, alteration, or physical degradation can generally be detected through visual perception; observing that a piece of paper has been torn, is starting to degrade or that the text on the page has been modified in some way. Physical properties of paper-based materials and general stability, along with other contextual information, aid in establishing a document's authenticity and reliability.[1]

[1] According to Duranti (1995), the terms authenticity and reliability aid in establishing a records trustworthiness with reliability meaning the record represents an action or process and is, "treated as the fact of which it is evidence" (p. 7). Authenticity, by Duranti's definition (a1995) and in context to records, is, when a document is

By their nature electronic records are subject to rapid change and can be easily altered, through both intentional and unintentional actions. These changes are mediated through complex systems of computer hardware and software. Changes to a file or bit stream may be imperceptible through visual inspection –one may not understand that a magnetic field has changed or erased bits from a disk by simply looking at the disk. The bits representing informational content are not easily understood or perceived without hardware and software to interpret and present them, or as Heslop et al. (2002) assert, to render a "performance of the record". In order to maintain authenticity of the original record, processes employed in managing electronic records should be consistent and traceable, account for inherent volatility, provide transparency, avoid unintentional or opaque data and metadata changes in order to maintain provenance[2] and trustworthiness – characteristics that are crucial to maintain in archival records regardless of information carrier.

The University Archives (abbreviated as UA for the remainder of this paper) at the University of Illinois Urbana-Champaign has a well-established documentation and workflows related to acquiring and processing paper-based material; however, their policies, procedures and workflows for stewarding born digital materials is less mature due to the relatively recent influx of digital materials accessioned by the Archives which require a significantly processes than previously established for paper-based materials. Staffing changes at the University Library and nascent and evolving born digital content stewardship communities of practice have also necessitated review, coordination and development of policies, procedures and workflows for

---

what, "it claims to be" (p. 7). These are key components in establishing record trustworthiness, where a record can be trusted as evidence and stands as what it purports to be (Duranti, 1995).

[2] Another fundamental principle in archives, provenance refers to, "the individual, family or organization that created or received the items in a collection" (Pearce-Moses, 2005); it can also refer to the chain of custody.

stewarding born digital materials at the University of Illinois Urbana-Champaign University Library (henceforth addressed as the Library) and UA.

Project Development

Project Goal and Scope

The goal of this project is to develop the first iteration of a set of policies, procedures, guidelines and workflows to address issues specific to stewardship and preservation of born digital materials held in UA collections and establish a base of formal policies, procedures and workflows for the Preservation Unit's Born Digital Content Preservation service staff to consult when assisting others with born digital Library collections materials. These policies, procedures and workflows are intended to lay the foundation for future projects in which the Preservation Unit consults or collaborates with other Library units to acquire, process and preserve born digital collections.

Informed in part by the framework established in the publication "AIMS Born Digital Collections: An Inter-Institutional Model for Stewardship" (AIMS Work Group, 2012) this project is intended to be the first iteration of establishing a working relationship between the UA and the Library's Digital Preservation Coordinator to create policies, procedures and workflows which address critical digital lifecycle and archival actions at the pre-accession and accession processing stages. This project is scoped to these particular phases of archival processing per the AIMS whitepaper statement (2012), "decisions made at the beginning of the process will have a direct impact on later outcomes" (p. 1) which may affect provenance, reliability and future preservation efforts. The RGL report *Attributes of Trusted Digital Repositories* (2002) asserts that preservation actions should begin as early as object creation to influence positive

preservation outcomes stating, "preservation requires active management that begins at the creation of the material" (p. 18). Influencing content creation can assist in mitigating later digital preservation problems through advising content creators about particular issues which may include educating about file format selection and metadata creation.

Born digital physical media received by the UA often runs the gamut of computing history; media that have a projected lifespan of five to ten years may have aged well beyond that lifespan become obsolete, and are in need of "retrospective preservation" (Digital Preservation Coalition, 2008, p. 93) and thus will require more resources of time, technology, and funds. Although the time to influence the donor at the content creation stage has long past, stabilizing born digital collections as much as possible upon receipt by the collecting repository is a key initial step. Gretchen Gueguen, Digital Archivist at University of Virginia and one of the principal authors of the AIMS report advises that content stabilization attention should be given early within archival processing, stating "appraisal and accession steps are crucial – stabilize and you can arrange and describe later" (Gueguen, April 2013). Assuredly, if media is unreadable due to failure or unavailability of necessary hardware and software or knowledge to use the hardware and software the content suspended on declining digital media will not be accessible to arrange and describe.

Other factors influencing the scoping of the project are time and departmental focus. One project goal is to develop an initial set of policies, procedures and workflows for work between the UA and the Preservation Unit. The project began in earnest in January 2013 with completion of documentation initially scheduled for May 2013. In April 2013, I (Tracy Popp, Digital Preservation Coordinator) requested a project extension to accommodate additional material and subsequently re-scheduled the completion to August 2013. As the project has a finite timeline,

production of a fully comprehensive set of documents covering each aspect of processing born digital materials throughout their lifecycle was deemed impractical within the timeframe established. Additionally, most of the documentation created targets digital preservation concerns – overtly and indirectly – giving priority over other types of documentation that may be important from a curatorial standpoint, but which are less critical in terms of preservation. Although tangentially influential in future digital preservation efforts, actions related to arrangement and description, discovery and some access aspects are out of scope in terms of Preservation Unit responsibilities. Content managers such as archivists and curators are more familiar with the intellectual contents of their collections and are better suited to address intellectual arrangement and description. Preservation Unit staff may work with other Library staff on discovery and access points and may advise on tools related to these actions, but these issues are not the sole responsibility of Preservation Unit staff and, in terms of this project, are considered out of scope.

Project Stakeholders

As the intent of the project is to develop a base set of policies, procedures, guidelines and workflows for stewarding and preserving born digital content held in Library collections, with an initial focus scoped to materials in the UA collections and to formalize the Born Digital Content Preservation service within the University Library's Preservation Unit, the project stakeholders comprise representatives from UA and the Preservation Unit:

- o Dr. Christopher Prom – Assistant University Archivist, Professor of Library Administration
- o Kyle Rimkus – Preservation Librarian, Assistant Professor of Library Administration

        o   Tracy Popp – Digital Preservation Coordinator and CAS project manager

Given that the scope of this project is limited to pre-accession and accession stages and establishing workflows between the UA and Preservation Unit, this stakeholder list is thus limited to representative stakeholders from UA and Preservation Unit. In future work the representative stakeholders will expand to other areas within the Library including representatives from special collection units such as the Sousa Archives and Center for American Music. Stakeholders external to Library staff, such as collection donors and Library and Archives users, may also be represented in future work via interactions with collection managers and archivists as the policies, procedures and workflows also affect this population.

Meeting With Stakeholders and Establishing Expectations

The project is intended to serve academic, professional and practical needs, and in fulfilling the roles of CAS student and Digital Preservation Coordinator, I served as primary manager of project development and project documentation author. Project management and discussion on documents produced occurred through periodic in-person meetings between Chris, Kyle and me, through email, and through a shared cloud storage service account where project stakeholders had access to various deliverables and work in progress.

In the project kickoff meeting I asked Chris and Kyle about their expected outcomes. Chris identified the following needs for UA, summarized as:

- Create a set of policies, procedures, guidelines and workflows to manage the activities associated with preserving born digital records; primarily, the management of activities between UA and Preservation Unit relating to the transfer of content from obsolete storage media to networked storage in preparation for ingest into the Medusa Digital

Preservation repository[3]

- Retention of computer disk images generated from migrated digital storage – recommendations on why and when to retain images, and how to best store these images in a storage-efficient, low-cost manner

- Generation of preservation metadata, documentation and audit trail as events and actions occur along the workflow

- Recommendations of tools (if extant) for improved processing, archival arrangement, and description

As supervisor of the Preservation Unit, Kyle articulated his expectations for the project, summarized as:

- Develop a collection of workflows, policies and procedures on which to base future born digital collection preservation actions

- Develop a documentation framework for reformatting born digital media in coordination with UA from which to generalize service provisions for other Library Units

I have served in the role of Digital Preservation Coordinator since starting the job in October 2011 and had previously met with both Kyle and Chris; these expectations aligned with my own expectations and goals for this project.

New Position at the Library

A major factor in developing this project was that the Digital Preservation Coordinator position had not existed in the University Library prior to my employment in October 2011; it is a new position in the Preservation Unit and the University Library. Although digital preservation activities occurred in various locations throughout the Library, efforts were often decentralized, with practices and workflows largely developed ad hoc. Prior to October 2011, the Library's

---

[3] Medusa is the given name of the Library's Fedora-Hydra-Blacklight based digital preservation repository currently under development. Related policies, procedures and workflows for Medusa are also in development.

Preservation Unit did not have permanent staff dedicated to digital preservation or born digital media reformatting activities. During the spring of 2010, a pilot project was conducted to gauge the needs for and feasibility of a digital media reformatting program. This pilot resulted in the transfer of nearly five hundred floppy disks - 3.5" and 5.25"; primarily Windows\PC-formatted (Popp, 2010) - and aided in justification for developing a new full-time position within the Preservation Unit.

The Digital Preservation Coordinator position was established within the Preservation Unit in 2011 to provide, in part, a centralized service point to assist Library units with stewarding data stored on obsolete storage media. This position works in a highly-collaborative manner with Library Units as a matter of necessity – the Digital Preservation Coordinator should have a strong understanding of the curatorial goal behind maintaining the content and the respective Library Units often do not have the necessary equipment or expertise to acquire, process and steward born digital content. To underscore the importance of collaboration in stewarding born digital materials the authors of the AIMS report (AIMS Work Group, 2012) observed that, "As [we] worked collaboratively to design procedures for accessioning and processing born digital materials, we discovered that these cannot be isolated activities carried out by one department, or by one staff specialty, or apart from the rest of the archival management workflow" (p. 1). This need for collaboration highlights the need for documented procedures, policies and clear workflows. I recognize the importance of formally establishing policies, procedures and managing workflow and expectations regarding born digital collection materials preservation and defining the Born Digital Content Preservation service as a key part of proving the service as stable and trustworthy.

Born Digital Media Reformatting Lab

The cornerstone of digital content reformatting activities within the Preservation Unit is the born digital media reformatting lab. The lab is designed as a centralized service within the Library for assisting the distinct Library and Archives Units with preservation of and access to born digital content. Specifically, it is designed to provide in-house reformatting capabilities for a variety of computer-based physical media created in PC and Mac computing environments including:

- 5.25" floppy disks

- 3.5" floppy disks

- Iomega ZIP and JAZ disks

- Optical Media: e.g. data CDs and DVDs

- Hard disk drives including internal and external devices and various types of interfaces

- Flash media including jump\thumb\flash drives and a variety of media cards

The lab has various tools and software for content analysis, viewing, select file migration activities and, in the future, will include emulation tools and a user-access workstation for accessing content that has particular hardware and/or software requirements and dependencies which are not easily replicated or ported to contemporary computing environments.

Presently, the lab's physical space is very limited – with activities occurring in the Digital Preservation Coordinator's workspace in Room 44 in the Main Library. Planning is currently underway to move the Preservation Unit to another location within the Library which will include a larger space with a dedicated Born Digital Media Reformatting lab and digital content access workstation. This larger lab area will allow digital media reformatting operations to grow and create greater workspace for additional staff. Successful program growth with consistent

and reliable outcomes is dependent upon established and vetted policies, procedures and workflows to ensure that all working within the born digital media lab consistently produce secure, verifiable and trustworthy digital objects.

Methodology

Publication Analysis

To start the process of identifying, planning and designing policy, procedural, and workflow documents, I began by analyzing five publications that focus on major archival processing activities, with four of the five relating to processing born digital collections: *AIMS Born digital Collections: An Inter-Institutional Model for Stewardship* (AIMS Work Group, 2012), the multi-institutional effort *Born Digital: Guidance for Donors, Dealers and Archival Repositories* (Forstrom et al., 2013), Gregory Hunter's *Developing and Maintaining Practical Archives* (Hunter, 2003), J. Gordon Daines' "Module 2: Processing Digital Records and Manuscripts" (Daines, 2013) and Martin Gengenbach's Master's thesis paper, *The Way We Do It Here: Mapping Digital Forensics Workflows in Collecting Institutions* (Gengenbach, 2012). These publications were selected based on their respective approaches to identifying steps or workflows in archival processing. Although Hunter's book is primarily focused on processing paper-based materials, and the others born digital, it was consulted as a basis for identifying basic archival processes and comparing how the general terms and processes mapped to those established in publications that address born digital content in particular.

I began my analysis with *AIMS Born digital Collections: An Inter-Institutional Model for Stewardship* framework due to its influence on my initial work practices. This publication is the result of a collaboration developed by participants at four cultural heritage institutions, located in

the US and UK, representing the experience and methodologies employed in processing hybrid collections of born digital and paper-based archival materials. The framework creators recognized the need to "acknowledge established practices and infrastructure within archive institutions for paper-based collections, the existence of hybrid collections (those consisting of both digital and paper-based materials), and the existence of legacy materials transferred in the past and still stored on donors' physical media." (AIMS Work Group, 2012, p. iii). The authors worked together to reach, "consensus on the scope, purpose and definitions of the main archival activities…" (AIMS Work Group, 2012, p. vii). They identified four high-level functions, "that should be thought of as sequential steps in a very high-level workflow."(p. 1): collection development, accessioning, arrangement and description and discovery and access. The publication expands on these functions through objectives which are then "further detailed through expected outcomes, decision points and tasks" (AIMS, 2012, p. 2). One marked difference in the AIMS framework compared to standard archival processing workflows is that appraisal is not a defined as a discrete function. Under this framework appraisal activities are iterative, with analysis and actions occurring at several functional points; it is part of, "any or all of the first three functions of the framework" (AIMS, 2012, p. 2).

The authors note that the framework, by design and through the process of working with institutions of varying size, and respective strengths, is flexible and, "that is by its nature not institution-specific." (AIMS, 2012, p. iii) It has also been developed to comply with varying established standards applicable to working with materials in cultural heritage institutions – notably the Open Archival Information System (OAIS), which is a fundamental standard for digital preservation systems. In most of the publications reviewed, flexibility is a term frequently encountered; expressed as a requirement to make guidance documents flexible enough to aid

varying types of institutions, and to withstand workflow modifications often necessary in the rapidly evolving domain of processing born digital materials.

Out of a requirement to make the framework usable in broad contexts, the AIMS partners designed it and tested it via practical application. The publication includes case studies from each partner institution where the partners implemented the framework, which include information about tools used at each partner institution and high-level workflows for processing a specific collection.

The publication *Born Digital: Guidance for Donors[4], Dealers, and Archival Repositories* focuses predominantly on the "issues and concerns related to born digital acquisitions" (Forstrom et al., 2013, p. 4). The underlying concern is preservation, with authors providing recommended actions for entities submitting materials to an institution and the institution itself in order to maintain the provenance and reliability of born digital objects. This front-loading of focus on preservation allows for considerations and actions on maintaining the longevity of born digital materials to occur before content arrives at a repository. The idea of working with donors or creators prior to transferring content to an institution aligns with the AIMS guidelines of preservation concerns and donor engagement being addressed early within the archival process – indeed, prior to formal acquisition - in order to "ensure the integrity and usability of digital materials over time" (Forstrom et al., 2013, p. 5). Compared to the AIMS publication, *the Born Digital: Guidance…* publication aligns some activities which fall under AIMS Collection Development activities and places them within a major activity identified as the "Initial

---

[4] In the Forstrom et al publication, donors is defined as, "refer[ring] to the person, family, organization, estate, or other entity that sells, gives, deposits, loans, or otherwise transfers born-digital materials to an archival repository" (Forstrom et al., p. 3). In this paper, donor is used in the same manner.

Collection Review" (Forstrom et al, 2013, p. 5). For example, the collection review and survey is situated in a pre-acquisitions activity in *Born Digital: Guidance…* whereas similar activities within the AIMS publication are carried out in the Collection Development stage. Although the actions are identified in sections using different terminology in each document, the underlying intent of the action within the workflow is similar – to maintain "the integrity and usability over time" (Forstrom et al, 2013, p. 5). Although the inconsistency in terminology may be confusing, basic archival processing activities aligned with acquisition are demonstrated including suggestions on how to physically gain control of the media and associated contents.

Gregory Hunter's *Developing and Maintaining Practical Archives* (Hunter, 2003) is a general *how-to-do-it manual* outlining the process of establishing and managing archives. Hunter's publication provides guidance through the high-level archival actions of selection and appraisal, acquisitions and accessioning, arrangement, description, preservation and access and guides readers with archival principles and theory. The text is designed to aid readers through issues related primarily to paper-based archives, although sections on digital records and other content-delivery formats, such as audiovisual materials, are provided. In contrast to the publications explicitly addressing digital records processes preservation topics are primarily focused on understanding issues such as environmental control that largely affect paper such as humidity, light and air quality - although these environmental issues certainly also affect physical digital media. Consultation of this text was more in the interest of identifying format agnostic and basic archival processes and comparing how the general terms and processes mapped to those established in publications that specifically address born digital content.

J. Gordon Daines III's *Module 2: Processing Digital Records and Manuscripts*, which is available as a stand-alone publication or as part of the *Archival Arrangement and Description*

series, expands upon and provides practical examples for addressing the statement from the AIMS report that, "the challenges of stewarding born digital material demand new strategies as well as a redefinition of archival workflows" (AIMS Work Group, 2012, p. i). Of particular interest is section two where Daines "discusses how and why archival processing workflows can be adjusted in order to accommodate digital records and manuscripts." (Daines, 2013, p. 91). In section two, Daines describes the tasks and sub-tasks in traditional archival processing, based on the steps provided in Kathleen D. Roe's, *Arranging and Describing Archives and Manuscripts* which include: accessioning records, gathering contextual information about the records, performing a conservation assessment, establishing an arrangement scheme/intellectual arrangement, arranging the records physically, if necessary, describing the records and creating access tools. (Daines, 2013, p. 100). Daines provides an example of a born digital processing workflow based on traditional archival processes, but re-arranged to accommodate the inherent instability and needs of electronic records. Although he doesn't cite the report when providing the example workflow, it does seem that Daines follows the AIMS framework as he advocates for a pre-accessioning or collection development phase that is not reflected in the traditional archival workflow, but is present in the AIMS framework.

Gengenbach's Master's thesis paper, *The Way We Do It Here: Mapping Digital Forensics Workflows in Collecting Institutions* (Gegenbach, 2012), presents the results from a study, "to explore the ways in which collecting institutions have implemented digital forensics tools and practices into workflows for acquiring and preserving born digital content." (Gengenbach, 2012, p. 21). This study provides example workflows created through information gathered from interviewing nine participants about their born-digital processing methodology. Although the study focus is on implementation of forensics tools within born digital processing

workflows the examples provided within the report illustrate how these institutions are applying archival principles to their respective workflow and identifying where forensics tools are best applied.

A number of the study participants were active in developing the AIMS framework and have implemented AIMS framework processes within their workflows (or, conversely, their workflows informed the basis of the AIMS framework). This paper was of particular interest due to my use of digital forensics methods in processing born digital materials. Examination of the examples provided further guidance as I considered how to formalize my workflow and develop documentation to support the various archival activities.

There are several commonalities among the publications that focus on born digital content processing. In each, actions and processes that address traditional archival processing activities such as acquisition and appraisal are included – although not all of the born-digital publications focus on every major aspect of archival processing. Daines (Daines, 2013) references "basic principles of archival arrangement and description as described in Kathleen D. Roe's book *Arranging and Describing Archives and Manuscripts*" (p. 91) in mapping electronic records processing to traditional archival processes. The AIMS Framework developers agreed early on that the, "framework would need to acknowledge established practices and infrastructure within archives institutions for managing paper-based collections…. (p. iii)' and that, "In line with our general research methodology, this work translates traditional archival principles and practices into a born-digital context" (p. vii). However, the AIMS partners recognized that mapping traditional archival terms to a born digital context would not be trivial, noting challenges of aligning terminology between the US and UK and between archives and digital library domains (p. viii). In *Born Digital: Guidance*… (Forstrom et al., 2013) the authors

represented two professional groups: archivists and special collection curators. They recognized that terminology used reflects, "an unavoidable professional bias" (p. 3). Although task recommendations are not explicitly mapped to traditional archival processing, the authors use terms originating from traditional archival processing such as acquisition. In Gengenbach (Gengenbach, 2013) although not specifically articulated as aligning with traditional archival processes, workflows provide evidence of mapping to traditional archival methods through the vocabulary used and identified major steps such as acquisition, arrangement and description and access.

While aligning with some traditional archival processing practices, the born digital publications have reordered and added aspects which better recognize concerns specific to born digital materials. For example, each born digital publication emphasizes engaging with donors and creators prior to committing to acquisition. Daines notes (Daines III, 2013), "archival processing of hybrid and digital-only collections begins before the decision to acquire materials" (p. 114). Pre-acquisition steps may assist with providing additional context to collections, which in turn can aid in arrangement and description activities and long-term preservation goals. Daines states (Daines III, 2013), "We cannot wait for materials to arrive at our repositories. We must proactively engage with the creators of digital manuscripts and records to ensure that appropriate information is being captured before and during accession" (p. 125). In his thesis, Martin Gengenbach observed (Gengenbach, 2013), "In what will be a continuing theme in many of the interviews I conducted, the work of acquiring born-digital content from donors begins long before the material arrives at the archives" (p. 27). Although Gengenbach determined that pre-acquisition work was out of scope for his thesis, he remarked (Gengenbach, 2012), "the importance of developing relationships with donors early on to facilitate the transfer of born-

digital content to collecting institutions is a theme that has been widely discussed in the archival literature" (p. 27-28) and provides examples of said literature, including a citation of the AIMS Work Group paper, which defines its first major step as "collection development" (AIMS Work Group, 2012, p. 5). In AIMS nomenclature collection development is, "the actions and policies of an institution to acquire material for end-users" which, "form the basis for subsequent actions and decisions undertaken by the institution as they accept stewardship for and legal ownership of materials from a donor, creator, or seller" (AIMS Work Group, 2012, p. 4). Collection development activities were initially referred to as "pre-accessioning intervention" as, "it highlights attempts to determine what actions should be undertaken or what information should be gathered in order to lay a solid foundation for the long-term stewardship of born digital archives" (AIMS Work Group, 2012, p. 4).

In contrast to traditional archives workflows, which, if conceptualizing actions in a linear fashion, may place conservation and preservation activities after acquisition and arrangement, born digital publications emphasize preservation considerations earlier in the process. AIMS advises for rapid, unencumbered content stabilization (AIMS Work Group, 2012), "in particular, accessioning and processing workflows need to allow for and enable digital materials to be transferred to managed storage as soon as possible to ensure preservation of bitstreams" (p vii). Forstrom et al advise on careful transfer of content to the repository in order to preserve acquired content (Forstrom et al, 2013), "in order to ensure that born-digital materials arrive at repositories in good condition and accompanied by appropriate documentation, it is vital to convince donors, dealers, repository staff, and others to be mindful of how they handle, document, ship, and receive digital media and files" (p. 4). Daines also notes the importance of careful transfer of contents as a preservation concern, noting (Daines, 2013), "the sub-task most

impacted by digital records and manuscripts is that of taking physical control of the materials. The key lies in taking custody in a way that preserves the authenticity of the digital records and manuscripts" (p. 101). Daines also re-aligns tasks associated with traditional conservation activities to the accessioning step (Daines , 2013, p. 104). Although not explicitly stated, in reviewing the workflows as documented by Gengenbach (Gengenbach, 2012) each institution stabilizes born-digital content by migrating the bitstreams to managed networked storage prior to arranging and describing content; this step is often necessary to begin engaging with the content as some level of access must be provided to read contents stored on the physical media.

Although archives and other cultural heritage repositories have been encountering born digital materials within the recent past, broadly applicable best practices are not yet fully established, mature or many not yet exist. Initially tasked to identify best practice in processing digital collections, partners in the AIMS project recognized early in the project process (AIMS Work Group, 2012), "that the development of best practice within born-digital stewardship was not yet possible" (p. ii) due to lack of available tools, and rapidly shifting technological and institutional environments . In recognition of these limitations, "the AIMS Framework, therefore, was developed to define good practice in terms of archival tasks and objectives necessary for success" (AIMS Work Group, 2012, p. ii). Forstrom et al (Forstrom et at, 2013) also use the term good practice instead of best practice in providing recommendations.

With the lack of firm, best practice and ever changing working landscape, the born digital publications note that workflows and practices are ever-shifting, which necessitates review and iteration as well as a need to implement less than perfect solutions in favor of progress. Daines advocates for starting with small steps and to avoid inaction (Daines III, 2012), "don't let the perfect become the enemy of the possible. Look for short-term solutions that that can help bridge

the gap between current practice and a long-term solution" (p. 126). Gengenbach observed

through various interviews with individuals at institutions engaging in born digital archival

processing that (Gengenbach, 2012),"many participants described workflow creation as an

iterative process, gauging not when a workflow was finished, but when it was usable for their

institution, with the understanding that it would continue to develop" (p. 71).

Faced with developing good practice and processes, current publications also reflect

language of flexibility; terms such as "recommendations" and "guidelines" signal that born

digital publication authors are taking less of a prescriptive approach and offering more examples,

guidelines and suggestions to others embarking on stewarding born digital materials. Daines

provides a sample mapping of archival processes to business processes and notes (Daines, 2013),

"the application of various tasks and sub-tasks described as part of the archival processing

business process will vary from institution to institution" (p. 100) thus providing readers with

recommendations instead of imposing a strict prescriptive methodology. Gengenbach

(Gengenbach, 2012) remarks that documenting generic and standardized, "workflows for

capturing and preserving born-digital content helps to minimize unforeseen changes that force

archivists to be reactive, rather than proactive" (p. 72). The AIMS Framework is developed for

scalability and flexibility; "in developing the AIMS Framework, the project would apply a

practitioner-based research approach by developing a model based on real case studies of

collections at each institution. Applying our theories would confirm or challenge the initial

framework, which could then be used as a model around which to build individual workflows

and processes within each partner's organization" (AIMS Work Group, 2012, p. ii). Forstrom et

al use the term recommendation and also take a general approach (Forstrom et al, 2013), "these

recommendations are not meant to be universal and do not necessarily reflect the official policies

of the co-authors' institutions. Rather, they offer broad, useful guidance for donors, dealers, and repository staff involved in the acquisition of born-digital materials." (p. 4)

Documentation Needs Analysis and Workflow Diagram

After gathering requirements and expectations from stakeholders, and concurrently analyzing publications focused on archival processing workflows I began a workflow and documentation needs analysis. I began by reviewing the documentation UA established for processing born digital materials. This stage of analysis aided in identifying gaps, determining areas within the workflow identified as logical stages for the Born Digital Content Preservation service to consult or manage specific activities, and to ensure the Born Digital Content Preservation service was operating in complimentary manner with Archives working methods and procedures.

In comparison to the recommendations identified in the publications listed above, particularly the *AIMS Framework* and *Born Digital: Guidance…*, the UA documentation was, understandably so, written to address born digital content transferred to and already acquired by the UA. In its current form is untested as scalable solution; documentation refers to processing methods which are only applicable to UA staff such as identifying the Archives network share as a content storage location. Donor rights and expectations on access, media and file retention were not addressed in the documentation, although these issues may be addressed in a Deed of Gift. How viruses are handled is addressed in staff documentation, but there is no disclosure of action with a target audience of the general public/donors. Also, documentation developed for transferring contents was developed prior to the Digital Preservation Coordinator position. Some documentation is outdated and does not reflect current practice nor disclose the use of digital forensics technology in media access and preservation activities. Processes and actions should be

made transparent to address ethical concerns and to ensure consistency in processing collections.

The University Archives documentation examined in this stage of analysis can be found in

Appendix A: University Archives Documents.

On the Preservation Unit's side, there was little to no formal documentation about the

Preservation Unit's policies, procedures and workflows with regard to born digital content. I had

compiled many notes and had informational documentation which I referred to, but

documentation was largely incomplete not written with a wider audience in mind. Thus, there

was an abundance of opportunity to create documentation relating to the Born Digital Content

Preservation service's activities.

Workflow Diagram

The workflow diagram is another planning document created as part of this project. A

workflow is a step-by-step illustration of a process, often used to illustrate business processes to,

"document responsibility of tasks" and processes (Richards, 2010). In this project, the workflow

diagram graphically represents select actions identified in the publication analysis stage and

aligns actions with entities responsible for completing particular actions. The workflow is

vertically oriented and uses the swim lane diagramming method of identifying units responsible

for particular parts of the process. There are four swim lanes, representing four entities which are

part of the born digital content stewardship process:

- **Donor:** a term defined in "Born Digital: Guidance for Donors, Dealers, and Archival Repositories" (Forstrom et al., 2013) as, "the person, family, organization, estate or other entity that sells, gives, deposits, loans or otherwise transfers born digital materials into an archival repository."(p. 3)

- **Archives:** represents the actions and responsibilities completed by an individual within the University Archives

- **Preservation:** Represents the actions and responsibilities undertaken by the Digital Preservation Coordinator within the Preservation Unit

- **Media Reformatting Vendor:** An external service provider contacted by the Digital Preservation Coordinator to complete select tasks which cannot be done within the Preservation Unit.

The documents in the workflow represent points where action, decision, or guidance may occur or may be beneficial. The workflow contains the most detail in areas where the Born Digital Content Preservation service is primarily responsible.

Within the workflow, several documentation types are identified in relation to points of action. These documentation types vary based on the intended documentation function and intended audience. Function outlines what the documentation is intended to communicate – are there step-by-step procedures involved? Is the documentation intended to disclose methods of content transfer to a broad audience? The audience represents who the documentation is written for and intended to assist. Audience is generally broadly indicated as public (anyone external to the Institution) and internal (those within the institution or within Library and Archives Units.) The types of documentation created and identified within the workflow are:

- **DOCUMENT**: Asset created completed or created through archival\reformatting actions; may become part of the object's archival record; may record digital object provenance information

- **POLICY:** rule to guide decisions and convey institutional intent; implemented in procedures; developed and implemented internally, but also functional as an informational guide for donors

- **STATEMENT:** official expression of views and approach; intended for a public audience

- **GUIDELINE:** General advisory document based on policy; intended as an internal document

- **PROCEDURE:** Step by step actions to complete an action; internal document use to produce consistent results and output

- **INFORMATION:** Informational document; can be for an internal or public consumption

- **FORM:** Document used to gather information; may be used internally, or between a Library or Archives Unit and a donor

The workflow diagram was produced using Microsoft Visio, a diagramming and vector-graphic illustration program. Most of the detail lies within the pre-accessioning and accessioning activities. Due to the complexity, the original diagram is very large – a legible print of the entire diagram would measure approximately 25" x 46.5". Due to the size, the entire diagram is not included with this paper. A PDF version will be included in the supplemental materials.
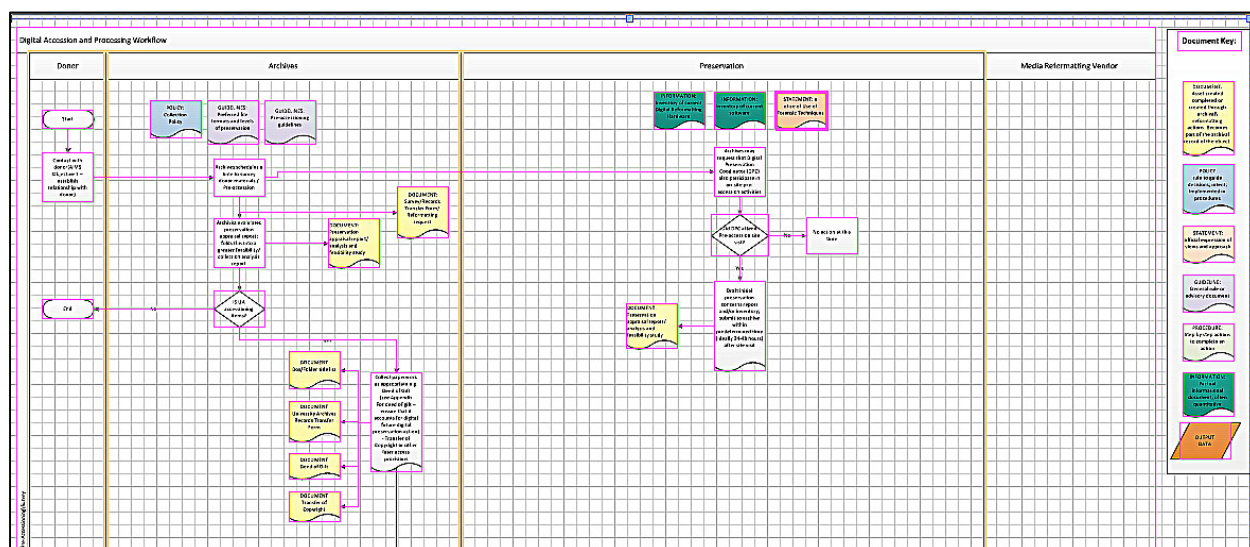

Figure 1: Screen capture from a section of the Workflow Diagram produced in Visio

Priority Matrix

Given time limitations and the desire to establish initial procedures and policies I developed a documentation priority matrix, which contained a list of documentation identified in the workflow diagram. Within the matrix I identified priority levels of high, medium and low. Priority was established based on criteria such as: the overall impact of establishing the document – would its creation result in potentially increasing the digital object viability within

its lifecycle? Would a lack of documentation, guidance or any actions overlooked at earlier stages of processing affect the viability or reliability of the object or negatively affect preservation actions undertaken at a later time? Other points considered when establishing priority was the number of representatives from other Library and Archives Units or committee review required to reach consensus prior to implementation. If a document was related to activities which collection managers would be responsible for or required approval from a larger group in order to reach a consensus, due to time constraints, the document was given a low-priority score. If I was able to author the documentation and the scope of action and authority was largely limited to the work completed within the Preservation Unit over which I had authority, and had high preservation impact it received a higher priority. Due to the size, the entire matrix is not included with this paper. An example of the worksheet is provided in Figure 2. A copy of the Excel-produced worksheet will be included in the supplemental materials.

The priority matrix also served as documentation tracking tool. I used it to identify various stages of document development – tracking documentation which existed in some form, but which may require editing to address born digital content concerns, tracking documentation slated to be included in the CAS and other development notes. As such, the priority matrix was used as a working document and was modified as documentation was edited, or created. It is a centralized location where I could quickly note the status of a particular document.

Document Type	Document Title	Author(s)	Does this document exist in any form?	If exists File location/URL:	In its current form does this document require editing?	Does this document refer solely to born-digtal collections?	Notes On Document	Priority
POLICY	Collection Policy	University Archives; Preservation Unit advisor for born-digital content	yes	MISSING from Web site: http://archives.library.illinois.edu/about/policies/collection-policy/	possibly, Chris to advise	no	Collection Policy	Medium - depends upon what role born-digital plays and if were disclosing born-digital collection methods here
STATEMENT	Issues and Ethics of working with born-digital materials	DLARSCAG group?	no			yes	University of Illinois Libraries Statement on Issue and Ethics of Working with Born-Digital Materials; (Is this too broad to develop at this time for library-wide application? A document such as this might be better vetted in the DLARSCAG group and may be out of scope for this project); may also cover	Low
POLICY	Preferred file formats and levels of preservation	University Archives; Preservation Unit advisor for born-digital content	yes	http://archives.library.illinois.edu/deposit-records/transfer-electronic-records/	no	yes	Achives has some file formats listed as preferred; I would like to have at least some baseline of preferred file formats and levels of preservation to give a better understanding of preservation and expectation management;	High
GUIDELINES	Pre-accessioning guidelines	University Archives; Preservation Unit advisor for born-digital content	no				dependent upon and points to several other documents; is an overarching document to guide in making accessions decisions	Low

Figure 2: Partial view of Priority Matrix

Out of Scope Documentation:

Although partially based on publication analysis not all of the recommendations made in the publications were expressed in the workflow diagram or priority matrix. Based on the criteria listed in the priority matrix and time constraints, some points within the workflow identified as requiring documentation were determined to be out of scope for this project and will be reconsidered at a future date. Documentation that was primarily out of the authority of the Preservation Unit, not related exclusively to born digital materials, or closely related to other documentation produced and serving as exemplars were not created for this project. Documentation focused on interactions between the collection manager and donor are excluded or addressed at a very high-level. One is example of an out of scope document is modifying the UA collection policy to address the collection of born digital materials. This document is under the authority of the University Archivists and was written in a manner that is intended to be

broad and format-agnostic. Further analysis and discussion may occur at a later date to ensure that the current collection policy is adequately broad enough to cover born digital materials.

Documentation outlining procedures for transferring content from various media formats including optical media, specific hard disk drive interfaces such as SCSI and flash card media have not been created at this time. Other procedural documents detailing the transfer of data from particular media types such as 3.5" floppy disks and IDE/SATA hard disk drives have been created. Although the details vary based on source media type, most of the procedures are very similar. The procedural documentation produced at this time serves as an exemplar and template for future transfer procedure documentation.

Documentation Created

This section contains high-level information about the documentation created for this project. Documentation is divided into two sections: pre-accessioning level and accessioning level. In the pre-accession section, the documentation relates to actions that may occur when a donor and collecting entity display interest in acquiring or donating a collection of materials. The documentation situated in this section provide information to donors, guide Archivists and other collection managers when appraising born digital collections for acquisition, and outline activities that the Preservation Unit may engage in to assist with collection appraisal. The accessioning level relates to policies, procedures, guidelines and other documentation intended to guide Library and Archives Units through born digital collections acquisition activities. Documents are briefly described within this section with a link provided to the entire document found in Appendix B: Document Manual.

Pre-Accessioning Level Documentation

INFORMATION: Disclosure of Use of Digital Forensics Technology

See Appendix B: **INFORMATION: Disclosure of Use of Digital Forensics Technology**

About:

Digital forensics, which evolved within the law enforcement and legal fields, is a branch of forensic science encompassing the capture and analysis of contents stored on digital materials. It employs a number of techniques and tools that address cultural heritage concerns of maintaining provenance, and ensuring trusted, reliable, and verifiable data. The Born Digital Content Preservation service uses some digital forensics techniques in its data migration activities.

With the use of these techniques, early adopters within the cultural heritage field assert that disclosure of the use of these techniques and notifying the donor of what may be revealed when, for example, a forensic disk image is made available to researchers, is key for maintaining trust between the institution and donor. (Kirshenbaum et al, 2010, p. 51) Kirshenbaum et al emphatically make the point that the implications of using such techniques should be carefully monitored and made transparent:

> *"An understanding of these ethical concerns and established mechanisms by which to address them must be embedded in the organization's professional culture, in its employment policies and staff handbooks, in the managerial regimes that monitor the behavior of staff, and in the transparency and public availability of a repository's policies governing the transfer, preservation, management, and delivery of potentially sensitive materials." (Kirshenbaum et al, 2010, p. 51)*

In following with this ethical obligation, this document is meant to inform that the Preservation Unit, acting on behalf of Library and Archives collecting units, employs digital forensics techniques. This documentation is written at a high-level; details on the specific methods used and exact information uncovered is not provided, although there is an invitation for those concerned to contact the Digital Preservation Coordinator regarding methods used. The document also notes that digital forensics techniques are used in a different manner than typical in the law enforcement context – that is, the Digital Preservation Coordinator is not looking for evidence to prove a legal case – and that use is primarily to retain provenance information and to provide trusted digital objects to researchers.

Future Work:

This document could be revised to provide more detail regarding guidance on handling sensitive data, redaction methods and other security and access control issues. Kirshenbaum et al state that," "Privacy is at the heart of the ethical issues surrounding born digital materials." (Kirshenbaum et al, 2010, p. 53) In its current iteration, this document is meant to provide some level of transparency regarding internal procedures, and provide an avenue for donors and others to learn about the processes if more information is desired.

GUIDELINES: Levels of Preservation and File Format Support

See Appendix B: **GUIDELINES: Levels of Preservation and File Format Support**

About:

These guidelines serve to define base levels of digital preservation within the institution and identify base levels of preferred file formats for general content forms. These initial levels of preservation and preferred file formats are intended to be the starting point from which to develop more robust digital preservation services. Born Digital Content Preservation service

actions are largely contingent upon the services provided by the Library's Medusa digital preservation service which is still in development. The guidelines will be reviewed periodically to ensure they accurately reflect the Library's digital preservation environment.

The document provides a brief introduction on challenges inherent to preserving digital objects. The intention of this document is to provide a basic level of expectation management – establishing the Born Digital Content Preservation service's current level of ability and accepted responsibility. This document is for information purposes; it is not meant to serve as a map for donors to migrate file formats to a preferred file type prior to acquisition. This approach is explicitly discouraged as text is included indicating that data could be lost if and provenance disrupted if files are migrated without care or without knowledge of what is modified. The guidelines are not intended to act as prescriptive procedural manual for acquiring born digital content.

This document is based on preservation level and file format guidelines developed for the Library's institutional repository, the Illinois Digital Environment Access to Learning and Scholarship (IDEALS) (Shreeves, 2008, Shreeves, 2008b and Shreeves, 2009) and guidelines created for the University Archives (Prom, 2012). The IDEALS source document, "IDEALS Preservation Policy" (Shreeves, 2009) and this document use terminology and concepts from the documents, "Trusted Digital Repositories: Attributes and Responsibilities" (RGL, 2002) , "Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects" (OCLC/RLG Working Group on Preservation Metadata, 2012), and the tutorial which coalesces the preceding documents, "Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems" (MIT, 2012).

Future Work:

As more research is completed within the institution on file formats, polices set in place for the Medusa preservation repository, and preservation capabilities increase through further research and practice, these guidelines will likely be superseded by born digital stewardship policies created at the institutional level.

GUIDELINES: Pre-Accessioning guidelines

See Appendix B: **GUIDELINES: Pre-Accessioning guidelines**

About:

Per the AIMS Work Group report on stewarding born digital collections, collection managers should consider the resources and activities required to steward born digital collections at various levels. It is advisable for an archivist or specialist trained in born digital content stewardship to author a feasibility study or technical appraisal on born digital collections identified as of interest to the collecting institution. Although stewardship of these collections is important when making acquisition decisions, the AIMS report advises that, "acquisition of born digital material should be based firstly on a curatorial appraisal of its fit within the collection development policies of the institution". (AIMS Work Group, 2012, p. 7).

The intent of these guidelines is to assist collection managers when considering acquisition of born digital collections. These guidelines are intended to be interpreted as such – guidelines - and are not prescriptive nor take precedence over the collection manager's judgment and expertise in making acquisition decisions. However, the decision to acquire collections of born digital materials does not occur in a vacuum – often, many resources and varying types of expertise are required to steward born digital materials. Thus, these guidelines are intended to help collection managers consider the overall cost and support which may be needed to keep the collection viable in perpetuity. A report or documentation generated by specialists, such as the

Digital Preservation Coordintaor, may also inform donors about how their collections will be handled and preserved and may also be beneficial if a collection manager is seeking financial support to assist in processing a particular collection.

Future Work:

This documentation may be revised once the Digital Collections Advisory Committee establishes authority and develops its own guidelines for collection acquisition. Presently, this Committee is in the development stage and has yet to develop workflow or guidance regarding when a collection manager may consult the committee when making acquisition decisions, or if this is an acceptable committee function.

DOCUMENT: Pre-Accession Preservation Appraisal Report

See Appendix B**: DOCUMENT: Pre-Accession Preservation Appraisal Report**

About:

This report is intended to offer guidance to collection managers about collection preservation concerns prior to acquisitioning a collection. The Digital Preservation Coordinator may be asked to reformat storage media from a collection being considered for acquisition so that the collection manager may access and appraise content stored on inaccessible media. As this report is intended to assist in appraisal decisions it will be written at a high level, outlining some of the long-term preservation concerns specific to the collection based on a brief survey or small sample. Examples of some of the criteria the Digital Preservation Coordinator will consider are included in this document.

GUIDELINES: Transfer of Storage Media and Computer Systems from Offsite to Library

See Appendix B: **GUIDELINES: Transfer of Storage Media and Computer Systems from Offsite to Library**

About:

This document provides basic computer and digital media handling advice, noting some of the issues to be aware of when moving these items from an offsite location to the Library. It is intended to mitigate damage while the computer or storage media is in transit. Collection managers are also encouraged to contact the Digital Preservation Coordinator for transport supplies and consultation.

Accessioning Level Documentation

FORM: Born Digital Content Preservation: Request for Reformatting

See Appendix B: **FORM: Born Digital Content – Request for Reformatting**

About:

This form is intended for use between the Digital Preservation Coordinator and a collection manager, and is used to gather information from collection managers about storage media reformatting projects and their intentions for reformatted content. This iteration draws from a similar form published in the AIMS white paper (AIMS Work Group, 2012) in use at Stanford. This form, however, is used a bit differently in that the primary use is internal and not to survey donors about their collection and its contents.

The form contains the following sections:

- **Request Administration** – contains basic administrative information such as collection manager\requestor contact information, storage media quantity, and reformatting purpose

- **Media Type** – contains list of media types and free-text fields to add information about media types not listed

- **Content** – fields and queries related to the content on the disk intended to aid the Digital Preservation Coordinator in assessing the content after reformatting and to assist in identifying particular types of content

- **Content Rights** – prompts the requestor to consider which rights the rights-holder has granted the institution for preservation, access and other actions which may be covered by copyright. Although the Preservation Unit is not responsible for clearing rights and making rights-based decision, rights questions are often posed to us. This section is intended to document clearance from the collection manager and to prompt collection managers to consider preservation rights issues when working with donors.

- **Reformatted Content Landing** – queries the requestor about storage location of the contents recovered from the storage media. At this point, digital storage locations may vary depending upon the requesting unit's resources. Once the Medusa digital preservation repository is in production, this section will likely change as Medusa will likely be the end storage area.

- **Internal Use** – for the Digital Preservation Coordinator to take notes about the project

Future Work:

This form will likely be modified once work begins with other collection managers. This length of form may be better suited for collection managers to use when gathering contextual information from a donor.

PROCEDURE: Beginning a Media Reformatting Project

See Appendix B: **PROCEDURE: Beginning a Media Reformatting Project**

About:

Primarily an internal document used by the Digital Preservation Coordinator and her assistants, this provides step-by-step procedures for initiating a born digital media reformatting project. It details entering information into the YYYY_ReformattingStats.xls spreadsheet and creating the organizational structure for the reformatting project. The directories created at this stage organize content related to the media group[5]; the lowest organizational level is item level.[6]

---

[5] A media group is defined as a collection of media received by the Digital Preservation Coordinator for purposes of reformatting. Media groups are organized to represent a collection received at a particular point in time and are separated by creator, or subject. If at a later time a media reformatting project is requested that represents an accrual to a collection, a new media group is established. This is again for internal organization and reporting – the

There are several directories created which relate to the item level. These are:

- **ImageCapturedFromDisk** – directory where disk image is placed

- **OriginalFiles** – directory where contents exported directly from the storage medium or disk image are stored

- **MigratedFiles** – directory where copies of files stored in the OriginalFiles directory transformed into a more widely accessible file format. Migrated files stored in this directory have been migrated within the Preservation Unit for access purposes.

- **ObjectDocumentation and ObjectJPG subdirectory** - directory where documentation relating the item level administration is stored. Documentation may include the ObjectJPG directory where a .jpg image of the source media is stored, notes indicating any errors encountered during the reformatting process and other documentation which may provide provenance or auditing information.

The documentation also serves as a data dictionary for the ReformattingStats spreadsheet, and includes notes on which fields in the spreadsheet are required (M indicating mandatory) and which are not (NR indicating not required). This procedural document assists those undertaking reformatting actions by providing step-by-step instruction for organizing reformatted content.

The directory structure created is intended to organize various types of content generated during reformatting including administrative documentation, disk images, un-modified files recovered from the storage media, and files which have been migrated to an accessible file format from an original file format. It is also organized in a manner which is intended to assist collection managers as they work with the reformatted content and re-organize content in preparation for ingest into a repository, for storage on a file server or for access.

---

requesting collection manager does not need to maintain this directory structure; however, she should understand what the media group encompasses in order to better understand reports relating to specific media groups.

[6] Item level represents discreet storage media. Item level directories related to a single disk and maintain a one-to-one relationship.

POLICY: Addressing Computer Viruses, Malware and other Potentially Harmful Content

See Appendix B: **POLICY: Addressing Computer Viruses, Malware and other Potentially Harmful Content**

This policy notes the approach taken by the Digital Preservation Coordinator when reformatting media to mitigate damage to collections by computer viruses. In AIMS whitepaper and other born digital stewardship publications virus scans and removal are important digital collection stabilization tasks. In short, the policy states that collections will be scanned for viruses as part of the transfer. If a virus is found, it will be removed from the reformatted content before moving the content from the quarantined forensic workstation.

POLICY: Disk Image Creation and Retention

See Appendix B: **POLICY: Disk Image Creation and Retention**

About:

This documentation explains in brief about forensically imaging computer disks, the benefits of creating disk images, and practices within the Born Digital Content Preservation service. Content about forensically imaging disks, benefits of doing so, choice of disk image file format, and other concerns (.AFF) were drawn from forensics science (AccessData, 2012), (Common Digital Evidence Working Group, 2006), (Farmer and Venema, 2005), (Garfinkel et al, 2006) and library and information science publications (Gengenbach, 2012), (Kirshenbaum, Overden, Redwine, 2010), (Woods, Lee, and Garfinkel, 2011). It is intended to disclose the Born Digital Content Preservation services' use of forensic disk imaging techniques, as well as to guide collection managers when considering retaining a disk image. Within the documentation is also a caution on deleting disk images – once a collection manager has made the decision to

discard a disk image and indicated this decision in writing, the disk image cannot be recovered.

The source media would need to be imaged once again in order to restore the disk image.

PROCEDURE: Transferring Data from Windows-formatted and Mac OSX- (HFS)

Formatted 3.5" Floppy Disks

See Appendix B: Error! Reference source not found.

About:

Intended primarily for use by the Digital Preservation Coordinator and her assistants this

document is a step-by-step procedural for transferring data from Windows-formatted and Mac

OSX- (HFS) Formatted 3.5" Floppy Disks. Included within the document are the expected

outputs after performing this action, step-by-step instructions and pointers to other documents to

consult once the disk image has been created. Although this is intended to be an internal

procedural document it will be made accessible to other Library Units in order to provide

transparency about transfer processes.

PROCEDURE: Transferring Data from 5.35" Floppy Disk using the Catweasel

See Appendix B: **PROCEDURE: Transferring Data from 5.35" Floppy Disk using the**

Catweasel

See Appendix B: **OUTPUTS FROM THIS PROCEDURE:About:**

Intended primarily for use by the Digital Preservation Coordinator and her assistants this

document is a step-by-step procedural for Transferring Data from 5.25" Floppy Disks using the

Catweasel disk controller card. Included within the document are the expected outputs after

performing this action, step-by-step instructions and pointers to other documents to consult once

the disk image has been created. Although this is intended to be an internal procedural document

it will be made accessible to other Library Units in order to provide transparency about transfer processes.

PROCEDURE: Transferring Data from an Internal Hard Drive: IDE/ATA/PATA/SATA

See Appendix B: **PROCEDURE: Transferring Data from an Internal Hard Drive: IDE/ATA/PATA/SATA**

About:

Intended primarily for use by the Digital Preservation Coordinator and her assistants this document is a step-by-step procedural for Transferring Data from internal hard disk drives with an IDE/ATA/PATA or SATA interface. Included within the document is a list indicating which hardware and software is required to complete this procedure, step-by-step instructions with example images and pointers to other documents to consult once the disk image has been created. Although this is intended to be an internal procedural document it will be made accessible to other Library Units in order to provide transparency about transfer processes.

PROCEDURE: Transferring Data from FireWire/Mac Device

See Appendix B: **PROCEDURE: Transferring Data from FireWire/Mac Device**

About:

Intended primarily for use by the Digital Preservation Coordinator and her assistants this document is a step-by-step procedural for Transferring Data from external hard disk drives with a FireWire interface (400 or 800) or from Macintosh computers with FireWire transfer-capable operating systems. Included within the document is a list indicating which hardware and software is required to complete this procedure, step-by-step instructions with example images, and pointers to other documents to consult once the disk image has been created. Although this is

intended to be an internal procedural document it will be made accessible to other Library Units in order to provide transparency about transfer processes.

PROCEDURE: Create Disk Image Using FTK Imager

See Appendix B: **PROCEDURE: Create Disk Image Using FTK Imager**

About:

Intended primarily for use by the Digital Preservation Coordinator and her assistants this document is a step-by-step procedural for using FTK imager software to create a forensic disk image[7]. Included within the document are the expected outputs after performing this action, fields within the outputs, step-by-step instructions with example images, and pointers to other documents to consult if imaging particular formats. Although this is intended to be an internal procedural document it will be made accessible to other Library Units in order to provide transparency about transfer processes.

PROCEDURE: Extracting Files from Disk Image Using FTK Imager (version 3.1.0.1514)

See Appendix B:

PROCEDURE: Extracting Files from Disk Image Using FTK Imager (version 3.1.0.1514)

About:

Intended primarily for use by the Digital Preservation Coordinator and her assistants this document is a step-by-step procedural for using FTK Imager[8] to extract files from a disk image.

---

[7] A disk image is a "sector- by-sector copy of the data from physical storage media, including modern hard drives, optical disks, USB storage devices, and even portable devices such as iPods, digital cameras, and mobile phones" (Woods, Lee, Garfinkle, 2011, p. 57). A forensic disk image will often include acquisition metadata as a means of recording and retaining provenance information.

[8] FTK Imager is free forensics imaging software created by Access Data and is available here: http://www.accessdata.com/support/product-downloads

Included within the document are the expected outputs after performing this action, fields within the outputs, and pointers to other documents to consult if imaging particular formats. Although this is intended to be an internal procedural document it will be made accessible to other Library Units in order to provide transparency about transfer processes.

PROCEDURE: Receiving Vendor Reformatted Content

See Appendix B: **PROCEDURE: Receiving Vendor Reformatted Content**

About:

Intended primarily for use by the Digital Preservation Coordinator and her assistants this procedural document details how to incorporate vendor-reformatted content into the reformatting workflow to produce consistent outputs. Included within the document are the expected outputs, a list of hardware used in this procedure, notes to record vendor cost, and recording other vendor-related items. Although this is intended to be an internal procedural document it will be made accessible to other Library Units in order to provide transparency about transfer processes.

PROCEDURE: Generating File Metadata and Reports Using TreeSize Pro (version 5.5.5.816)

See Appendix B: **PROCEDURE: Generating File Metadata and Reports Using TreeSize Pro (version 5.5.5.816)**

About:

Intended primarily for use by the Digital Preservation Coordinator and her assistants this procedural document details how to use TreeSize Pro to export reports about reformatted items or media groups. Included in this report are expected outputs from this procedure. Although this is intended to be an internal procedural document it will be made accessible to other Library Units in order to provide transparency about transfer processes.

INFORMATION: Mapping Preservation Output to Archives to Medusa SIP

See Appendix B: INFORMATION: Mapping Preservation Output to Archives to Medusa SIP

About:

This diagram serves as an example of how the reformatted content directory structure that is generated through born digital content reformatting can be mapped to a.) the current UA working directory structure and b.) mapped to a proposed UA Submission Information Package, or SIP, which will be ingested into the Library's digital preservation repository, Medusa.

One of the goals in developing the Medusa archive is to comply with the ISO standard (ISO 14721:200) OAIS Reference model, defined by recommendation CCSDS 650.0-B-1 of the Consultative Committee for Space Data Systems (CCSDS, 2012). The Reference model diagram is a high-level conceptualization of a digital preservation system, and highlights the functions of an OAIS-compliant digital preservation system. The term Submission Information Package, abbreviated as SIP, is drawn from the OAIS Reference model. The SIP is an information package submitted to an OAIS system by a producer used to construct or update the Archival Information Package within the OAIS – a producer being one who provides information to be preserved. The information package may contain Content Information and associated Preservation Description Information. Within the package is Packaging Information which functions to identify and delimit the Content Information and Package Description information which serves to, "facilitate searches for the Content Information" (CCSDS, 2012).

The diagram represents mapping output supplied by the Born Digital Content Preservation service to the UAs' current working area – a directory within the Unprocessed directory stored on one of the Library's file servers. Currently, the Unprocessed directory serves as a working space, storage and delivery area for the UAs' electronic records. The diagram also

represents how the UA may choose to map the directories and contents submitted by the Born Digital Content Preservation service to an example SIP. The diagram thus represents the transfer of content extracted from physical storage media, which is then uploaded to the UA's working area for processing. Once processed, an Archives' employee would then map relevant information to the directory structure and\or content model established for the SIP and submit the SIP to Medusa. Medusa would verify the SIP based on the established content model, which has not yet been created.

The diagram also suggests directories within the SIP where content produced by the Born Digital Content Preservation Service may be transferred. For example, the ImageCapturedFromDisk and OriginalFiles directories may map to the Preservation directory within the UAs' SIP as these directories contain the least modified representations extracted from source storage object. Documentation generated through reformatting actions such as reports, Representation Information[9] and a source media disk image may map to the Documentation directory within the SIP. Files which have been modified or migrated prior to ingest into Medusa may map to one of the subdirectories within the Processed directory in the SIP.

This diagram is meant to serve as an example of how the UA may map to a SIP. It is not intended to be prescriptive, nor is there a requirement that the UA retain all documentation

---

[9] Representation Information is another digital preservation concept from the Open Archival Information System and is defined as, "any information required to understand and render both the digital material and the associated metadata. Digital objects are stored as bitstreams, which are not understandable to a human being without further data to interpret them. Representation information is the extra structural or semantic information, which converts raw data into something more meaningful." (Digital Curation Centre, n.d.)

produced by the Born Digital Content Preservation service. A group of documentation that may not be retained is the reports generated relating to particular media groups. Media group level reports generated by TreeSize Pro may not be relevant once the discrete directories within the media group are reorganized by the UAs' staff. These reports are to assist UA staff in analyzing the reformatted content.

Once the Preservation Unit-organized media groups are sent to the UA contents may be rearranged to conform to the arrangement of other electronic contents within a particular collection, thus breaking apart a media group. As an example, a collection may contain a number of electronic files, some which are immediately accessible to the UA and which do not require the services of the Born Digital Content Preservation service, and other content which is stored on inaccessible media, which requires assistance from the Born Digital Content Preservation service. A UA staff member may create an organizational scheme for the files which do not need reformatting, and may incorporate content restored from physical media once it has been migrated to the UA's networked storage. As the Born Digital Content Preservation service staff only knows about the physical media which the UA submitted for reformatting, reports are limited to that domain.

GUIDELINES: Retention of Original Storage Media

See Appendix B: **GUIDELINES: Retention of Original Storage Media**

About:

After recovering the informational content from storage media a decision must be made about retaining the carrier itself. These guidelines are intended to assist collection managers in making the decision to retain or dispose of original media carriers. Reasons are provided for considering retaining the source item, including a list developed by the National Archives and

Records Administration Committee on Intrinsic Value. The guidelines also prompt collection managers to consider care if the original storage format is retained, and how to securely dispose of physical media that are identified for deaccessioning.

Future Work:

This project was not undertaken with a "once and done" mindset. As noted in several publications (AIMS Work Group, 2012), (Kirshenbaum, Ovenden, and Redwine, 2010), (Gengenbach, 2012) there is still work to be done on a community level in developing best practice for processing born digital collections. With work presently being undertaken in this area, acknowledgement that there are not yet established best practices and an evolving born digital processing environment within the University Library, change is indeed inherent in the process. In describing the documentation produced, I have also indicated some areas where I anticipate future work. Indeed, it is expected that all documentation will be reviewed on a regular basis to ensure that 1.) it is still relevant and reflects our working environment and practices 2.) it aligns with community practices, particularly if a strong statement for particular best practice has evolved.

Part of gauging the documentation's efficacy is implementing and using it. Some of the documentation created is born from empirical research and daily practice; others have been developed based on recommended good practice as indicated by the AIMS framework and other publications. The next stage for this documentation is to implement it in a production environment and measure its efficacy in assisting with born digital stewardship.

To maintain transparency and engender trust between Archives and Library faculty and staff, donors and researchers, documentation will be made available via a wiki. Publication of the

documentation in professional contexts, and continued development is also intended to contribute to the development of the born digital stewardship community of practice.

Other future plans are to broaden the development and implementation of workflows to other units within the Library who are working with born digital materials. A working group will be established to aid in the development of documentation, particularly in those areas of processing not addressed or deemed out of scope for this paper.

Policies, procedures and workflows will also likely change once the Medusa Digital Preservation repository is in testing and production mode. Migration, emulation, metadata extraction and other content and file-related issues will be addressed and expanded including how to manage content stored in problematic file formats such as proprietary formats, or files which do not contain file extensions and are not easily detected using file identification tools such as FITS or JHOVE. Identifying where in the workflow these problems are addressed and those responsible for problem solving will again cause a shift and require documentation updates.

Conclusion:

This project is representative of the first stage of procedures, policies and workflows. As indicated several times throughout the paper, this iteration of documentation produced is intended to be the ground floor for institutional practice in stewarding born digital content at the University of Illinois' Library and Archives. Based on nascent frameworks and developing practice, this is the starting point in systematically addressing and working in coordinated manner toward the stewardship of born digital content within the institution.

The process will be necessarily iterative – with the documentation having a lifecycle of its own including stages of creation, review, and revision as it is published and implemented. Documentation may need to be created and modified as stewarding other types of content, such as born digital research data and storage media within general collections, are systematically addressed and processes scaled to other Library units. Evolving practice and the development of new knowledge and capabilities within the Born Digital Content Preservation service will also necessitate change, as will development of good- and best-practice within the born digital stewardship community at large.

References

About | BitCurator. (n.d.) Retrieved from http://www.bitcurator.net/aboutbc/

AccessData. (2012). *FTK Imager: User guide.* Retrieved from

http://marketing.accessdata.com/acton/attachment/4390/f-000d/0/-/-/-/-/file.pdf .

AIMS Work Group. (2012). *AIMS born-digital collections: An Inter-institutional model for*

stewardship. Retrieved from

http://www2.lib.virginia.edu/aims/whitepaper/AIMS_final.pdf.\

CITES Security. (n.d.) *CITES – Disk Scrubbing.* Retrieved from

https://security.illinois.edu/content/disk-scrubbing

Common Digital Evidence Storage Format Working Group. *Survey of disk image storage*

formats: Version 1.0.(September, 2006). Retrieved from

http://www.dfrws.org/CDESF/survey-dfrws-cdesf-diskimg-01.pdf

Consultative Committee for Space Data Systems (CCSDS). (2012). *Reference model for an open*

archival information system (OAIS): Recommended practice CCSDS 650.0-M-2.

Retrieved from http://public.ccsds.org/publications/archive/650x0m2.pdf

Digital Curation Centre.(n.d). *What is representation information?* Retrieved from

http://www.dcc.ac.uk/node/9558#sthash.5x7RHw4t.dpuf

Digital Preservation Coalition. (2008). *Preservation management of digital materials: The*

Handbook. Retrieved from http://www.dpconline.org/advice/preservationhandbook

Duranti, L. (1995). Reliability and Authenticity: The Concepts and Their Implications.

Archivaria, *39*, 5–10. Retrieved from

https://courses.ischool.utexas.edu/megan/MeganCollection2/files/13353/Duranti_Reliabil

ity_and_Authenticity.pdf

Farmer, D., & Weiste, V. (2005). *Forensic Discovery.* Boston, MA: Addison-Wesley

  Professional. Retrieved from http://www.porcupine.org/forensics/forensic-discovery/

File: PATA Cable.jpg – Wikipedia, the Free Encyclopedia. (2010, September 10). Retrieved

  from http://en.wikipedia.org/wiki/File:PATA-cable.jpg

File: SATA Cable.jpg – Wikipedia, the Free Encyclopedia. (2005, May 9). Retrieved from

  http://en.wikipedia.org/wiki/File:SATA_Data_Cable.jpg

Forstrom, M., Kuhl, N., Thomas, S., John, J. L., Barnard, M., Redwine, G., Shaw, S. (2013,

  January). Born Digital: Guidance for Donors, Dealers, and Archival Repositories. *Born*

  Digital | Media Commons Press. Retrieved from

  http://mediacommons.futureofthebook.org/mcpress/borndigital/

FTK Imager. – Forensics Wiki. (2010, December). Retrieved from

  http://www.forensicswiki.org/wiki/FTK_Imager

Garfinkel, S. L., Malan, D. J., Dubec, K., Stevens, C.C., & Pham, C (2006, January-February).

  Advanced forensic format: An Open, extensible format for disk imaging. *Advances in*

  Digital Forensics II: FIP International Conference on Digital Forensics, 17-31. New

  York: Springer. Retrieved from:

  http://dash.harvard.edu/bitstream/handle/1/2829932/Malan_AdvancedForensic.pdf?seque

  nce=4

Gengenbach, M. (2012). The way we do it here: Mapping digital forensics workflows in

  collecting institutions. (master's thesis). University of North Carolina – Chapel Hill,

  Chapel Hill, NC. Retrieved from

  http://digitalcurationexchange.org/system/files/gengenbach-forensic-workflows-2012.pdf

Gueguen, G. (2013, April 16). *Management of Born-Digital Special Collections*. Presented via

ASERL Webinar: "Intro to Digital Preservation #3. Retrieved from

http://www.aserl.org/wp-content/uploads/2013/04/Intro_DP_2013-3_Born_Dig_Docs.pdf

Heslop, H., Davis, S., & Wilson, A. (2002, December). *An Approach to the Preservation of Digital Records.* Retrieved from http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm16-47161.pdf

Hunter, G. (2003). *Developing and maintaining practical archives: A how-to-do-it manual: Number 22.* New York, NY: Neal-Schuman Publishers, Inc.

IDE vs SATA Hard Drives. (2010, February 5). Retrieved from

http://blog.shophde.com/2010/02/05/ide-vs-sata-hard-drives/

Individual Computers Catweasel. (n.d.) Retrieved from

http://en.wikipedia.org/wiki/Individual_Computers_Catweasel

Intrinsic value in archival material. (1999). Staff Information Paper Number 21. Washington, D.C., National Archives and Records Administration. Retrieved from http://www.archives.gov/research/alic/reference/archives-resources/archival-material intrinsic-value.html

Kozierok, C. (2001). *PC guide – Ref – Hard disk head sliders, arms and actuator*.

Retrieved from http://www.pcguide.com/ref/hdd/op/act_Parking.htm

Jordan, A. (n.d.). *Transferring data from 3.5" floppies.* Retrieved from

http://archives.library.illinois.edu/staff/documents/digital/1.5.pdf

Kirschenbaum, M. G., Ovenden, R., Redwine, G., & Donahue, R. (2010). *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington, D.C.:Council on Library and Information Resources. Retrieved from http://mith.umd.edu/wp-content/uploads/whitepaper_borndigital.pdf

Pearce-Moses, R. (2005). *A Glossary of Archival and Records Terminology – Provenance.*

Retrieved from http://www2.archivists.org/glossary/terms/p/provenance

Popp, T. (2011). *Floppy disk migration pilot project: Phase one report.* Retrieved from:

http://www.library.illinois.edu/prescons/pdfs/Floppy_Disk_Migration_Pilot_Project.pdf

Preservation and Conservation – Mission statement. (2013). Retrieved from

http://www.library.illinois.edu/prescons/

Product Downloads | Access Data – FTK Imager version 3.1.3. (2013, June 13). Retrieved from

http://www.accessdata.com/support/product-downloads

Prom, C. (2012a). *Digital checklist.* Retrieved from:

http://archives.library.illinois.edu/staff/documents/digital/digitalchecklist.docx

Prom, C. (2012b). *Digital workflow.* Retrieved from: http://archives.library.illinois.edu/staff-

resources/digital- workflow/

Prom, C. (2012c). *Generating Preservation Metadata.* Retrieved from

http://archives.library.illinois.edu/staff-resources/digital-workflow/generating-

preservation-metadata/

Prom, C. (2012d). *Transfer electronic/digital records.* Retrieved from

http://archives.library.illinois.edu/deposit-records/transfer-electronic-records/

Research Libraries Group. (2002). *Trusted Digital Repositories: Attributes and Responsibilities:*

An RLG-OCLC Report. Retrieved from

http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf

Richard, F. (2010). *Illustrate workflow process with a swim lane diagram.*

Retrieved from http://www.brighthubpm.com/monitoring-projects/85265-illustrate-

workflow-process-with-a-swim-lane-diagram/

Shreeves, S. (2009, November). *IDEALS Digital Preservation Policy.* Retrieved from

https://services.ideals.illinois.edu/wiki/bin/view/IDEALS/IDEALSDigitalPreservationPol

icy

Shreeves, S. (2008). *File Format Recommendations.* Costello, K. (Ed.). Retrieved from

https://services.ideals.illinois.edu/wiki/bin/view/IDEALS/FormatRecommendations

Tessella Digital Preservation. (n.d.) *Media Obsolescence.* Retrieved from

http://www.digitalpreservation.com/challenges/media-obsolescence/

Woods, K*,* Lee, C.A.*, &* Garfinkel, S. *(*2011*). Extending digital repository architectures to*

support disk image preservation and access. In Proceedings of the 11th Annual

International ACM/IEEE Joint Conference on Digital Libraries (pp. *57–66*). New York,

NY: Association for Computing Machinery. Retrieved from

http://delivery.acm.org/10.1145/2000000/1998088/p57-

woods.pdf?ip=130.126.148.111&acc=ACTIVE%20SERVICE&CFID=183273555&CFT

OKEN=57850102&__acm__=1361743179_baeef023ac036414432420121ef30ea6

Zlib home site. (2013, April). Retrieved from http://www.zlib.net/

Appendix A: University Archives Documents

Below are the workflows for processing digital records the University Archives has in place as of January 23, 2013.

Digital Checklist

This text was retrieved from

http://archives.library.illinois.edu/staff/documents/digital/digitalchecklist.docx (Prom, 2012a)

The formatting from the source document may have been modified to better fit the formatting of this paper.

University Archives Electronic Records Management Form

Collection Title: _____

Record Series Number: _____

Accessioning Tasks

Accession date: _____

Accessioned by: _____

Actions Taken	**Check**	**Date**
virus check		
disk contents transferred to Unprocessed folder		
weeded for personal information/non-archival materials		
preservation metadata created and results placed in Preservation folder; material copied to Preservation folder		
disk contents copied to Nearline folder and compressed		

disk contents copied to Online folder (if applicable)		
Archon collection record and digital record completed		
print form and transfer digital copy to Preservation folder		

Final Drive Location (ex. Libsysdigi\UA\Preservation):

Comments: please detail digital management aspects outside of the scope of the checklist (i.e. sent e-records to DCC, particular digital formats could not migrate, etc.)

Digital Workflow

This text was retrieved from http://archives.library.illinois.edu/staff-resources/digital-workflow/ (Prom, 2012b)

The formatting from the source document may have been modified to better fit the formatting of this paper.

This page outlines the general processing, preservation, and access workflow for electronic records. In order to process born digital and digital surrogate records in accordance with archival standards, process materials using the below steps.

Be sure to record all steps on the Digital Checklist. Keep this form in the unprocessed folder until processing is complete, save as a pdf, then transfer it to the preservation folder. Also, print a copy of this completed form and put it in the record series' administrative file. *The links and menu to the left provide detailed information on each process.*

Note: all newly created e-records series must have folders created by either Chris or Angela. Before processing, have one of these individuals create the folders required for your processing project.

1. **Storage for Unprocessed Electronic Records** All born digital and digital surrogate records should be transferred unprocessed to \\lib-archives-fs.ad.uillinois.edu\UniversityArchives\archives\unprocessed under the newly-created record series folder. If the computer operating system does not automatically open Teracopy upon transfer, right click on the open program (Teracopy), select "Use TeraCopy as default copy handler" and "Always test after copy." Electronic records acquired on CDs or DVDs must be ripped using the specified programs and placed in the same unprocessed records folder.

2. **Archon Collection Record** If the series is comprised of hybrid records (i.e. analog and digital), generate a collection record for the analog records. A digital collection record will be generated at a later point in the workflow process. If the series is entirely digital and is not associated with an analog series, create a collection record for the series. The record series title should be "Electronic Records of…" to indicate born digital records, or "Digital Surrogates of…" for electronic copies of original analog records. Most of the descriptive information for stand-alone electronic records series will be generated later in a digital collection record.

3. **Expunging Personal Information and Non-Archival Records** Before generating preservation metadata, make sure the record series contains no personal information. Private records often found in donated materials include social security numbers, banking and credit card information, student records, and medical records. For larger electronic records ingests, Firefly can quickly scan and identify files that contain social security and bank numbers. Once the program is installed, select "Configure," and search for the correct scan location. Delete any identified files that contain personal information. Save

the results in the Preservation folder. In the case of complex records, run TreeSize Pro to gain a better sense of file structures and type. Delete any material not worth archiving. Save the results of TreeSize Pro in the Preservation folder.

4. **Generating Preservation Metadata** Process all electronic records using NARA's File Analyzer, found in [Preservation Metadata](). Upon successful creation of preservation metadata, transfer the files from the Unprocessed folder to the Preservation folder. The Unprocessed record series folder should be deleted. Send Angela or Chris an email with the record series number to be deleted.

5. **Generating Access Copies** Depending upon restrictions of copyright and donor request, electronic records may be restricted, available upon request, or publicly available. All documentation for these processes is located on the [Public Access]() page. At a minimum, make a nearline access copy, which will be a complete or close-to-complete copy of the received files, minus only those files that have access restrictions. In some cases, online access copies will consist of content free of copyright or privacy concerns, arranged and described in a way that is useful to users. Before materials are made publicly accessible, Bill, Chris, or Ellen must review the records placed online. Upon permission, update the Archon collection record to signify permission was granted. Under the appropriate record series, select "Other" and make a note of the date permission was granted for public accessibility, and the archivist granting permission.

Providing public access to digital records requires greater attention to copyright and privacy restrictions than when working with analog records. If you have any questions or concerns, see Bill or Chris!

Transferring Data from 3.5" Floppies

This text was retrieved from http://archives.library.illinois.edu/staff/documents/digital/1.5.pdf

(Jordan, n.d)

The formatting from the source document may have been modified to better fit the formatting of this paper.

Digital Archives Program

Procedure 1.5: Transferring Data from 3.5" Floppies

Author: Angela Jordan

Transferring Data from 3.5" Floppies

1. PC floppies can be ripped using the external floppy drive located in the University Archives' corner staff desk (reading room) and placed in the Unprocessed Records Folder under the appropriate record series (Libsysdigi->Archives->Electronic Records Repository->UA->Unprocessed E-Records)\

2. Once files are ripped, run these files through the NARA File Analyzer (C Drive>Program Files>usnationalarchives-File-Analyzer-74e0b69>bin>fileAnalyzer.jar). If necessary, create a shortcut on the desktop for future use. See the Creating Preservation Metadata tab on the website and document 2.0 for instructions on how to use the NARA File Analyzer.

3. Make sure to label each physical CD and/or DVD so they correspond to the files located in the Electronic Records Repository.

To migrate information from 5.25″ floppies contact Tracy Popp, the Digital Preservation Coordinator, at tpopp2@illinois.edu for assistance.

Note: We currently have no way to transfer data from Mac floppies and other storage devices such as zip drives and cartridges at the Archives. Contact Tracy Popp, the Digital Preservation Coordinator, at tpopp2@illinois.edu for assistance. If Preservation cannot help, keep these storage devices with the record series and make a note in the finding aid.

Generating Preservation Metadata

This text was retrieved from http://archives.library.illinois.edu/staff-resources/digital-workflow/generating-preservation-metadata/ (Prom, 2012c)

The formatting from the source document may have been modified to better fit the formatting of this paper.

1. Using Teracopy, move the folders/files from the Unprocessed folder to a the Preservation folder.

2. Open NARA File Analyzer (C drive\Program Files\usnationalarchives-File-Analyzer-74e0b69\bin\fileAnalyzer.jar). If necessary, create a shortcut on the desktop for future use.

3. Provide the following information on the opening screen:

    o Root Directory: find the appropriate record series in the Electronic Records Repository (\\lib-archives-fs.ad.uillinois.edu\UniversityArchives\archives\xxxxxxx\Preservation)

    o Action to perform: Get MD5 Checksum By Name

    o Max Items: depends upon the size of the ingest

    o Filter Criteria: leave blank unless you want to scan particular folder items

4. Hit "Analyze" at the bottom of the screen.

5. Once an Analysis is complete, the results will open in a new tab to the left of the existing tabs.

6. Open the "Details" Tab, which provides a file-by-file description of the processed files.

7. Export the information from the details tab to the record series' Preservation folder located in the Electronic Records Repository, saving it as "xxyyzzz_metadata.txt" (example: "1505050_metadata.txt" for record series 15/5/50).

8. Copy the files from the unprocessed folder to the preservation folder. Once copied, delete the files being from the unprocessed folder.

For multiple ingests of the same record series:

Use these instructions for a record series that already has preservation metadata and additional material is being added to the series. The user can also use the below steps to filter look only at matching items, mismatched items or missing items.

1. Run the file analyzer on the new material, as per steps 1-5 above.

2. In Step 6, save the results as xxyyzzz_metadata2.txt and rename existing metadata.txt file to metadata1.txt

3. Open both metadata.txt and metadata2.txt with a text editor, such as NoteTab.

4. Cut and Paste the results from both metadata.txt and metadata2.txt into a new file. Save the file, using the filename "xxyyzzz_metadata.txt."

5. Retain all three metadata files in the directory, to allow for redundancy if there is a problem at a future time.

Appendix B: Document Manual

INFORMATION: Disclosure of Use of Digital Forensics Technology

Digital forensics, a branch of forensic science encompassing the capture and analysis of contents stored on digital materials, employs a number of techniques and tools beneficial to digital preservation; tools that help retain metadata such as creation dates, and avoid changing or writing back to source media. Cultural heritage institutions have begun adopting some digital forensics tools and methods as they complement the activities important to stewarding born digital collections including preservation, provenance, and discovery.

Digital forensics tools can uncover a variety of information stored on media including deleted files, sensitive and personal information such as bank and credit card numbers, internet search information and downloaded information. This ability to uncover files which may not have been intended by a creator to be seen by others or available to researchers implicates those in cultural heritage institutions who use the tools as ethically obligated to disclose use of this technology. In compliance with ethical disclosure this statement serves to disclose that The Born Digital Content Preservation service and third party vendors may use digital forensics techniques such as forensic imaging and software to view and extract disk image contents. Collection managers are advised to notify donors of born digital content of our use of these technologies.

Although we are using this technology it is not our intention to provide access to potentially sensitive materials or investigate contents in the same manner as law enforcement; the tools are used primarily to retain provenance and data integrity and to prevent access to donors' potentially sensitive information. However, in the course of processing born digital contents archivists and other technical staff within the Preservation Unit will have access and be able to view all of the contents recovered from the media - this may include sensitive or very

personal information. Although professional staff will take care to redact personal information such as bank and credit card information, due to the volume of content often stored on digital media it may not all be found and maybe inadvertently disclosed to researchers through access.

Prior to accessioning materials, curators should fully disclose our use of this technology to donors and should address access rights and actions relating to:

- Who is responsible for locating and redacting sensitive information?

- How digital contents will be accessed – is it acceptable to allow researchers to interact with a forensic disk image? Is online access allowable? What files are considered within scope of access?

- Does the donor wish to have a technical specialist from the Library review the digital content prior to acquisition?
  Note: this has the potential to be quite costly and time-consuming. A collection manager and donor should carefully consider accessioning digital content if there is excess worry about disclosure of sensitive or personal information such as bank and credit card numbers, personal contacts, pornography downloaded to the computer, or illicit content.

It is ultimately the collection manager's responsibility to work with the donor on these issues. Particular concerns or any agreements on file retention, or file return should be documented. The Digital Preservation Coordinator should be made aware as soon as possible if there are any particular concerns about accessing content from born digital storage devices.

If you would like details about the specific technologies employed in the Born Digital Content Service please contact the Digital Preservation Coordinator (tpopp2@illinois.edu).

GUIDELINES: Levels of Preservation and File Format Support

Introduction:

The Born Digital Content Preservation service is committed to the overarching mission of the Preservation Unit to, "assure long-term access to the physical and intellectual contents of the Library's collections."[10]

As with most materials, there are challenges inherent to preserving digital objects. Unlike traditional analog objects such as books which are eye-legible and do not require additional tools to decode their contents, digital objects are dependent upon computer software and hardware in order to render the intellectual content. This dependency situates the digital object on very tenuous ground; software and hardware computing environments continually shift as technology evolves and are prone to rapid obsolescence. As software and hardware become obsolete access to the intellectual contents of objects dependent upon these environments is threatened; the digital objects then follow in obsolescence. Relative to their function, digital objects can also be complex and dynamic, attributes which contribute to the challenge of preserving the object. These challenges are often contingent upon, but certainly not limited to, the digital object's file format.

In acknowledgement of the inherent challenges, this policy serves to define levels within the digital preservation service and identify preferred file formats for general content forms in order to develop more robust levels of preservation in the future. Born digital content Preservation service actions are largely contingent upon the services provided by the Library's Medusa digital preservation service which is still in development. This policy will be reviewed periodically to ensure it accurately reflects the Library's digital preservation environment.

---

[10] http://www.library.illinois.edu/prescons/

Levels of Preservation:

The Born Digital Content Preservation service's ability to provide long-term digital preservation is linked to the ongoing development of the Library's Medusa digital preservation service; at present, Medusa offers bit-level preservation. As the Medusa digital preservation service matures and evolves so too will the related policies developed for born digital content preservation.

Bit-level preservation actions maintain the viability of the object, ensuring that the bits comprising the digital object remain unaltered and readable from the storage media by a computer system. These basic actions are intended to maintain the fixity and authenticity of digital objects, but do not cover actions required to make the file functional or human-understandable.

Born Digital Content Preservation service actions are intended to complement and coalesce with Medusa bit-level preservation actions. During born digital media reformatting the following actions are performed:

- **Regular virus and malware checks**

- **Creation of checksums**

- **Creation of preservation metadata**

- **Secure transfer between intermediate storage media**

Bit-level preservation within the Medusa preservation environment includes the following actions:

- **Provision of persistent identifier for object and/or its metadata**

- **Creation of preservation metadata**

- **Secure storage and backup**

- **Regular fixity checks**

- **Regular virus checks**

- **Storage of original digital object**

- **Ability to extraction of digital object from storage system with original directory structure and file names intact.**

File Format Support:

Although the Born Digital Content Preservation service will extract all relevant content from media, and collecting units within the Library will accept a wide variety of formats, some file formats are preferred for their potential to remain viable, renderable, understandable and functionable in the future. These preferred file formats are identified in part by to the following:

- **Openly documented and/or non-proprietary**

- **Supported by a wide range of software**

- **Widely adopted; has a large user base**

- **If compressed, has been compressed using a lossless compression method**

Digital object complexity also affects preservation outcome. Files that contain embedded, dynamic or highly-customized content can be more difficult to preserve beyond basic bit-level. The file formats listed below are preferred; however, if donor files are not in this format there is no expectation for them to convert files before acquisition. In fact, it is preferable that **donors DO NOT alter files** in attempt to meet guidelines; file migration can result in data loss and change of important metadata information.

The following file formats have been identified as preferable by the University Archive[11]'s and

IDEALS[12] due to increased potential for long-term preservation and access.

If your files types are not listed

please DO NOT convert them to a preferred format as this action may adversely affect file

quality and metadata information. Contact an archivist or the Digital Preservation

Coordinator if you have questions.

Textual Formats (including Office/Presentation Formats):	
Highest Preference:	
Name	*File Extension*
Comma-Separated values	.csv
Plain Text	.txt
XML	.xml
XHTML	.html
OpenOffice.org formats	see http://www.openoffice.org/
Microsoft Office open XML format	.docx
Moderate Preference:	
Name	*File Extension*
Rich Text	.rtf
HTML	.htm, .html

---

[11] http://archives.library.illinois.edu/deposit-records/transfer-electronic-records/
[12] https://services.ideals.illinois.edu/wiki/bin/view/IDEALS/FormatRecommendations

Microsoft Word	.doc
Microsoft Excel	.xls
Microsoft PowerPoint	.ppt
Portable Document Format	.pdf
Postscript	.ps
Image Formats:	
Highest Preference:	
Name	*File Extension*
Tagged Image File Format	.tif, .tiff
JPEG	.jpg
Moderate Preference:	
Name	*File Extension*
Graphics Interchange Format	.gif
Portable Network Graphics	.png
Portable Document Format	.pdf
JPEG 2000	.jp2
Bitmap (for Web graphics)	.bmp
Audio Formats:	
Highest Preference:	
Name	*File Extension*
Waveform Audio File Format	.wav
Audio Interchange File Format	.aiff

Free Lossless Audio Codec	.flac
Ogg (container format)	.ogg
Moderate Preference:	
Name	*File Extension*
Advanced Audio Coding	.aac
MP3	.mp3
Real Audio (for streaming media)	.ram
Windows Media Audio	.wma
Video Formats:	
Highest Preference:	
Name	*File Extension*
Audio Video Interleaved	.avi
MPEG-4 Video	.mp4
Windows Media Video	.wmv
Quicktime	.mov
Moderate Preference:	
Name	*File Extension*
MPEG-2 Video	.mp2
Data Files and Databases: (Convert tables to files with fixed-length fields or fields defined by delimiters)	
Acceptable:	
Name	*File Extension*

data spreadsheets files	XLS, XSLS, ODX, etc
database formats	DB, FP7, MDF, MDB
Email: (Include and retain email attachments)	
Acceptable:	
Name	**File Extension**
E-mail	.eml
Mbox	.mboxo, .mboxrd, .mboxcl, and .mboxcl2
XML	.xml

DOCUMENT: Pre-Accession Preservation Appraisal Report

This report is intended to offer guidance to collection managers about collection preservation concerns prior to acquisition. The Digital Preservation Coordinator may be asked to reformat storage media from a collection being considered for acquisition so that the collection manager may access and appraise content stored on inaccessible digital storage media. As this is intended to be a report to assist in appraisal decisions it will be written at a high level, outlining some of the long-term preservation concerns specific within the collection based on a brief survey or small sample.

When drafting this document the Digital Preservation Coordinator will analyze the content from a long-term digital preservation point of view. She will not analyze the intellectual content of the files. She will appraise the content based on the following factors:

1. Identifying long-term preservation challenges including:

    a. Digital storage space required

    b. Potential costs incurred in maintain the content

    c. Impediments to accessing the content. This may include:

        i.   Content stored on media that is not accessible by the Born Digital Media Reformatting lab

        ii. Content stored on media that is damaged or appears to be in poor physical condition

        iii. Content locked in proprietary or unusual file formats (file types will weigh heavily in the assessment; see item #2.)

    d. Curation actions necessary to maintain the content – what will the collection manager need to do before content is suitable for ingest into the digital preservation repository?

    e. Intellectual property rights concerns – a high-level analysis of potential rights concerns

2. File types – an analysis of the types of files found within the acquisition:

    a. Is the content stored in proprietary file formats which may be difficult to access?

    b. Are there many files in unusual formats?

    c. Are there many files without easily identifiable file extensions?

    d. What is the average age of the file?

    e. What steps might be necessary to access the content? This may include:

        i. Purchasing software

        ii. Migrating content from its original file format to one deemed acceptable within our repository for long-term preservation

GUIDELINES: Pre-Accessioning guidelines

Per the AIMS Work Group report (AIMS Work Group, 2012) on stewarding born digital collections, collection managers should consider the resources and activities required to steward born digital collections at various levels. It is advisable for an archivist or specialist trained in born digital content stewardship to author a feasibility study or technical appraisal on born digital collections identified as of interest to the collecting institution. Although stewardship of these collections is important when making acquisition decisions, the AIMS report advises that, "acquisition of born digital material should be based firstly on a curatorial appraisal of its fit within the collection development policies of the institution". [13]

Thus, when appraising born digital content for acquisition:

1.) Consider collection fit within collection development policies (collection manager's responsibility)

2.) Consult with appropriate library committees and/or specialists to draft a feasibility or technical appraisal prior to making a final decision. In cases of small, homogenous acquisitions (for examples, collections consisting of less than 10GB of content in stored in contemporary and familiar file formats) a technical appraisal will likely be sufficient. In cases of unusual or large acquisitions the feasibility study or technical appraisal process may warrant the advice of the nascent Digital Collections Advisory Committee.[14] Unusual acquisitions can include:

   a. Content that was created more than fifteen years ago

---

[13] AIMS Work Group. (2012). *AIMS Born Digital Collections: An Inter-Institutional Model for Stewardship.* Retrieved from http://www.digitalcurationservices.org/files/2013/02/AIMS_final.pdf

[14] This committee is still in development. Its charge and area of responsibility is currently in flux.

b. Content created in a non-Windows or non-Macintosh environment

c. Collections of heterogeneous content – this may include collections that have many different file format types, and many different file content types (e.g. collections that contain word processing documents, audiovisual material, data set material, etc.)

d. Collections containing files which are dependent upon particular software and hardware for access.

e. Collections which have over 10GB of heterogeneous content.

3.) Refer to the document **GUIDELINES: Levels of Preservation and File Format Support** for guidance on varying levels of preservation and common file formats.

GUIDELINES: Transfer of Storage Media and Computer Systems from Offsite to Library

Computers systems and storage media are fragile. Even if intended to be portable, systems should be handled with care; especially if working with equipment that is more than several years old and whose operating history is unknown. A hard disk may be operating fine upon inspection at an offsite location, but transport may cause damage to delicate physical media causing irrecoverable data loss.

Mechanical and electrostatic damage are often the greatest concern when moving computer equipment and storage media. If you are planning on transporting media and equipment from an offsite location to the main library keep these guidelines in mind. Hard drives, both external and internal, especially drives over ten years old, are subject to damage from jolting, vibration, or other movement which can cause parts within the drive enclosure to contact the delicate data-storing platters enclosed within the drive and cause irrecoverable data loss. Some older hard drives, primarily those used in PCs in the early '80s, may require manually moving the drive head to a safe spot via program commands (known as "parking"[15]) prior to moving the hard disk and computer. If you are moving equipment manufactured in the early '80s please request assistance from the Digital Preservation Coordinator prior to transport.

It is especially important to ensure that drives, or parts containing drives such as computer chassis, are tightly wrapped and secured before transporting them. If items are being shipped via commercial shipper it is imperative that they are wrapped in at least three inches of padding on all sides and secured in a sturdy shipping container to avoid adverse effects from

---

[15] See: http://www.pcguide.com/ref/hdd/op/act_Parking.htm about parking and to learn more about various hard drive parts, general operation, and technologies.

rough handling. If items are shipped long distances or subject to cycling temperatures, they

should be allowed to acclimatize before attempting to power them on.

If you would like supplies such as anti-static bags or hard drive protection boxes or

would like moving consultation assistance please contact the Digital Preservation Coordinator,

Tracy Popp (tpopp2@illinois.edu)

Protect hard drives and media from electrostatic and physical damage

- If you have an internal hard drive with exposed circuitry, place it in an anti-static bag or hard drive protection box.

- If not already in one, place optical media in a jewel case to avoid scratching disk surface or breaking disk.

- Place all other media in an anti-static bag, unless they are already housed in some type of protective case.

- If you have multiple pieces of media, and they each have a hard outer plastic shell and are in good condition, you may place multiple items within a single anti-static bag; however, ensure that no item is abrading another or causing damage if there are multiples within a single bag. Do not overfill bag.

- Before transporting, wrap item in several inches of bubble wrap and place items in a clean, sturdy shipping container. In lieu of bubble wrap, use foam within the shipping container, but ensure that the items are securely packed on all sides and are not abrading against other items. Do not stack or pack multiple items together without separating them with bubble wrap or foam.

- If moving equipment that is particularly sensitive, valuable, or older than twenty-five years please contact the Digital Preservation Coordinator. Insurance may also need to be procured for these types of items. Please contact someone in Library Shipping for information about transport insurance.

If moving laptops, desktop systems and peripherals:

- Unplug all cables, neatly wind them, and keep them together in a bag or some other container. Label bag appropriately to avoid loss.

- Keep all cables with their corresponding computer system.

- Please keep separate bags for peripherals or label peripheral power supplies and cables. Power adapters and cables can have different voltages and can cause damage if plugged

into the wrong device.

- If possible, remove hard drives from desktop chassis and place in hard drive protection box, or anti-static bag prior to transport. If neither is available or removing the hard drive is impractical, would cause damage to the system if removal was attempted, or you are unfamiliar with working with computer system internal parts, leave the drive in the chassis and wrap chassis very tightly in several layers of bubble wrap.

- Place wrapped items in a clean, sturdy container, preferably one with handles for easy moving. Do not overfill container.

Commercial-grade 2.5" hard drive protection box:

Commercial-grade 3.5" hard drive protection box:

FORM: Born Digital Content – Request for Reformatting

Request Administration Section

Requesting Library Unit: _____

Unit Contact Person: _____

Unit Contact Email: _____

Unit Contact Phone: _____

Date of Request: _____

Date Reformatted Contents Needed By: _____

Number of Items: _____ **Rush Order?** Y / N

What is the purpose of reformatting? (Select one)

Patron Request - Access   Access   Preservation   Preservation and Access

Media Type Section

Types of media (select all that apply):

   Optical Media (CDs/DVD)   Floppy Disks (3.5" /   5.25")

   Zip/Jaz Disks   Flash media   Hard disk drives (internal)

   Hard disk drive (external; USB / Firewire)   Data Tape

   Computer System (laptop / desktop)   I Don't Know

Other: _____

If hard drive(s) what is the storage capacity of the drive(s)? _____

   _____

_____

Content Section

Approximate Year Content Was Created: _____

_____

Collection Subject:_____

Is a there a call number or some other unique identifier (e.g. a record series number)

assigned to the item?   **Y /   N**   **If yes, please list the call number/unique**

identifier if it is not affixed to the item:_____

_____

_____

_____

_____

In what operating system environment was the content created?

     Windows   Macintosh (pre-1998)   Macintosh (post-1998)

     Unix/Linux   I Don't Know   Other: _____

What types of files may be found on the media (select all that are applicable):

Text (e.g., word processing files, email)   Images (e.g. .TIFs, JPEGs)

Software/Multimedia   Audio (e.g. .WAV files, MP3 files)

 Video (e.g. .WMV, .AVI, .mpeg)   Databases/Data sets

Programs/Scripts (e.g. Java source files)

Complex files (e.g. files containing more than one content type such as a PDF with embedded

media)

Other (please describe):   _____

What are your expectations for reformatted content?

- **Maintaining the original file structure, look, and functionality** – it's important to retain how the original object was experienced and functioned; If the original content is a program or series of events, the program or series of events should be preserved to interaction and functionality (e.g. maintaining the various audio track configurations within a digital audio workstation project; maintaining the functionality within a database to manipulate data within the database; keeping macros, formulas and other dynamic and program-like functionality within a spreadsheet document.)

- **Maintaining the content** – original fonts, layouts and other design elements aren't as important as long as I can read the content; If the original content is a program or series of events retaining the results of the event or final production output is acceptable (e.g. saving a report from a database; outputting content stored in a production desktop publishing program file to a digital file; outputting the results from digital audio workstation environment to a single audio file.)

- I don't know yet.

Are there particular types of content or a specific file that you're looking for?

No \ Yes If yes, describe the content or provide specific file names: _____

_____

_____

_____

Content Rights Section

What is the legal copyright basis? (Select one)

Protected by U.S. Copyright   Protected under Statute   Protected under

License

Protected by international copyright law (Berne Convention, Canadian copyright law, etc.)

Public Domain   Other   None   I Don't Know

Does the Library have the right to copy or modify this content for access? (Select one)

No \ Yes   Rights Basis: Granted Permission by Rights-Holder*   Fair Use

Public Domain   Orphaned Work   Other   I Don't Know

(if permission granted by rights-holder include permissions statement with item)

Does the Library have the right to copy or modify this content for long-term preservation?

(Select one)

No \ Yes   Rights Basis: Granted Permission by Rights-Holder*   Fair Use

Public Domain   Orphaned Work   Other   I Don't Know

(if permission granted by rights-holder include permissions statement with item)

What access rights limitations are there on this content?

Access is restricted   Access is open and unrestricted

Access is restricted to campus-use only   Access is restricted to library use

only

Access is restricted to library use only on a computer that does not have

network\Internet access

Is there the potential for sensitive information to be found on the media? (e.g. tax records,

credit cards, social security numbers, personal health information, student records, etc.)

Yes/   No

Any other rights concerns with providing access\performing preservation actions to this

object and all derivatives? Please explain:_____

Reformatted Content Landing Section

Where will reformatted contents be stored?

*Path to networked file share:*_____

External hard drive / portable media provided by Unit.

Type: *External hard drive   flash/jump/thumb drives*

*Other:*_____

Portable media provided by Preservation:

    optical media

    lend external hard drive

(subject to availability; hard drive MUST be returned to Preservation Unit within 72 hours of receipt)

Other:_____

Do you have a preferred naming convention for the root folder of the reformatted content (e.g. a record series number, call number, etc.)?_____

Internal Use Only:

Date content received:_____ Received by:_____ Rush order? Y/N

Notes:_____

PROCEDURE: Beginning a Media Reformatting Project

Begin this step once media has been received from a collecting unit. Previous project tasks, including surveying media, generating an appraisal report, or creating a project plan, may have already been completed. Before starting this step verify on the R:\DigitalMediaReformatting share that directories and content related to this project have not already been created.

Tools Required:

- **External hard drive**
- **Computer Workstation**
- **XXXX_Reformatting spreadsheet**

Receive Media

I. ***Log into a computer attached to the Library's network. Open the ReformattingStats.xls*** spreadsheet found here:
\\libsysdigi2.library.illinois.edu\Reformatting\2013_ReformattingStats.xls
If you cannot access the spreadsheet please see Tracy.

II. ***Enter project administrative information into Excel spreadsheet***
   a. If necessary, create a worksheet corresponding to the content's originating unit. Use the three-letter code specified below for the unit, followed by an underscore and the year. For example, "AHX_2013," if the content came from the University Archives. If the unit does not have a corresponding abbreviation, create a three letter code which roughly approximates the naming pattern exhibited below (e.g. Fine and Applied arts, would become FAA_2013):
   i. Standard Unit Abbreviations:

AFX	African/Afro-Americana
ACE	Agriculture, Consumer & Environmental Sciences
AHS	Applied Health Sciences
ARX	Architecture and Art
BIX	Biology
DOC	Gov docs
CHX	Chemistry

STX	Circulation/Bookstacks
CLX	Classics
BEL	Business and Economics
CMX	Communications
EDX	Education and Social Sciences
ENX	Engineering
HPN	History, Philosophy, and Newspaper
IHX	Illinois History and Lincoln Collections
LLX	Literature and Languages
MAX	Map and Geography
MTX	Mathematics
MPA	Music and Performing Arts
NHX	Natural History Survey
RBM	Rare Books and Manuscript Library (special collections)
REX	Reference
SPX	Slavic and East European
UGX	Undergraduate
UHX	University High School
VEX	VetMed
SCX	Scholarly Commons
HID	History Department
SOU	Sousa Archives
IAS	International and Area Studies

b. In the first row for this particular project, complete the following fields. The first row of a new media group received will contain this information. It should not be repeated beyond the first row corresponding to this particular media group. *(Note: This is a good case for using a database rather than a spreadsheet to avoid information redundancies; this spreadsheet will serve as the basis for creating future database fields).*

(M) preceding a field name means that the field is mandatory and must be filled in;
(NR) indicates that the field is not required, but if it is available, it would be useful to include.

AdministrativeInfo

i. Requesting_Unit (M) – The name of the holding library unit
ii. RequestingUnit_ContactPerson (M)- The contact person's name who is requesting reformatting work
iii. Contact_Phone (M) – The contact person's phone number
iv. Contact_Email (M) – The contact person's email address
v. Date_Received (M) – the date the media was received by the Preservation Unit
vi. Total_Items_Received (M) – The number of items received by the Preservation Unit (at this point a more fine-grained inventory, counting and separating items by format, is not necessary – this will be done at a later step. However, do note if items beyond what is being reformatted, such as accessories or cables, is included with the media group.)
vii. Request_Needed_By (NR) – Date requested by the unit contact person for reformatted content receipt
viii. RushOrder? Y/N (NR)- Indicate "Y" if this is a rush order; "N" or blank if it is not

III. *Manually create directories on an external hard drive*

a. Attach an external hard drive to the networked computer you are using.
b. Identify the external drive's drive letter.
c. If the drive has been used for media reformatting the past, it should contain a MediaReformatting directory. If it does not, create it in the root directory.
d. In the MediaReformatting directory look for a folder that corresponds with the originating library's unit abbreviation (see: Sec. II.a.i). If there is not one, create it using the standard unit abbreviations.

e. If you are creating the folder also create a subdirectory corresponding to the current year (e.g. 2013).

f. Create another sub-directory and give it the collection name (e.g. BELFloppyDisks). Use camel casing between words. Camel casing is when multiple words are strung together without spaces and the first letter of each word in capitalized. This convention is meant to increase readability without adding spaces to a file name.

   i. In this directory create two additional sub-directories:
      *CollectionName*Content
      *CollectionName*Documentation
      e.g. BELFloppyDisksContent, BELFloppyDisksDocumentation

g. Given the examples above the following directory structure is generated:

[root of E:\] MediaReformatting

      ---> BEL

          --->2013

              --->BELFloppyDisks

                  --->BELFloppyDisksContent

                  --->BELFloppyDisksDocumentation

IV. ***Begin laying groundwork for reformatted digital object file and directory structure.***

a. The following fields in 2013_ReformattingStats.xls will be completed in this section:

ProjectInfo

   i. **ProjectName (M) -** Create a name for this project; will often be the name of the collection or subject. It is for the Preservation Units internal use only; consulting an authority file for directory naming is not necessary.

   ii. **ItemID - (unitMailcode_sequentialnumber) (M)** – three-digit unit mail code established in section I.a.i in this document. Establish a unique ID used for tracking within the Preservation Unit:

      1. In the ItemID field enter the three characters unit abbreviation, an underscore and a sequence of seven numbers. The numbers should be sequential and padded with leading zeros. For example, the first

item received from Scholarly Commons would have a leading SCX, an underscore and 0000001 (e. g. SCX_0000001).

2. If items have been previously reformatted from Scholarly Commons use the next row/sequence of number available. As an example the last item reformatted for Scholarly Commons was assigned SCX_0000081. The next item would then be SCX_0000082.

iii. **ItemNumber (1 of #) (M) -** Enter the number of items. If the item is part of a series indicate what part it is in the series. E.g. if a series contains three disks you would enter 1 in the field that corresponds with disk one, 2 in the field that corresponds with disk two and 3 in the field that corresponds to disk three. All three disks should use the same ItemID. E.g. The three disks that comprise item SCX_0000082 would each have SCX_0000082 in their ItemID field. This is to ensure that disk series are grouped together and subfolders created for each disk in the series.

iv. **ItemName (labeltext - nospaces) (M) -** Enter the item's name. This may be a title or label transcribed from the disk. If there is no text on the disk name the item by its media type and follow that with unlabeled. E.g. an unlabeled floppy disk is called FloppyUnlabeled.

v. **First10CharsOfItemName (M) – autogenerated -** In the First10CharsOfItemName the name provided in ItemName is truncated to ten characters using an Excel command. E.g. =LEFT(K71, 10)

vi. **CollectionSubject (M) -** enter the collection subject using the first four characters of the collection subject's last name and the first two characters of the subject's first name. If the CollectionSubject is not a person's name enter the first six characters that best represents the collection (e.g a collection of disks from Fine and Applied Arts may be called FAAdis)

b. **A number of fields in the spreadsheet are created through concatenating data from other fields.** For example, the DirFileName field is created using this formula: =TEXT(J2,"000")&"_"&L2&"_"&M2. What this formula does is combines the text from the **itemNumber (1 of #) column,** formats it to a three digit number, adds an underscore after the numbers, add the text from the **ItemName (labeltext - nospaces) column**, appends another underscore and then adds the text from the **First10CharsOfItemName column**.

c. A number of fields include text used to create a directory path that will be used in a batch file to create the directories via automation.

i. DirFileName (M) – autogenerated

ii.   DirPath (M) – autogenerated

iii.   Make Directory Command (MD) And Root Path(M)

iv.   Full Path (M) – autogenerated

v.   ImgSubDir (M) – enter ImageCapturedFromDisk

vi.   CreateImgSubDir (M) – autogenerated

vii.   OrigSubDir (M) – enter OriginalFiles

viii.   CreateOrigSubDir (M) – autogenerated

ix.   MigSubDir (M) – enter MigratedFiles

x.   CreateMigSubDir (M) – autogenerated

xi.   DocSubDir (M) – enter ObjectDocumentation

xii.   CreateDocSubDir (M) – autogenerated

xiii.   PicSubDir (M) – enter ObjectJPG

xiv.   CreatePicSubDir (M) – autogenerated

Auto generated fields and their formulas:

a.   **DirFileName** field is created using this formula:
     =TEXT(J2,"000")&"_"&L2&"_"&M2 (fields: itemNumber (1 of
     #)_First10CharsOfItemName_collectionSubject)

b.   **DirPath** is created using this formula: =CONCATENATE(I2, "\", N2) (fields:
     Item_ID (unitMailcode_sequentialnumber)\ Dir/FileName)

c.   **FullPath** is created using this formula: =CONCATENATE(P2,O2) (fields: **Make
     Directory Command (MD) And Root Path, DirPath)**

d.   CreateImgSubDir is created using this formula: =CONCATENATE(Q2,"\", R2)
     (fields: **Full Path\ImgSubDir)**

e.   CreateOrigSubDir is created using this formula: =CONCATENATE(Q2,"\",T2)
     (fields: **Full Path\OrigSubDir)**

f.   CreateMigSubDir is created using this formula: = CONCATENATE(Q2,"\",V2)
     (fields: **Full Path\MigSubDir)**

g.   CreateDocSubDir is created using this formula: =CONCATENATE(Q2,"\",X2)
     (fields: **Full Path\DocSubDir)**

h.   CreatePicSubDir is created using this formula: =CONCATENATE(Q2,"\",Z2)
     (fields: **Full Path\PicSubDir)**

i.   The **Make Directory Command (MD) And Root Path** field contains the
     Windows command line command MD (make directory) and the path to the
     directory where you'd like to create the file structure. For our workflow this will
     be in the "*CollectionName*Content" directory that we created in Sec. III.f.i. For
     example, if you were working on this spreadsheet for the example in Sec. III, it
     would end up like this:

     MD

E:\MediaReformatting\BEL\2013\BELFloppyDisks\BELFloppyDisksContent\

ENTER THIS INFORMATION ONLY ONCE – it can be copied to the other object fields in this collection, so there is no need to retype it.

j. The **ImgSubDir** field contains the text which will be used to name the ImageCapturedFromDisk directory. This directory is where the captured disk image is stored. Type ImageCapturedFromDisk in the first field and copy down the column as corresponds to the number of items you're transferring.

k. The **OrigSubDir** field contains the text which will be used to name the OriginalFiles directory. This directory is where files in their original format have been extracted from the disk image. Type OriginalFiles in the first field and copy down the column as corresponds to the number of items you're transferring.

l. The **MigSubDir** field contains the text which will be used to name the MigratedFiles directory. This directory is where files that have been migrated from their original format to an updated and/or normalized file format. Type MigratedFiles in the first field and copy down the column as corresponds to the number of items you're transferring.

m. The **DocSubDir** field contains the text which will be used to name the ObjectDocumentation directory. This directory is where reports or documentation related to the digital object (NOT ENTIRE COLLECTION) are stored. Type ObjectDocumentation in the first field and copy down the column as corresponds to the number of items you're transferring.

n. The **DocSubDir** field contains the text which will be used to name the ObjectDocumentation directory. This directory is where reports or documentation related to the digital object (NOT ENTIRE COLLECTION) are stored. Type ObjectDocumentation in the first field and copy down the column as corresponds to the number of items you're transferring.

o. The **PicSubDir** field contains the text which will be used to name the ObjectJPG directory. This directory is where digital pictures to the digital object are stored. Type ObjectJPG in the first field and copy down the column as corresponds to the number of items you're transferring.

V. ***Begin Creating the Directory Structure on the External Drive or Network Share***
   a. Copy the text from the FullPath, CreateImgSubDir, CreateOrigSubDir, CreateMigSubDir, CreateDocSubDir, and CreatePicSubDir columns for the collection you are working on. Open a text editor (e.g. Notepad) and paste the text into the document. Save the file as fileCreate.bat anywhere on the drive (you're explicitly providing an absolute drive path, so it should not matter where the file

is saved.)

b. Run the batch file to create the directories before disconnecting the external hard drive. If you run this process after you've connected it to another machine, it will not run correctly, unless you explicitly map the external drive to the same drive letter as the previous computer.

VI. *Inventory Media and Organize for In-House or Vendor Transfer*

*Note: the following steps are not required to be completed sequentially; you may choose to label each of the media within a media group, photograph each media and then enter inventory information about each piece into the spreadsheet. For the sake of creating a workflow, they have been listed sequentially. Just ensure you work in an organized way and that each item is accounted for and labeled appropriately.

a. Identify a logical place to begin labeling media sequentially. Write the ItemID and, for disk series with more than one item, Item Number on a small Post-it Note. Attach it to the item. Work through in order that the items are received or at least in some grouped sequence (especially if you have multiple containers within a larger box) to maintain some sense of original order.

b. After the media is labeled, you may sort media by various formats to ease workflow and to separate media types for in-house or vended transfer. ***Once reformatting is complete return the media to the collection or storage box in sequential order by label, which should resemble original order as received from the holding unit.***

c. Identify media that can be transferred in-house and which requires vendor attention. Separate those storage media which require vended services.

d. Photograph the item. Ensure that the Post-it is viewable, but does not obscure any label or textual information on the media.

e.   At this stage items are evaluated for in-house or vendor formatting by their format and our ability to access the format in-house with currently available equipment.

VII. **Item Inventory and assessment:**

a. Open the 2013_ReformattingStats.xls spreadsheet. The following fields in 2013_ReformattingStats.xls will be completed in this section: *MediaInfo*

    i. **Quantity (M) -** Number of items this record corresponds to. Often, this will be one, but in cases where one does not gain much by distinguishing between items (such as punched card media) you can process a group of items rather than creating separate records and directories for each discrete item.

    ii. **TranscribedLabelInfo (M)** - If there is information on the label, transcribe it here. Try to capture what appears to be the most information. We will also be photographing the media so no need to excessively document; can also type "See Photo Doc" if there is a lot of information. If no label information, do not enter anything into this field.

    iii. **OriginalFormat (M) -** The digital media format; use a controlled list:
1. 5.25" floppy disk
2. 3.5" floppy disk
3. 8" floppy disk
4. internal hard drive
5. external hard drive
6. desktop
7. laptop
8. DVD
9. CD
10. Iomega ZIP
11. Iomega JAZ
12. Flash Media
13. Other (if other, enter information about the format in the notes field)

    iv. **ManufacturerBrand/Model (NR) -** If available, enter the manufacturer name and model number of the source media. If the data is being reformatted from a computer system, enter the name and model number of the machine (e.g. Mac PowerBook 3400c)

    v. **ContentDate (if none, enter NONE) (NR) -** Approximate date of the content – only enter this if there is a date to transcribe from the storage media's label.

    vi. **StorageCapacity (M) -** maximum storage capacity of the media

vii. **If HDD, Connector Type (NR) -** the type of bus used to interface with the hard drive. This is an indicator of what type of port interface and cable is required for a host computer to work with the hard drive. Use a controlled list:
   1. USB
   2. FireWire
   3. ESDI
   4. SCSI (type of SCSI such as I,II, or III male, or 50-pin Centronic, can be entered free text)
   5. IDE
   6. SATA/PATA
   7. Other (free text other types)

viii. **If tape, type (NR) -** configuration of the tapes housing. Use a controlled list. Free-text may be used if the tape type is not represented.
   1. Open-Reel – 9-track
   2. Cartridge – (free text describing the cartridge type, if known)
   3. Cassette – LTO1
   4. Cassette – LTO2
   5. Cassette – LTO3
   6. Cassette – LTO4
   7. Cassette – LTO5
   8. **Cassette – (free text other)**

ix. **ReelSize (NR) -** If open-reel tape, record the size of the reel.

x. **Density (tape only) (NR) -** Record the density of the tape, if known (may be printed somewhere on the tape label)

xi. **Tape Length (tape only) (NR) -** footage or length of tape if known (may be printed somewhere on the tape label)

xii. **TPI (tape only) (NR) -** Tracks per inch; also a measure of density; record if known (may be printed somewhere on the tape label)

xiii. **Single or Double Sided? (NR) -** Is information recorded on both sides of the recording substrate, if known

xiv. **Is this a "flippie disk"? (NR) -** A flippy disk (sometimes known as a "flippy") is a double-sided 5¼" or 8" floppy disk, specially modified so that the two sides can be used independently (but not simultaneously; the disk must be flipped with the appropriate side facing the drive read head in order to be accessed). Use of "flippy" disks was most common during the 8-bit home computer era of the early to mid-1980s.

xv. **Notes (NR) -** Any free-text notes about the storage media.

b. Based on previously assessing media for in-house or vended service transfer, continue on to the either then In-House Transfer step or Vended Transfer step.

VIII. *In-House Transfer*

a. Copy 2013_ReformattingStats.xls spreadsheet to external drive \\Reformat_Interim\ReformattingMetadataTracking\ (external hard drive – media reformatting machine is not on the network so the spreadsheet needs to be available on the external hard drive in order to enter tracking data as reformatting)

b. Start the reformatting machine. Use KVM switch on desk to view the reformatting machine desktop (the KVM shares a single monitor, keyboard and mouse between FRED or LAMONT)

c. If the external drive is still plugged into main workstation, eject it

d. Plug external hard drive into the reformatting machine you will be using

e. If using the FRED, copy over the directory structure you created on the external drive to D:\MediaReformatting; if using LAMONT, leave the directories on the external drive.

f. While reformatting content note the date the content was transferred and how long the transfer process takes to complete. Record this information in the TransferDate and TimeToTransfer fields on the ReformattingStats_YYYY.xls spreadsheet.

g. Continue on to: **PROCEDURE: Transferring Data from (specific media)** depending upon the various media types.

IX. *Vended-Transfer*

a. Copy the directories to the R:\DigitalMediaReformatting into the appropriate location.

b. Contact the appropriate vendor per media type. Most hard disk drives will go to DriveSavers. The Digital Preservation Coordinator will keep a record of vendors to contact based on their media reformatting capabilities.

c. Once a vendor is selected and contacted, fill in the vendor reformatted fields on the ReformattingStats_YYYY.xls spreadsheet.

VendorInfo: **This information may also be recorded in the item treatment plan, but should also be recorded here for centralized location and reporting purposes.**

   i. **Send to Vendor for Reformatting? (NR) –** Indicate Y in this field if sending to a vendor.

   ii. **VendorName (M) –** Mandatory if item is being shipped to a vendor for reformatting. Name of the vendor receiving the media

  iii. **VendorPhoneNumber (M) -** vendor phone contact information

  iv. **VendorEmail (NR) –** vendor email contact information

  v. **Vendor_Job_No (NR)** – vendor job number or project number; the format will be different among vendors

  vi. **EstCost (NR) –** approximate cost of media transfer job; NOTE: may not be a one to one cost per item.

  vii. **Shipping Box No (NR) –** if more than one shipping box is being sent to the vendor, indicate what number box the media was shipped in (e.g. if shipping two boxes and an item is shipped in box one of two, type "1" into this field.

  viii. **Shipped Date (M) –** date item was shipped to a vendor

  ix. **Recoverable? (NR) –** indicate if vendor was able to recover contents from media

  x. **Vendor_JobNotes (NR) –** notes on vendor comments about the media assessment

  xi. **ActCost (NR) –** actual charges incurred from the vendor

d. Once content is received from the vendor proceed to **PROCEDURE: Receiving Vendor Reformatted Content**

POLICY: Addressing Computer Viruses, Malware and other Potentially Harmful Content

In order to protect the Library's digital collections and networked, production computing environment, born digital content are scanned for potentially harmful content in a controlled, non-networked computing environment. As part of the disk processing workflow, disk images and extracted files are scanned; disk images are subject to anti-virus scanning after they are created. Files exported from disk images are scanned as they are exported, and once again after all of the disk contents have been exported – prior to transferring the data to a networked environment. Processing records on a non-networked machine and in a controlled environment limits the potential for wide-spread damage or file infection should a virus be found; the range of harm limited to the local machine only. This protection saves already processed digital collections from infection, as well as shields administrative and production technology from possible damage.

The non-networked machine's anti-virus and software is updated on a regular basis. As of 2/19/13 the Microsoft Forefront Endpoint Protection version 1.143.2608.0 is installed on the FRED forensics workstation.

When a Virus, Malware or other Potentially Harmful File is Detected:

When the anti-virus software detects a problem the default action is to remove the offending virus and related virus or malware files from the computer. If the virus cannot be successfully removed, affected files may need to be deleted. A record of removal action will be made and will be included with the object documentation. The malware or virus will be removed from the disk image or transferred content only; the source media will not be altered.

Infected virus file removal may include altering both the Preservation Master image as well as the working\access files. If there is reason to retain potentially harmful files collection

managers should indicate in writing and ensure that they steward the harmful files appropriately; that is, the collection manager should find an alternate means of file storage beyond the preservation repository or other networked environment where a malicious file's content could damage other files.

POLICY: Disk Image Creation and Retention

For the purpose of capturing a complete and accurate record of contents reformatted from born digital media the Born Digital Content Preservation service employs forensic disk imaging methods. Forensic disk images are bit-by-bit mirror copies created of the original physical storage media. The images contain a complete picture of the storage media and include all features of the original media. These features may include: deleted files, unused sectors or storage space, any data stored outside of a disk partition.

The disk image also retains source media metadata and maintains the source media file permissions, change and access dates and other information that can be modified through less accurate approaches in data moving such as simply copying individual files and folders from source media to a new storage location. Forensic disk imaging also allows for the duplication of disks that have been formatted or created within a different computing environment than the host computing environment used to image the disk. For example, one may be able to extract data from a disk created on an obsolete Macintosh computer using a modern computer running the Windows operating system.

When imaging source media that contains complex information or an entire computing environment, such as a hard drive that housed a donor's host operating system and applications, collecting and retaining a forensic image can help maintain context – through retaining the structure and all files from an original disk, a researcher may better understand how the disk owner organized files, who else used the media, which users had access to particular files and other contextual information that may not be readily understood through investigating working files copied from a hard disk drive.

Imaging also limits use of the original storage media – one can view the proxy disk image rather than continuing to subject the original media to potentially damaging read/write processes. The original media may be too fragile to endure repeated read\write\access processes. Creating a forensic disk image allows users to interact with a surrogate that has been stabilized.

Reformatting Practices:

As part of the Born Digital Content Preservation service reformatting workflow, by default all media will be imaged in order to maintain an accurate and complete record of the source media and content. Sometimes, imaging is out of necessity – in reformatting 5.25" floppy disks the hardware and software first must create a disk image as there is generally no way to view the disk contents without first extracting an image.

Example where disk imaging might not be used:

- When the original storage media contains only files and directories of working documents which has been selected and transferred by a donor prior to acquisition; that is, it doesn't contain additional contextual information such as application and operating system data and may be representative of a file backup, the media is understandable to the host operating system, the content was selected and copied to a secondary medium by the donor AND a write-blocking device is available. For example, a USB flash drive containing only files and directories copied by the donor and used to organize the files formatted for use in a relatively modern Windows operating system may be connected to a USB-write blocking device.

Imaging Format:

Most images will be created in the Advanced Forensic Format (.aff). This forensic disk image format is an extensible open-source format that allows for retention of metadata about the image in addition to the image itself. AFF images may also be compressed using the lossless and open-source Zlib file compression algorithm[16]. Compression allows the disk image to consume less storage space than the original source media. For example, an .AFF disk image created

---

[16] See: http://www.zlib.net/

based on a 500GB source hard drive compressed to a 42GB file using a level-4 compression option in the FTK Imager software. This allows storage space to be used efficiently in cases where the disk image is retained as part of a long-term preservation object.

AFF is also supported by a variety of tools – both open-source and proprietary – and gaining traction in cultural heritage institution. The BitCurator[17] project, developed by the University of North Carolina – Chapel Hill, is an open-source digital forensic framework and toolset that works with .AFF images and recommends imaging in the .AFF format. FTK imager[18][19], a proprietary but free digital forensics imaging and file examination tool also creates and works with AFF files.

However, due to the digital reformatting lab's current hardware, some floppy disk media may be imaged in a raw format. As the Catweasel [20]floppy disk controller card does not create .AFF images and is presently the lab's method for imaging older, and non-Windows based floppies the disk image format will be a raw, one-to-one copy of the source media. These disk images will have an .img file extension.

Retention of Disk Image File:

The disk image file may contain valuable contextual information about the creator's working habits. It may also provide valuable information about applications used to create files – which can be important when files are exported from the image and prepared for preservation. The wealth of unmodified information potentially held within a disk image makes it a good candidate for retention as a preservation master. However, an image may also contain un-

[17] See: http://www.bitcurator.net/aboutbc/
[18]About FTK Imager: http://www.forensicswiki.org/wiki/FTK_Imager
[19] Download FTK Imager: http://www.accessdata.com/support/product-downloads
[20] About the Catweasel: http://en.wikipedia.org/wiki/Individual_Computers_Catweasel

redacted, sensitive information; thus, an image file should remain within a dark archive with no unmediated use.

An image file can also use a large amount of storage space. Considering the example of the 500GB hard drive compressed to a 42GB image file, the image still is a significant size. Additionally, ancillary materials and contents extracted from the raw image for researcher use will increase storage use. Depending upon the quantity of contents decompressed and extracted from a disk image the complete digital object could be larger than the original source media.

The source media collection manager ultimately maintains responsibility for deciding upon the retention of the disk image. However, the Digital Preservation Coordinator may advise that the image be retained due to factors such as importance of the source media creator, any unusual features captured within the image that cannot be represented well elsewhere, or other circumstances. The source media collection manager must verify **in writing** if they do not wish to retain the disk image as part of the born digital reformatting process. Once the disk image is deleted it is not recoverable – another image would need to be created. Thus, it is recommended that the disk image be retained at the very least until the reformatted digital content has been fully processed and is ready for preservation and\or access.

PROCEDURE: Transferring Data from Windows-formatted and Mac OSX- (HFS) Formatted 3.5" Floppy Disks

Note: If the floppy is not Windows or Mac OSX formatted, or is not successfully imaged using FTK Imager then consult: PROCEDURE: Transferring Data from 5.25" and 3.5" Floppy Disks using the Catweasel

OUTPUTS FROM THIS PROCEDURE:

- **Image file of source media (.aff file) -** Forensic disk image derived from source media

This procedure details the steps necessary to create a disk image using the following hardware and software environment:

- Hardware:

    a. FRED forensics recovery workstation
    b. External floppy disk drive with USB connection

- Software:

    a. FTK Imager (version 3.1.2.0)
    b. Anti-virus scanning software

1. **Note: The UltraBay 3d does not recognize the external USB floppy drive. Do not plug the USB connector into the UltraBay 3d; instead, use the USB ports which directly interface with FRED (the USB ports next to the power switches.)**

2. Take an external USB floppy disk drive from the equipment shelf

Figure 3: External USB floppy disk drive

3. Plug the USB connector into FRED – do NOT plug it into the USB port on the UltraBay 3d

Figure 4: External USB floppy plugged into FRED USB port

4. On FRED launch the imaging software (FTK Imager)

5. Select File > Create Disk Image

6. Select Source:

   From the Select Source box, select the radio button corresponding to **Logical Drive**.
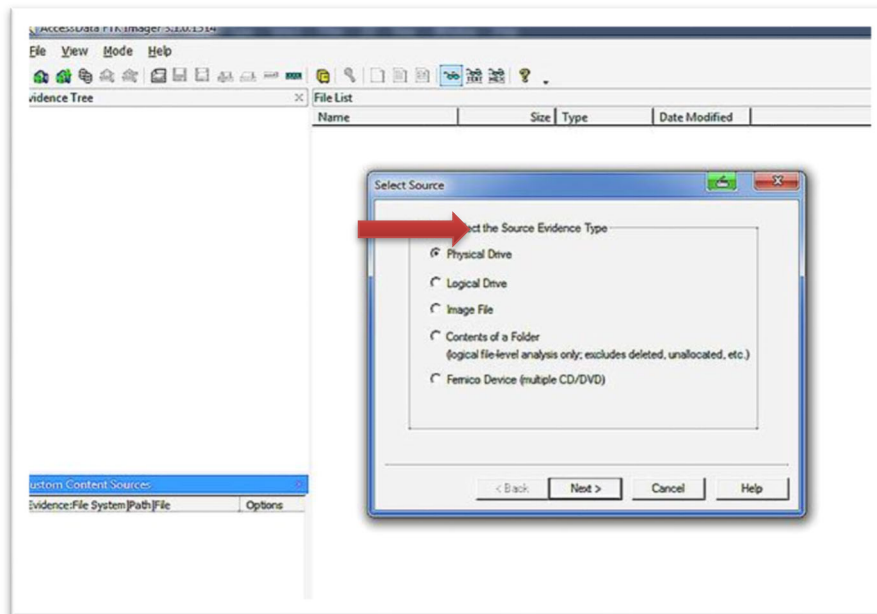
Figure 5: FTK Imager: Source selection box

7. Click Next.

8. Select Drive:

   Click on the down arrow in the box immediately below the text "Please select from the following available drives:" Select the drive which corresponds to floppy drive. Typically this is A:\.

9. Click Finish.

10. Create Image:

    Below the Image Destination(s) box, select the "Add…" button.

11. Select Image Type: Select AFF.

12. Evidence Item Information:

    In this dialog box enter disk image capture metadata/provenance information.

a. Case Number: If University Archives or some other unit has provided a record series number or unique identifier, use it. Otherwise, use a collection name – an arbitrary name provided in-house to reference the collection; e.g. Michael Hart collection

b. Evidence Number: internally-generated unique Item Identifier; this ID is generated earlier in the disk migration workflow as part of creating the reformatting directory structure:

e.g. AHX_0000009.

If this item is part of a series grouped together within a single Item Identifier, append an underscore and a corresponding sequential number to the end of the Item Identifier. If number is a single digit pad the number with one leading zero; .e.g. if this is the second disk associated with AHX_0000009 append _02 to the end of the evidence number: AHX_0000009_02.

c. Unique Description: blank

d. Examiner: enter your name in the last name, first name format: e.g Popp, Tracy

e. Notes: briefly document source media type here; e.g 3.5" Windows floppy

13. After entering Evidence Information click next.

14. **Image Destination Folder:**

Navigate to the ImageCapturedFromDisk directory related to this item. The directory should have been created earlier in the workflow. If it was not, create it now.

15. **Image Filename:**

Give the file the same name as the Evidence Number, which is the same as Item

Identifier and series number; e.g. given the example in step 12b the filename would be:

AHX_0000009_02

16. Image Fragment Size (MB): Enter 0

17. Compression: 0

18. Select Finish.

19. If not selected, check the following boxes:

Verify images after they are created – checks image hash signatures; will generate a

hash for the image if there is not one

Create directory listings of all files in the image after they are created – records the

file names and paths of the image contents; saved in tab-separated value (.tsv) format.

Precalculate Progress Statistics – provides imaging time and storage space required

20. Click Start. Let the imaging process complete.

21. Once imaging is complete and verified close FTK Imager status windows.

22. Remove source media, and return it to the storage space designated for its corresponding

collection.

23. Navigate to the disk image you've just created.

24. Run the anti-virus software on the file. Note any anti-virus activity and record it in the

"migrationNotes.txt" file in the "ObjectDocumentation" directory corresponding to the

floppy you're imaging. If the file does not yet exist, create it.

25. Proceed to the document **PROCEDURE: Extracting Files from Disk Image Using**

FTK Imager (version 3.1.0.1514)

PROCEDURE: Transferring Data from 5.35" Floppy Disk using the Catweasel

See Appendix B: **OUTPUTS FROM THIS PROCEDURE:**

- **Image file of source media (.img file)**

  Disk image derived from source media

- **migrationNotes.txt (.txt file)**

  This file will be created if there are any disk read errors encountered or if the disk is
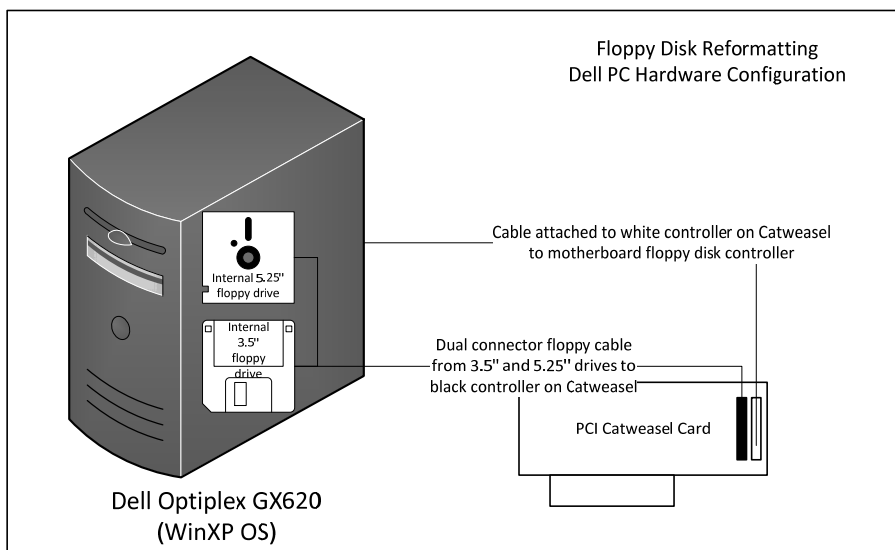
  unreadable via this procedure

Figure 6: Illustration of Catweasel and floppy drive setup on Lamont (expanded).

1. Boot the Dell Optiplex GX620 (Lamont) workstation.

   Ensure that you've prepared a directory structure on an external USB hard drive. The

   directory structure should have been created in Disk Migration Workflow step; you may

   need to copy the directory structure from a networked drive to an external drive in order

   to transfer the data from the non-networked Dell Optiplex.

2. Once the machine has completely booted attach an external hard drive where you've established the reformatted contents directory structure.

3. Double click on the "CatweaselFiles" directory on the desktop.

4. Insert floppy into appropriate drive

   a. 5.25" floppy drive

   b. 3.5" floppy drive

5. In the Imagetool 3.2 interface, if not already done, select the drive letter that corresponds with the appropriate disk drive format (i.e. 5.25" or 3.5")

6. If you are not sure of how the disk is formatted or in what operating system it was created, select the "Detect" button to get better idea of the type of disk in the drive. Make your best guess of options provided in the "Select Format" box. Examine the disk itself for clues on what it might be and how it might be formatted. Older and/or more esoteric formats may require multiple detections and/or readings to successfully capture a disk image.

7. In the ImageTool 3.2 interface select Image > Plain (*.bin, *.adf, *.d64, *.xfd)

8. Once the disk is detected select the "Read Disk" button.

9. In the "Write to File?" > "Show" dialog box select "All Files"
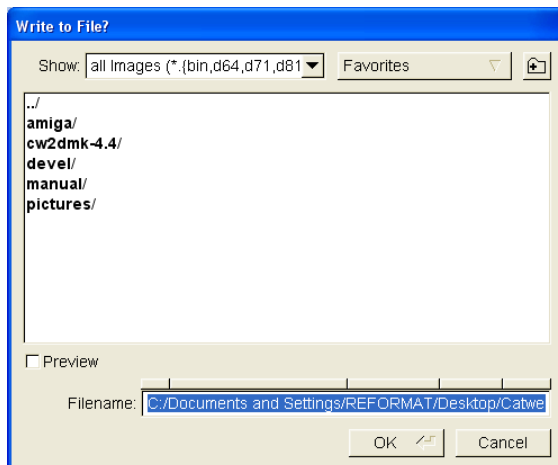
10. Navigate to the "ImageCapturedFromDisk" folder on the external hard drive that corresponds to the floppy disk you're reformatting. For example if the disk is labeled AHX_0000010 navigate to the ImageCapturedFromDisk subdirectory in the AHX_0000010 directory.

11. In the Filename box append the filename to the end of the path. Give the disk image the same name as the Item_ID; append an .img on the end of the name. For example, given the example above, append AHX_0000010.img to the end of the Filename path.

12. Select OK. The drive will begin reading the disk.

13. The Imagetool software provides several points of information during the imaging process including signal histogram, and a visualization of sector data. You can use these tools to gauge if there are weak or bad data sectors on the disk. Green within the visualizations mean good a good data read; yellow means slightly out of range, but probably still readable; red indicates an out of range signal or bad sector with a likely result of data loss or data unreadability within the sector(s) demarcated in red.
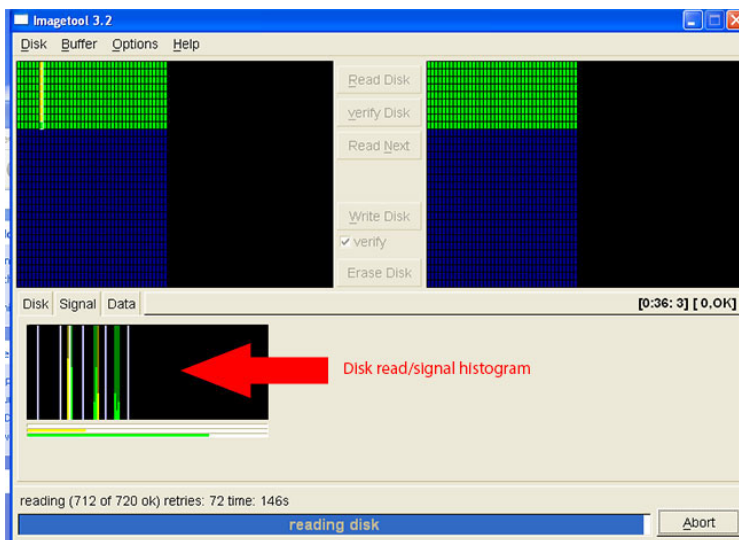
Figure 7: Catweeasel interface: signal histogram

14. **IN CASE OF READ ERROR:** the Imagetool will direct the floppy drive to re-read the disk until it all data signals are in range. Oftentimes, bad sectors will resolve after several re-reads. If bad sectors, particularly sectors flagged as yellow, do not fall within range after several retries hit the "Abort" button.

15. In the "ObjectDocumentation" directory corresponding to the floppy you're imaging create a "migrationNotes.txt" file.

16. Enter the following information:

    a. Date: mm/dd/yyyy

    b. Name: last, first

17. Note: document which tracks/sectors had read errors. You can find the bad track/sector information by clicking on the yellow or red colored blocks in the image tool interface and looking at the related track/sector location information on the "data" tab. If there is an error record that in the note field as well.
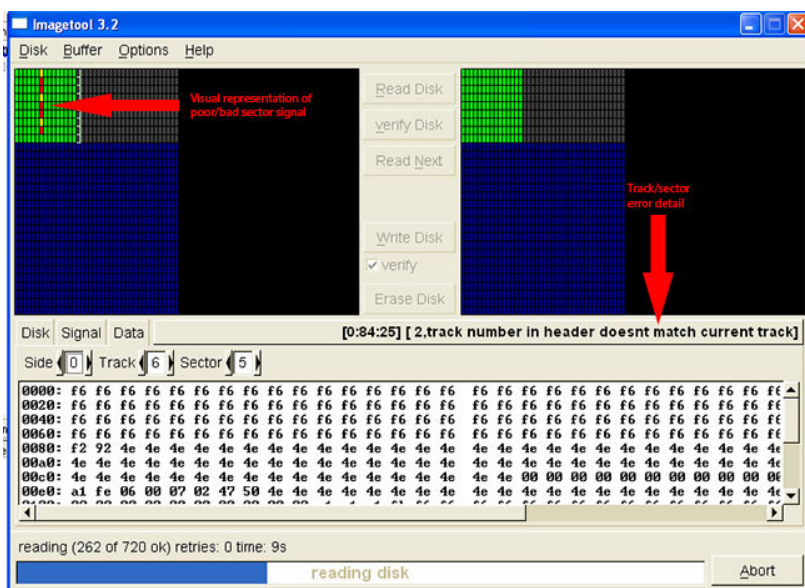
Figure 8: Catweasel interface with signal read error example (sectors flagged in red and yellow) and track/sector error detail.

18. If the bad sectors are contiguous and all seem to have the same error you can simply note the range of bad sectors within the note:

    e.g. track 6, sectors 0-7: [17, bad data checksum] error.

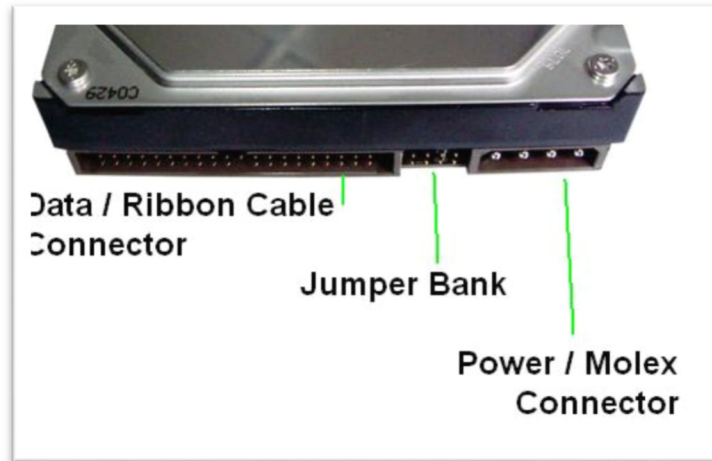19. Once complete continue on to **PROCEDURE: Extracting Files from Disk Image Using FTK Imager**

PROCEDURE: Transferring Data from an Internal Hard Drive: IDE/ATA/PATA/SATA

This procedure details the steps necessary to create a disk image using the following hardware and software environment:

- **Hardware:**
     a. FRED forensics recovery workstation
     b. Tableau UltraBay 3d Write Blocking forensics bridge
- **Software:**
     a. FTK Imager (version 3.1.2.0)

1. **WARNING: Ensure UltraBay 3D is TURNED OFF BEFORE connecting ANY drive cables or power.**

2. **CAUTION:** Handle internal drives with care. The drive heads within the drive enclosure are delicate and can cause data loss if handled roughly. Use anti-static bags and anti-static wrist strap to avoid drive damage due to static electricity, especially if you're working with a drive that has exposed circuitry.

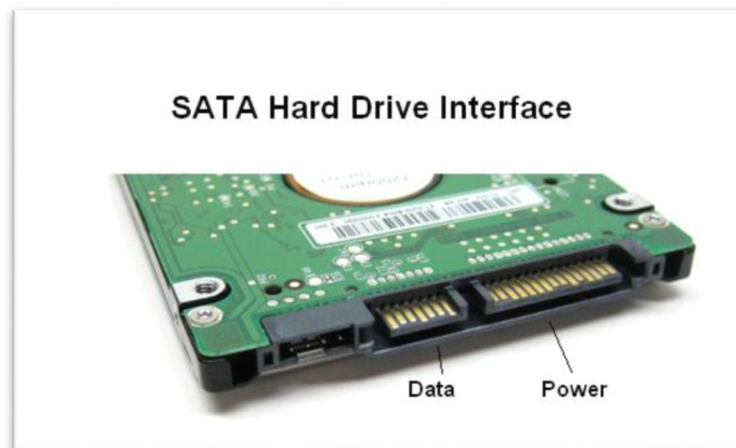3. **Determine if drive is IDE/ATA/PATA or SATA interface.**

a. IDE/ATA/PATA Interface:

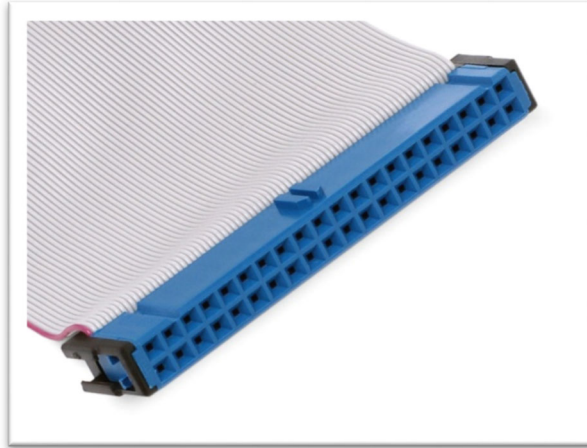Figure 9: From http://blog.shophde.com/2010/02/05/ide-vs-sata-hard-drives/

b. SATA interface:

Figure 10: From http://blog.shophde.com/2010/02/05/ide-vs-sata-hard-drives/

4. Locate the correct cable to interface from drive to Tableau write-blocker:

a. 40-pin IDE/ATA cable

---

[21] http://blog.shophde.com/2010/02/05/ide-vs-sata-hard-drives/

[22] http://blog.shophde.com/2010/02/05/ide-vs-sata-hard-drives/

23

Figure 11: From http://en.wikipedia.org/wiki/File:PATA-cable.jpg

b. 7-pin SATA cable

24

Figure 12 http://en.wikipedia.org/wiki/File:SATA_Data_Cable.jpg

5. Attach cable to drive. **CAUTION:** Be aware of notches and other cable orientation clues

so that drive pins are not bent as you insert the cable. **Do not force** the cable into the

---

[23] http://en.wikipedia.org/wiki/File:PATA-cable.jpg
[24] http://en.wikipedia.org/wiki/File:SATA_Data_Cable.jpg

drive connector. **DO NOT PLUG INTO ULTRABAY YET.**

6. Attach power cable to drive:

    a. If IDE/ATA drive interface connect Molex to 3M power adapter

    (insert adapter image here)

    b. If SATA connect SATA to 3M power adapter.

    (insert adapter image here)

7. On the FRED, pull out the imaging shelf immediately below the UltraBay 3D.

8. If hard drive circuitry is exposed, oriented to the bottom. Otherwise, place drive on imaging shelf in a way that avoids unnecessary cable crimping, bending or crossing.

9. **WARNING: AGAIN, Ensure UltraBay 3D is TURNED OFF BEFORE connecting ANY drive cables or power. (I'm not joking.)** Begin plugging the cables and power into the UltraBay 3D.

    a. Connect IDE drives to the IDE port on the UltraBay 3D.

    CAUTION: Be aware of cable notches and orientation so that you do not bend any pins on the UltraBay as you insert the cable. Do not force the cable into the port.

    b. Connect SATA drives to the SATA port on the UltraBay 3D.

    CAUTION: Be aware of cable notches and orientation. Do not force the cable

into the port.

10. Power on UltraBay 3D. The LED lights on the UltraBay should cycle through when

powering on; once the UltraBay is ready the first four LED lights should be illuminated.

    a.   Power – indicates that the device has power

    b.   Dev – indicates that a device is detected

    c.   Host – indicates that the UltraBay 3D is connected to the host computer

    d.   WrtBlk - indicates that write-blocking is enabled

11. Continue on to **PROCEDURE – Create Disk Image Using FTK Imager**

PROCEDURE: Transferring Data from FireWire/Mac Device

This procedure details the steps necessary to create a disk image using the following hardware and software environment:

- Hardware:

    a. FRED forensics recovery workstation

    b. Tableau UltraBay 3d Write Blocking forensics bridge

- Software:

    a. FTK Imager (version 3.1.2.0)

1. **WARNING: Ensure UltraBay 3D is TURNED OFF BEFORE connecting ANY drive cables or power.**

2. **FireWire External Hard Drive (HD)**

    a. Using appropriate FireWire cable and adapter, connect FireWire external HD to Tableau UltraBay 3d Write Blocker FireWire port.

    b. If necessary, attach power to external HD unit.

3. **Apple Macintosh Computer with FireWire port**

    a. Verify that the target computer supports Target Disk Mode:

        i. The internal hard drive on the target computer must be an internal PATA or SATA drive set to master. FireWire Target Disk Mode will not connect to Slave ATA, ATAPI, or SCSI drives.

        ii. The target computer is running Mac OS 8.6 or later with FireWire software version 2.3.3 or later.

    iii.   These models can be used as target computers:

      1. iMac (Slot Loading) with Firmware version 2.4 or later

      2. iMac (Summer 2000) and all models introduced after July 2000

      3. eMac (all models)

      4. Mac mini (all models)

      5. Power Mac G4 (AGP Graphics) with ATA drive

      6. Power Mac G4 Cube

      7. Power Mac G4 (Gigabit Ethernet) and all models introduced after July 2000

      8. Power Mac G5 (all models)

      9. Mac Pro (all models)

      10. iBook (FireWire) and all models introduced after September 2000

      11. PowerBook G3 (FireWire)

      12. PowerBook G4 (all models)

      13. MacBook Pro (all models)

      14. MacBook models introduced before October 2008

b. Verify that target computer is powered off.

c. Plug in the computer's power adapter.

d. Using appropriate FireWire cable and adapter, connect target computer to
    Ultrabay 3D FireWire port.

Figure 13 Tableau UltraBay 3d - FireWire Port circled.

e. Before booting into Target Mode verify that there is no firmware password first
    by booting the computer:

    i.   Hold down the Option key and press the computer's power button.

    ii. If the computer's Start up Manager appears (a graphical representation of
         the hard disk or disks) there is no password. Shut off the computer and
         continue to the next step.

    iii. If a padlock icon and password field are displayed on screen then there is a
         password which will need to be removed prior to booting into Target Disk
         mode.

    iv. This password can be cleared by restarting the Mac and holding down the
         COMMAND+OPTION-P-R keys on boot.

      v.   Once the password is cleared shut off the computer and continue to the next step.

  f. Start the Mac computer in TARGET DISK MODE.

      i. Start up the target computer and immediately press and hold down the T key until the FireWire icon appears.

1. **Figure 14 Macintosh in Target Disk Mode: Example of FireWire icon**

4. Power on UltraBay 3D. The LED lights on the UltraBay should cycle through when powering on; once the UltraBay is ready the first four LED lights should be illuminated.

  a. Power – indicates that the device has power

  b. Dev – indicates that a device is detected

  c. Host – indicates that the UltraBay 3D is connected to the host computer

  d. WrtBlk - indicates that write-blocking is enabled

5. The hard disk of the target computer should be detected by the UltraBay Write Blocker.

6. Continue on to **PROCEDURE: Create Disk Image Using FTK Imager**

PROCEDURE: Create Disk Image Using FTK Imager

Use FTK Imager to create disk images of most media except for the following:

5.25" disks and non-Windows or Mac OSX formatted 3.5" floppy disks – See

instead: PROCEDURE – Create Disk Image Using Catweasel

3.5" disks – See instead: PROCEDURE: Transferring Data from Windows-formatted

and Mac OSX (HFS+) –Formatted 3.5" Floppy Disks

OUTPUTS FROM THIS PROCEDURE:

- **Image file of source media (.aff file)**

   Forensic disk image derived from source media

- **Drive migration metadata information (.aff.txt file)**

   Text file which contains metadata about:

   o the version of the application used to create the image

   o image administration information entered into the Evidence Item Information

      section

   o information about the source media including:

      ▪ source data size

      ▪ computed hashes

   o information about the resulting image

   o information about the image verification including:

      ▪ number of bad sectors found (if any)

      ▪ MD5 checksum

- ▪ SHA1 checksum

- **Directory listing of files in image (.aff.csv file)**

  .csv file containing the following fields:

  - o **Filename:** name of the file found within the image

  - o **Full Path:** relative path to the file

  - o **Size (bytes):** file size in bytes

  - o **Created:** file creation date (if available)

  - o **Modified:** date file was last modified (if available)

  - o **Accessed:** date file was last accessed (if available)

  - o **is Deleted:** indicator if file has been marked for deletion. If the field is "yes" then the file has been marked for deletion.

Drive Imaging:

1. On FRED launch the imaging software (FTK Imager)

2. Select File > Create Disk Image

3. Select Source:

   From the Select Source box, select the radio button corresponding to Physical Drive.
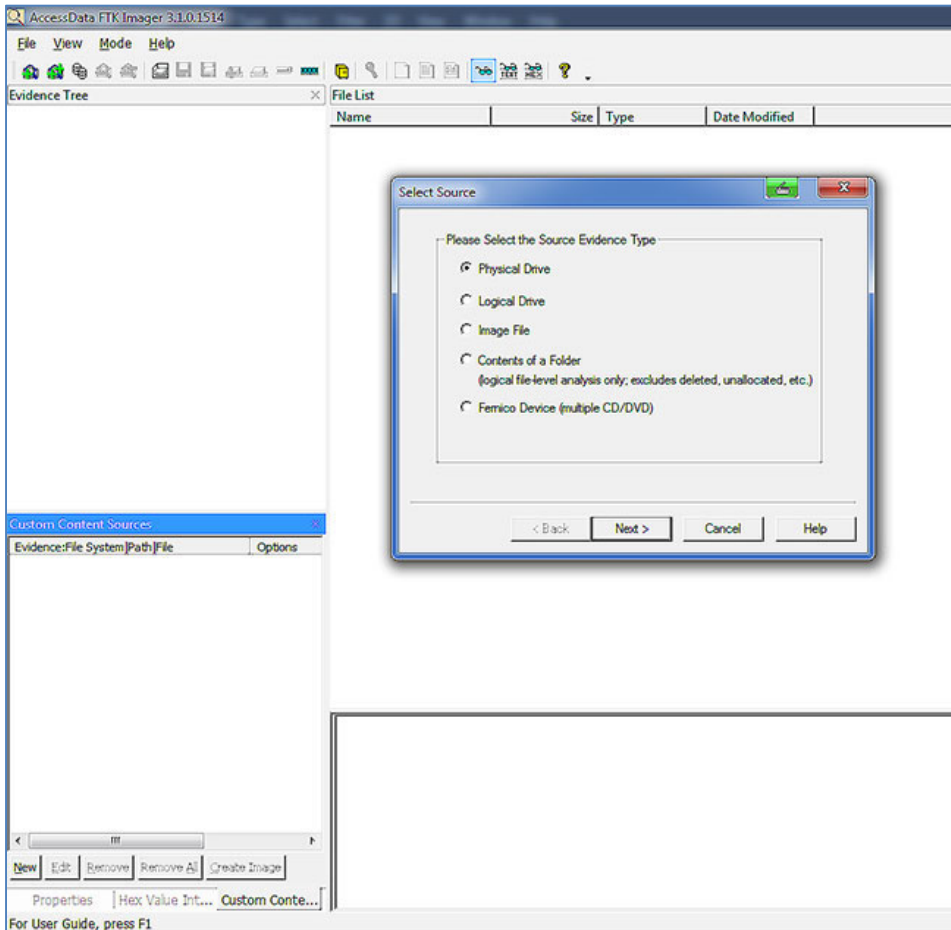
a. **Figure 15 FTK Imager Interface; Select Source option**

4. Click Next.

5. Select Drive:

   Click on the down arrow in the box immediately below the text "Please select from the following available drives:" Select the drive which corresponds to the drive you'd like to image. If the device is attached to the Tableau UltraBay 3D, it will likely be \\PHYSICALDRVE7.

6. Click Finish.

7. Create Image:

Below the Image Destination(s) box, select the "Add…" button.

8. Select Image Type:

Select AFF

9. Evidence Item Information:

In this dialog box enter disk image capture metadata/provenance information.

    a. Case Number: If University Archives or some other unit has provided a record series number or unique identifier, use it. Otherwise, use a collection name – an arbitrary name provided in-house to reference the collection; e.g. Michael Hart collection

    b. Evidence Number: internally-generated unique Item Identifier; this ID is generated earlier in the disk migration workflow as part of creating the reformatting directory structure:

    e.g. AHX_0000009.

    If this item is part of a series grouped together within a single Item Identifier, append an underscore and a corresponding sequential number to the end of the Item Identifier. If number is a single digit pad the number with one leading zero; .e.g. if this is the second disk associated with AHX_0000009 append _02 to the end of the evidence number: AHX_0000009_02.

    c. Unique Description: blank

    d. Examiner: enter your name in the last name, first name format: e.g Popp, Tracy

    e. Notes: briefly document source media type here; use terms from the following list:

1. internal hard drive – IDE

   internal hard drive – SCSI

   internal hard drive – SATA

   internal hard drive – 2.5"/1.8" form factor

   external hard drive – USB

   external hard drive – fire wire/Mac

2. flash media – (enter card type)

   floppy disk – 3.5"

3. optical media – CD

   optical media – DVD

   optical media – Blu-Ray

   ZIP disk

10. After entering Evidence Information click next.

11. Image Destination Folder:

    Navigate to the ImageCapturedFromDisk directory which corresponds to this item. The

    directory should have been created earlier in the workflow.

12. Image Filename:

    Give the file the same name as the Evidence Number, which is the same as Item

    Identifier and series number; e.g. given the example in step 9b the filename would be:

    AHX_0000009_02

13. Image Fragment Size (MB):

    a. Enter 0

14. Compression: 4

15. Select Finish.

16. If not selected, check the following boxes:

   Verify images after they are created – checks image hash signatures; will generate a

   hash for the image if there is not one

   Create directory listings of all files in the image after they are created – records the

   file names and paths of the image contents; saved in tab-separated value (.tsv) format.

   Precalculate Progress Statistics – provides imaging time and storage space required

17. Click Start. Let the imaging process complete.

18. Once imaging is complete and verified, close FTK Imager status windows.

19. **WARNING: turn the UltraBay 3D OFF before removing the source media.** Remove

   source media, and return it to the storage space designated for its corresponding

   collection.

20. **Remove the signal and power cables from the drive. If finished imaging return them**

   to the appropriate storage location.

21. Proceed to the document **PROCEDURE: Extracting Files from Disk Image Using**

   FTK Imager (version 3.1.0.1514)

PROCEDURE: Extracting Files from Disk Image Using FTK Imager (version 3.1.0.1514)

OUTPUTS FROM THIS PROCEDURE:

- **Image directory list .csv file into ObjectDocumentation directory**

- **Files exported from forensic disk image to OriginalFiles directory**

- **Hash listing (MD5 and SHA1 checksum) .csv file into ObjectDocumentation directory**

About FTK Imager:

FTK Imager is a free tool developed by AccessData for data preview and storage device forensic imaging. FTK Imager can also generate hash reports for files and disk images (including the files inside of disk images.)

In this step FTK Imager is used to export files and folders from forensic images for further curatorial work. Exporting files from the image allows access one to examine and manipulate a duplicate set of files while leaving the original files and forensic image intact. This set of exported files and folders can be considered the access copy or working copy of the data.

1. Select File > Add Evidence Item

2. Select Source: Image File

3. Browse to the image file location; typically images will be found within the **ImagesCreatedFromDisk** directory which was created when the disk image was generated.

4. Once the file is selected click "Finish". The image file structure is loaded into the Evidence Tree.

5. Export Image Directory List:

Once the image is loaded into the Evidence Tree right-click the top-most level (the image name). Select Export Directory Listing. This provides a full directory \ file listing of all data structures on the disk including unallocated space.

Fields included are:

    a. Filename,

    b. Full Path,

    c. Size (bytes),

    d. Created, Modified,

    e. Accessed, and

    f. Is Deleted

Fields will be populated based on metadata availability.
This directory listing is comprehensive than the content typically delivered to the requesting party as unallocated space or drive geometry pointers stored outside of a major data partition would likely not be transferred unless there is a request to do so.

6. Navigate to the digital object's "ObjectDocumentation" directory. Save the list here.

7. Name the directory listing after the image name, then, separated by an underscore, append dir List.

For example, if exporting from the 001_HartMi_IDE image, the dir list .csv file would be named: **001_HartMi_IDE_dirList.csv.**

8. Identify the directories\files you'd like to export. Generally, I export the top-most folder one level below the partition\file system level on the data-containing partition or partitions. FTK Imager will export all sub-files and directories within the folder.

9. Right click on the file \ directory level you want to export. Select "Export Files…"
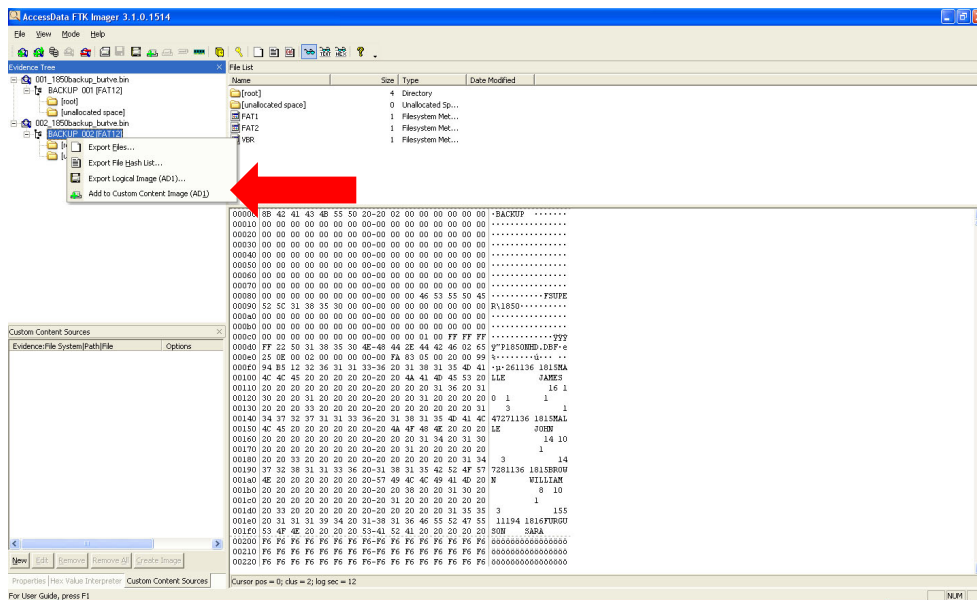

Figure 16: FTK Imager interface; right-click context menu shown in evidence tree.

10. Navigate to the **OriginalFiles** directory which corresponds to the image files from which you're extracting content.

11. Select "OK". You should see a progress bar indicating the file export status. Depending upon the number of files being exported this could take anywhere from a few to 10-15 minutes.

12. **Export File Hash List:**

    After the file export has completed export a file hash list by right clicking on the same directory \ file you previously exported and selecting "Export File Hash List…." The hash list .csv file contains **MD5 checksum** info, **SHA1 checksum** info and **corresponding file name** provided with complete path from the image.

13. Name the file hash list after the image name, then, separated by an underscore, add the top-most directory\file from which you're exporting, and append hash list.

    For example, if exporting from a directory structure called NONAME [FAT32] from the

001_HartMi_IDE image, the hash list .csv file would be named:

001_HartMi_IDE_NoName_FAT32_hashList.csv.

14. Export the file to the ObjectDocumentation directory corresponding to the digital

object\image you're currently working with.

15. Go back into the FTK Imager interface. Right-click on the image representation in the

"Evidence Tree" and select remove evidence item.

Note: If you attempt to mount a disk image which was derived from a disk formatted in an older

Mac operating system (pre-OSX; HFS) and are unsuccessful, re-try mounting the image as a

physical drive using the following steps:

1. In FTK Imager go to File > Image Mounting

2. In the Image File field: browse to disk image

    a. Select Mount Type: Physical Only

    b. Drive Letter: leave default value or value will be greyed out

    c. Mount Method: Block Device / Read Only

3. Select close

4. Mount the image from the physical share you've just created.

5. In FTK Imager go to File > Add Evidence Item

6. Select the drive that represents the image file you've just mounted.

7. Select Finish

8. In the Evidence Tree navigate down the tree to content found on the disk; it probably will

    be the disk name with HFS in brackets [HFS]

9. Export files, directory list and hash list as you would with other disk images using FTK.

10. After files are exported, run anti-virus scan on the files. If any virus activity is discovered, or viruses or malware are found, remove them. Document viruses removed or found in the migrationNotes.txt file in the "ObjectDocumentation" directory. If this text file does not exist, create it.

PROCEDURE: Receiving Vendor Reformatted Content

OUTPUTS FROM THIS PROCEDURE:

- **Original storage media**
- **reformatted content received from vendor (if successful transfer)**
- **Preservation metadata**
- **Content and transfer reports**
- **Original storage media disk image (optional)**
- **Photos of original storage media (optional)**
- **Complete SIP transferred to originating library unit (optional)**

Required Hardware:

- **Storage media containing reformatted content returned from vendor**
- **Non-networked computer system**
- **Second external hard drive**

1. Verify that media and all items shipped to vendor have been returned.

2. Verify that updated cost, invoice and purchase order information has been entered into the FY**Pres-Cons budget.xls.

3. Print corresponding purchase order; staple to invoice and file in "vendor invoices" vertical file.

4. Load the storage media returned from the vendor containing reformatted content on to the non-networked digital forensics machine

5. Run virus scan.

6. If viruses or malware are found, remove them. Document viruses removed or found in the migrationNotes.txt file. This file will be created in step eight as the virus scan is running.

7. While the virus scan is running, ensure that the directory structure corresponding to the digital object has been created on an external hard drive in preparation for transfer from the non-networked forensic machine to pre-determined end deposit location. The

directory structure should have been created when you first received the media from the requesting library unit.

8. Create a migrationNotes.txt file. Document that the item was sent a vendor for reformatting, note which vendor did the work, what they did, when the work was done and any errors or issues reported during data recovery. Place note in the "ObjectDocumentation" folder on the external drive corresponding to the original storage medium.

9. QA content received from vendor:

    a. Have the expected deliverables been returned? If not, contact the vendor.

    b. Check file dates. Have dates appeared to have changed to date of transfer? If so, make a note within the migrationNotes.txt file indicating that dates may reflect date of transfer not the true file creation date.

10. Transfer vendor reformatted content to the "originalFiles" directory within the appropriate directory structure on external drive attached to non-networked forensics workstation using TeraCopy; verify or generate checksums. If vendor has generated other deliverables including metadata, forensic image or photograph of the original storage media transfer them to the appropriate directory as follows:

    a. Metadata, reports, or other documentation > ObjectDocumentation

    b. Forensic image > ImageCapturedFromDisk

    c. Photograph of original storage media > ObjectJpg

11. Continue on to **PROCEDURE: Generating File Metadata Analysis Reports Using TreeSize Pro (version 5.5.5.816)**

PROCEDURE: Generating File Metadata and Reports Using TreeSize Pro (version 5.5.5.816)

**OUTPUTS FROM THIS PROCEDURE:**

- **TreeMap.jpg**
- **AllFileExtensionsReport.xls**
- **NoFileExtensionReport.xls**
- **Other reports such as duplicate files report as warranted**
  - **Item-level duplicate report**
  - **Collection-level duplicate report (compared to items)**

1. Launch TreeSize Pro.

2. From the toolbar browse to the reformatted digital object's "OriginalFiles" directory. Click on "Select folder". TreeSize Pro begins the scanning process from this folder.

3. Once scanning is complete you should see the file structure represented on the left side of the screen. File and directory navigation is largely done on the left side of the interface; finer detailed analysis and reporting is primarily completed on the center to right panel of the program interface.

4. In the left-hand tree navigation window, right click on the parent/highest directory. Select "Expand > Full Expand".

5. Make sure that the top-most level is selected.

6. In the reporting interface select the "Chart" tab. Click on the third icon – the tree map which displays a hierarchical chart for the currently selected folder.

7. Save the chart (ctrl+s or click on the floppy disk icon in the tool bar). Save the file in the MediaGroupDocumentation folder using the following naming convention:

Today's date:

YYYY_MM_DD_CollSu_ObjectNumberStart_ObjectNumberEnd_TreeMap. Select the .jpg file extension. e.g. 2013_03_13_HartMi_AHX000006_AHX0000022_TreeMap.jpg

8. Click on the Extensions tab.
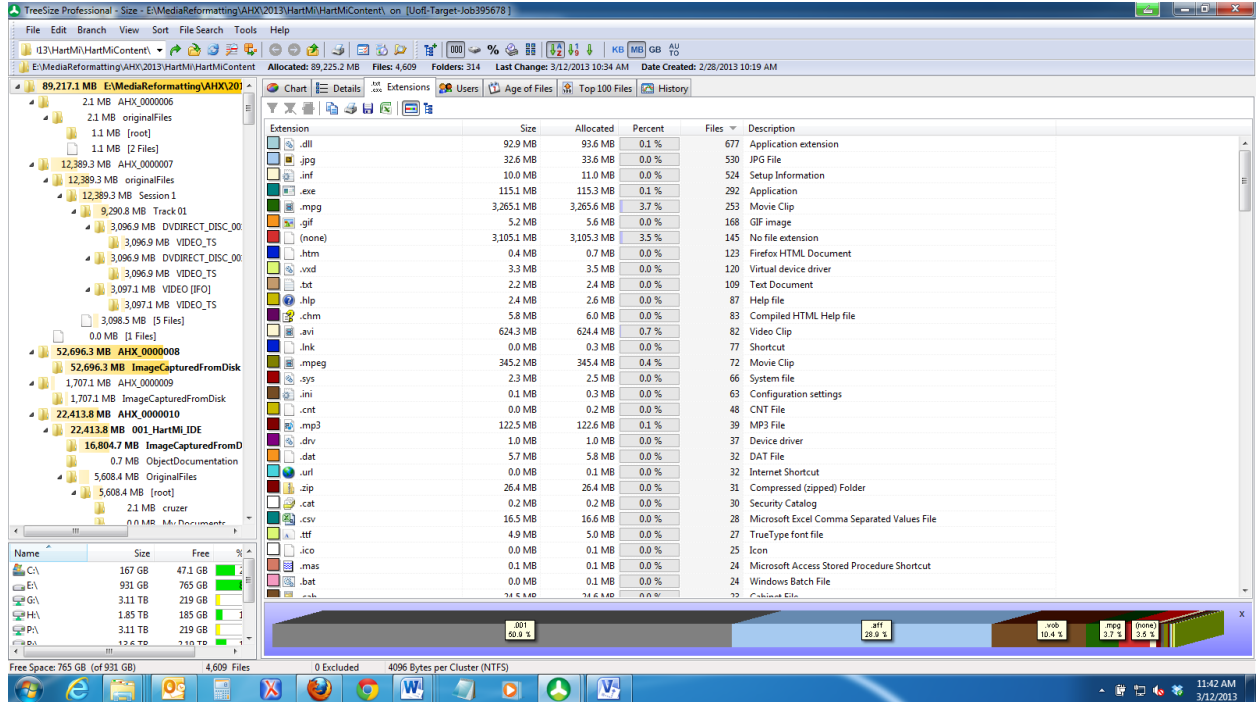
9. Click on the Files column; arrange from largest number of files to smallest.

10. From the toolbar click on the Excel icon. Export the file to the

    MediaGroupDocumentation folder using the following naming convention:

    YYYY_MM_DD_CollSu_ObjectNumberStart_ObjectNumberEnd_

    AllFileExtensionsReport.xls

    e.g. 2013_03_13_HartMi_AHX_0000006_AHX0000022_AllFileExtensionsReport.xls

11. The following fields are output in the report:

    a. Extension (the files' extension e.g. .txt)

    b. Size (the size of the files corresponding to a given extension)

    c. Allocated (the size on disk of the files corresponding to a given extension)

    d. Percent (percent of overall storage files represent)

    e. Files (number of files per type)

f. Description (basic description of the file type)

12. In the TreeSize extensions detail pane look for the (none) file extension; right click on

(none).
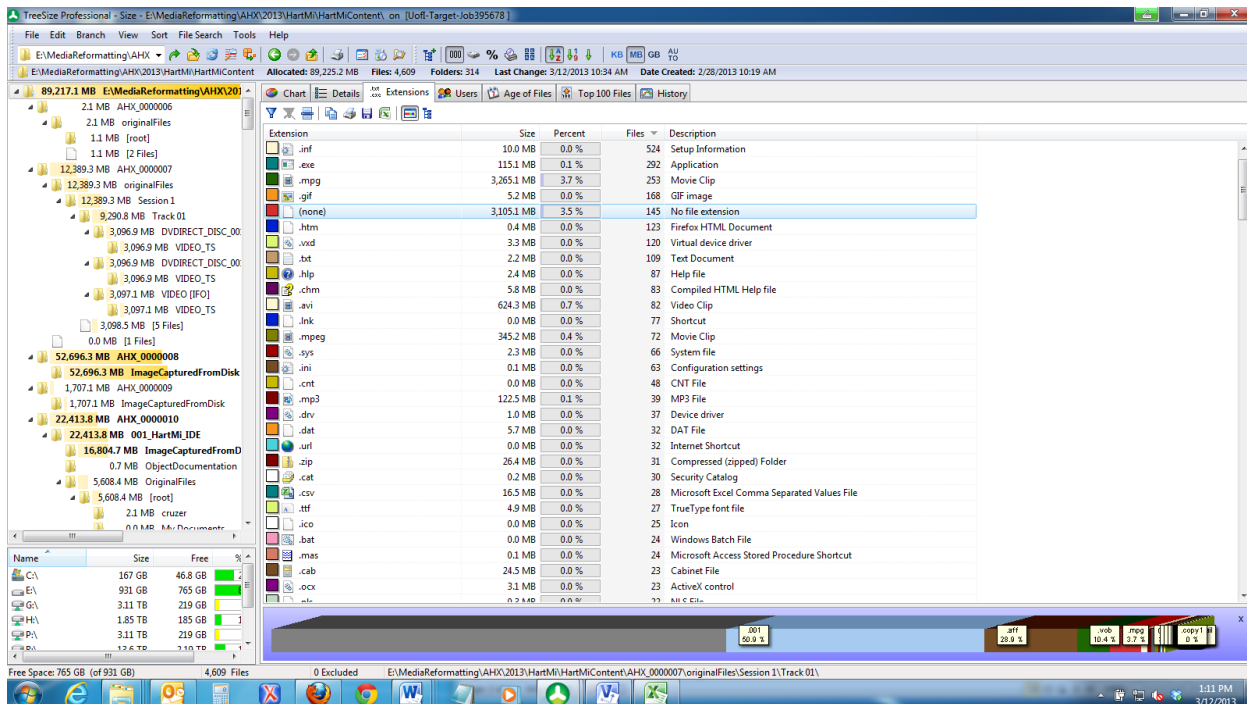
13. Select "Show all files of the select extension(s)". A new window will open.

14. Once the search has completed (let it run long enough to generate checksums) select the

"Save results" button on the far right side of the interface. . Export the file to the

MediaGroupDocumentation folder using the following naming convention:

YYYY_MM_DD_CollSu_ObjectNumberStart_ObjectNumberEnd_NoneFileExtensionR

eport.xls

2013_03_13_HartMi_AHX_0000006_AHX0000022_NoneFileExtensionReport.xls. This

provides a report of all of the files without a file extension and contains the following

fields:

a. Name (the file name)

b. Size (the size of the files corresponding to a given extension)

c. Last Change (the date the file was last changed if available)

d. Last Access (the date the file was last accessed if available)

e. File Type (The file type, e.g. "Text file")

f. Creation Date (the date the file was created)

g. File Version (The version number that is included in EXE, DLL OCX and similar binary files)

h. Path (the path to the file)

i. Extension (will be blank since there is no file extension)

j. MD5 Checksum (calculated MD5 checksum)

k. SHA256 Checksum (calculated SHA256 checksum)

15. To search for duplicate files, from the interface select File Search > Duplicate Files. A separate window will open.

16. Ensure that the "Compare using MD5-Checksum (more accurate) option is selected) in the search options. Depending upon the number of files this search could take a while.

17. Export report to SubjectMediaGroupDocumentation folder using the following naming convention:

YYYY_MM_DD_CollSu_ObjectNumberStart_ObjectNumberEnd_DuplicateFileReport. xls. This creates a report detailing all duplicate files found within a specified range and contains the following fields:

a. Name (the file name)

b. Path (the path to the file)

c. Size (the file size)

d. Last Change (the date the file was last changed if available)

e. Last Access (the date the file was last accessed if available)

f. Creation Date (date file was created)

g. MD5 Checksum

h. SHA256 Checksum

INFORMATION: Mapping Preservation Output to Archives to Medusa SIP
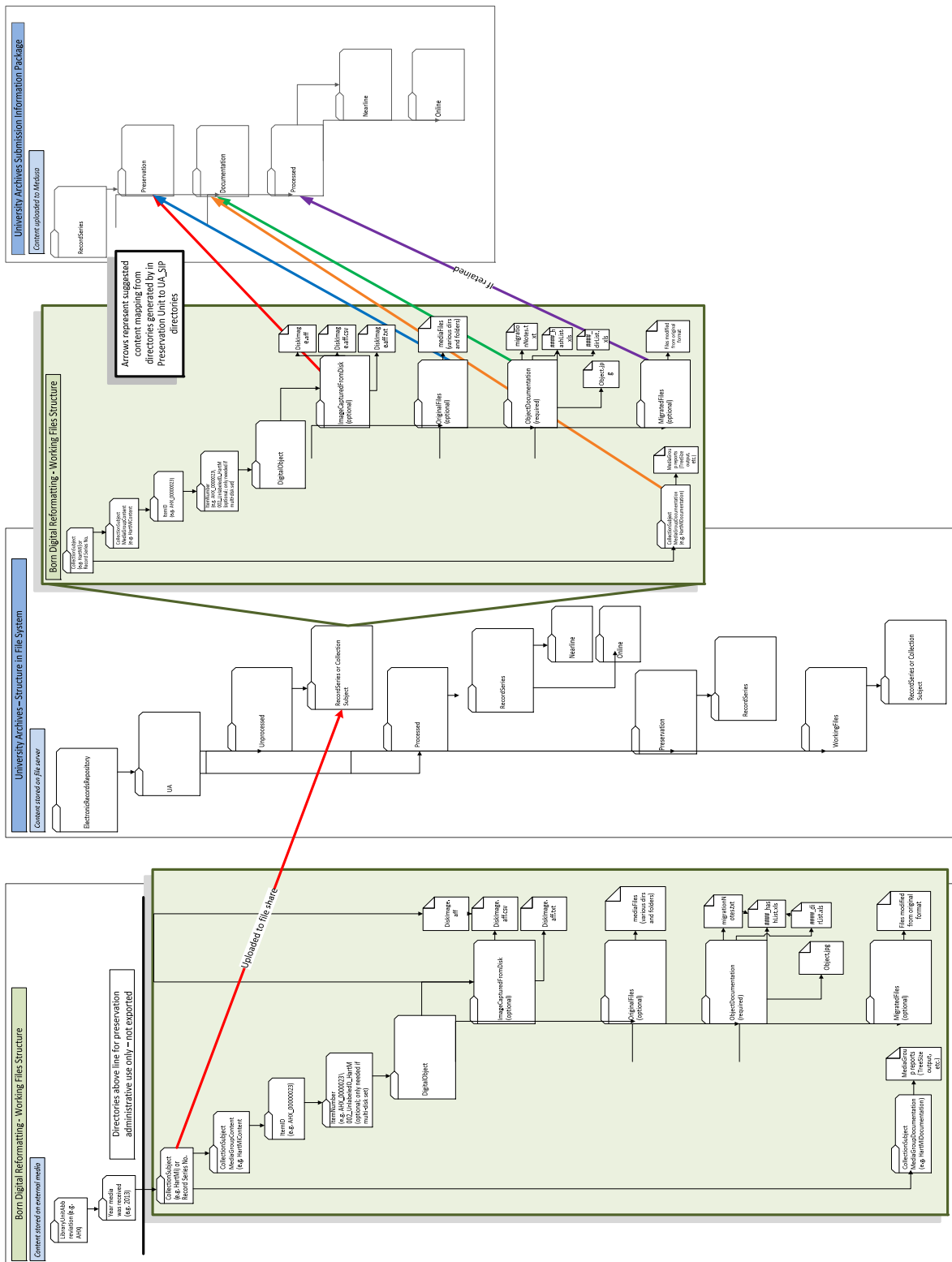
GUIDELINES: Retention of Original Storage Media

After recovering the informational content from storage media a decision must be made about retaining the carrier itself. The decision to retain or discard original, source storage media post-data-extraction lies with the collection manager, according to his or her own collection policies and professional judgment. These guidelines are intended to assist collection managers in making the decision to retain or dispose of original media carriers.

Understanding the Reformatting Process:

Reformatting analog materials for access and preservation generally includes generating a digital surrogate; a digital representative of the analog content. In some cases, the digital surrogate may not be of the same quality as the original analog item due in part to any number of technological and resource limitations and compromises. A strong argument for retention of the source analog object is that the digital surrogate is often not an exact representation of the source object; a researcher may wish to interact with the original object in order to ascertain evidentiary qualities not apparent in a digital surrogate.

In born digital reformatting there is often no degradation in quality from the source object to the reformatted object; the ones and zeros which represent the source object should be identical in the reformatted object. When storage media has been forensically imaged - a process where the entire storage object has been exactly replicated through a bit-by-bit copy process including unused storage blocks, disk drive metadata, disk drive structure and contents, and checksum files generated to verify the image reflects the source - the disk image is a "mirror image" of the original storage media. However, where the difference lies is in presentation – is the reformatted content experienced in a manner which is representative of the original? Do modern computing environments alter the presentation and experience in a way that significantly

affects researcher expectation? Is there a case to be made for retaining the original storage media and computing environment on which it is dependent?

Digital Media Fragility:

Compared to most analog information carriers such as paper, digital media are fragile and short-lived. Digital carriers require a machine intermediary to interpret the intellectual contents held on the media. These contents are also dependent upon particular hardware and software environments which render the contents human-readable. Any of these points within the computing ecosystem on which the content depends is prone to rapid obsolescence. The storage media itself is prone to mechanical and physical damage that is not easily detected through visual inspection and which can occur rapidly; relatively small amounts of damage can in turn result in catastrophic failure or irrecoverable data loss.[25]

Given than there is often little, if any, difference between the bits from the source and surrogate output, and the fragility of the physical media - especially unmonitored media – a case may be made to deaccession media after reformatting. However, the decision should not be made without giving sufficient consideration to the potential research value of the physical item and potential for engaging with the original.

Photo Documentation:

Regardless of the original physical object's fate, in most cases photo documentation of the object will be part of the born digital reformatting workflow. Photo documentation is especially important when the item contains hand-written labeling information, ancillary materials and if the item is slated for deaccessioning.

Objects Recommended for Indefinite Retention:

---

[25] http://www.digital-preservation.com/challenges/media-obsolescence/

Object has Artifactual Value:

An obvious case for retention is when the physical object has intrinsic value and is valuable as an artifact. According to a report authored by National Archives and Records Administration's Committee on Intrinsic Value, an analog (physical) object has intrinsic value if it has one or more of the following qualities:

- Physical *form that may be the subject for study if the records provide meaningful documentation or significant examples of the form*

- Aesthetic *or artistic quality*

- Unique *or curious physical features*

- Age *that provides a quality of uniqueness*

- Value *for use in exhibits*

- Questionable *authenticity, date, author, or other characteristic that is significant and ascertainable by physical examination*

- General *and substantial public interest because of direct association with famous or historically significant people, places, things, issues, or events*

- Significance *as documentation of the establishment or continuing legal basis of an agency or institution*

- Significance *as documentation of the formulation of policy at the highest executive levels when the policy has significance and broad effect throughout or beyond the agency or institution*[26]

Content Reformatting was Unsuccessful or Incomplete:

---

[26] http://www.archives.gov/research/alic/reference/archives-resources/archival-material-intrinsic-value.html

Libraries and archives may consider developing a retention policy for unique damaged media whose contents are partially or completely irretrievable.[27] Technologies may develop in the future that allow for the retrieval of data from previously unrecoverable media. One example of an institution advocating for the retention of media where contents may not have been successfully transferred – due to factors including unavailability of transfer equipment or media damage - is the Albert and Shirley Small Special Collections library at the University of Virginia.

Recommended Retention Period:

A collection manager may wish to retain media and systems for a temporary period. One example may be retaining the media until the Library's digital preservation repository has been well-tested and functioning in production mode. Although content uploaded to Medusa at this time is replicated and verified a safeguard against a worse-case scenario of data loss would be to return to the original carriers. Although data on the Medusa system is replicated retaining the original carriers for a temporary time period could provide an additional option for data recovery.

Media Retention - Rehousing and Storage Conditions Considerations:

Once the decision to retain the digital carrier has been made ensuring that the physical media is preserved as long as possible becomes a concern. One of the key reasons for migrating content from the carriers is due to the relatively short life-spans of digital physical media. Dust and other contaminants, environmental factors such as heat and humidity, magnetic fields, static electricity, physical damage and inherent vice can each drastically reduce carrier longevity. In general, digital media should be:

---

[27] http://www.archives.gov/research/alic/reference/archives-resources/archival-material-intrinsic-value.html

- Encased in a protective covering that shields the item from dust and other contaminants

- Encased in a manner which prevents or mitigates physical damage such as a placing an internal drive in a hard plastic case

- Handled in a manner that avoids physical damage including:

    o Limiting playback or use

    o Limit media access to properly trained staff

    o Avoid touching exposed media surfaces; do not open disk shutters or touch the surface of optical media

    o Careful handling of media; avoid dropping, especially hard drives and computer systems

- Maintained in an environment with consistent temperature and humidity – ideally, 68 degrees Fahrenheit and 40% relative humidity; avoid temperature and humidity cycling as much as possible

- Avoid exposure to magnetic fields – this can cause data loss or data corruption

- Shelve media vertically; do not stack media

Careful maintenance of the media may lead to increased life-expectancy; however, collection managers should not rely on digital physical media and should seek content stabilization. Solely rehousing born digital media is not a sufficient long-term content preservation strategy; retaining the physical media is often only part of preservation process and can engender a false sense of security if the content is not stabilized.

Media Enclosures:

The Born Digital Content Preservation service is able to provide consumer-grade replacement enclosures for storing the following formats:

- 3.5" internal hard drives

- 2.5" internal hard drives

- optical media such as data CDs and data DVDs

If a curator would like to maintain born digital media for long term preservation custom enclosures can be constructed by the Conservation Unit. If interested in discussing a custom enclosures project for born digital physical media contact the Digital Preservation Coordinator.

Deaccessioned Media:

Once the decision has been made to deaccession media there are disposal procedures which are intended to protect donors; specifically, content-carrying media, whether accessible or not, must be destroyed prior to disposal as a safeguard against sensitive data release. Consider this point carefully before making the decision to deaccession media as media and content will not be recoverable after destruction and disposal actions are taken.

Data Scrubbing Requirements and Media Disposal:

When considering discarding media collection managers are ethically obligated to adhere to any requirements established in acquisitions agreements and/or deeds of gift; has media retention or return to donor been agreed upon in either of these documents? If no specification has been made and the collection manager wishes to dispose of the media certain guidelines are followed to ensure ethical disposal of potentially sensitive or private information. University policy requires that prior to transferring or disposing of University-owned digital media, said media/devices must be sanitized in order to ensure that no sensitive data is revealed to unauthorized/inappropriate parties.

University Surplus provides media degaussing services which will "scramble" the magnetic particles and effectively removes data from the media. Surplus will also manage

disposal after degaussing. If the hard drive is still functional one can also scrub the disk prior to disposal. See this CITES guide on disk scrubbing: https://security.illinois.edu/content/disk-scrubbing.

If a collection manager wishes to have the Digital Preservation Coordinator manage the destruction and discarding of media after reformatting he or she must provide a written agreement indicating so before any action takes place. If the agreement is not received within sixty (60) days after reformatting has been completed the Digital Preservation Coordinator will return the media to the collection manager without taking any destructive action.

System Disposal

If the collection manager has acquired a computer system as part of an acquisition and no longer wishes to maintain the system after data transfer the system must also be disposed of in an appropriate manner.

U of I property:

If the system has arrived at the Library via another campus department, ensure that the originating department's property manager has transferred the system to the Library's inventory. The digital preservation coordinator can manage transfer of Library equipment to surplus AFTER given WRITTEN PERMISSION to do so by the collection manager. Again, once the data is scrubbed and the system is moved to surplus, this action cannot be undone.

Gifted System:

Ensure that the Deed of Gift allows for system disposal. If so, send a written request to the Digital Preservation Coordinator who will manage the disposal AFTER receiving WRITTEN PERMISSION to do so by the collection manager. Again, once the data is scrubbed and the system is moved to surplus, this action cannot be undone.
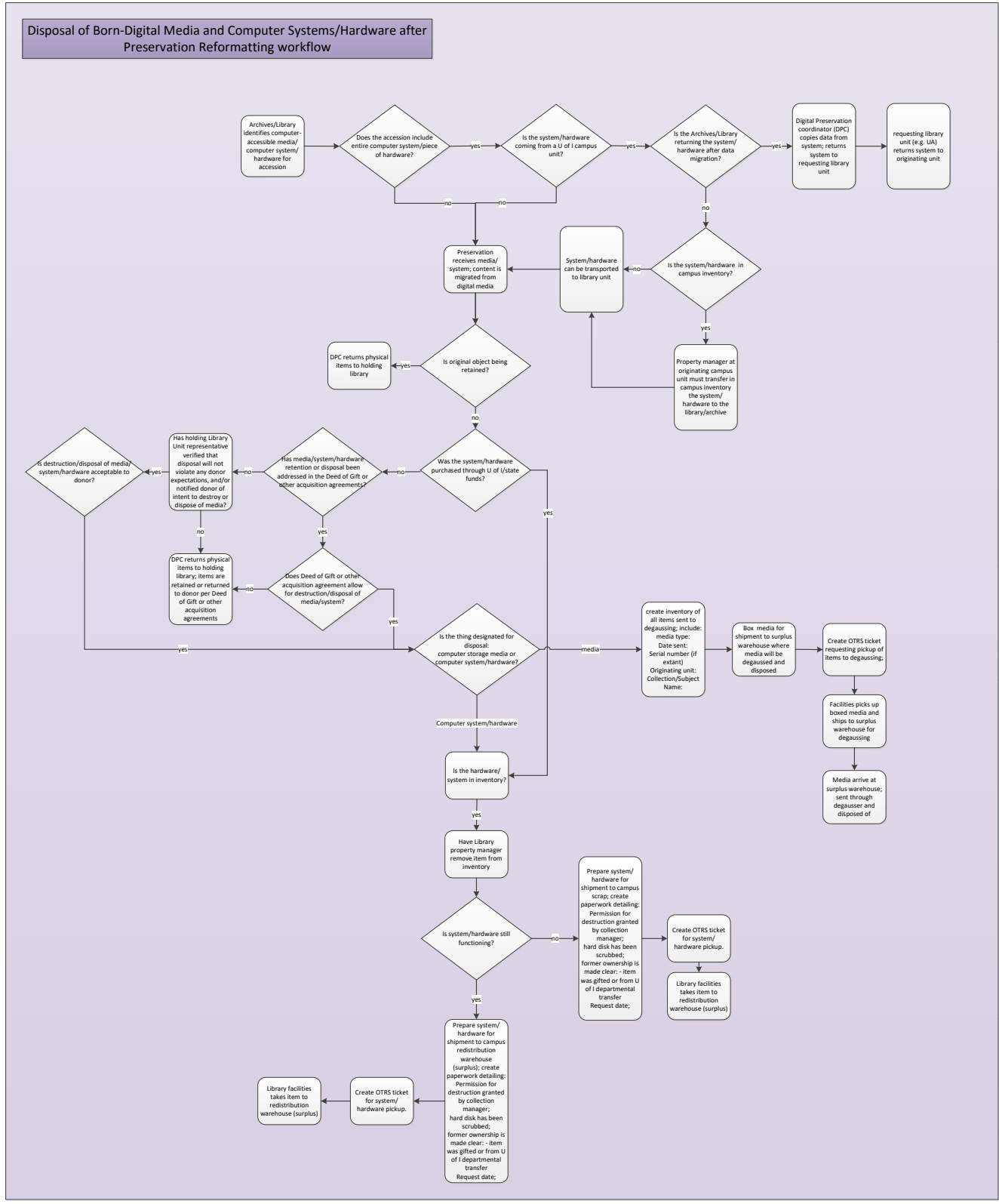
Figure 17: System and Media Disposal Workflow

Example Hardware Disposal notice

Disposal Type: Scrap Electronic

Requestor: Tracy Popp

Description: computer parts received as part of donation of Michael Hart materials; not in U of I inventory; permission to dispose of granted by donating party.

Date: 4/26/13

Hard Disk scrubbed? No hard disk to scrub.