Cornell University
ILR School

Practical Technology for Archives

Volume 1 | Issue 4 Article 4

5-2015

Getting to Know FRED: Introducing Workflows for Born-Digital Content

Alice Prael
University of Maryland

Amy Wickner
University of Maryland

Getting to Know FRED: Introducing Workflows for Born-Digital Content

Description

Special Collections and University Archives (SCUA) at the University of Maryland Libraries have accepted born-digital content from donors for the past ten years and these donations have grown exponentially over the last few years. Born-digital acquisitions will continue to grow as modern record-keeping moves to exclusively digital formats. Despite the volume of data acquired, almost none of it has been processed into the Libraries' digital collections for long-term preservation and access. Hard drives and other digital storage media have instead been processed like paper material and placed in boxes, often with printed copies of file inventories or of the digital contents. This can only be a temporary solution as digital storage is susceptible to degradation. As born-digital donations grow, so will the need to protect this data in long-term digital storage.

Keywords

born-digital content, Forensic Recovery of Evidence Device, FRED, protection, digital storage

Getting to Know FRED: Introducing Workflows for Born-Digital Content

Alice Prael
University of Maryland

Amy Wickner
University of Maryland

Born-digital at UMD

Special Collections and University Archives (SCUA) at the University of Maryland Libraries have accepted born-digital content from donors for the past ten years and these donations have grown exponentially over the last few years. Born-digital acquisitions will continue to grow as modern record-keeping moves to exclusively digital formats. Despite the volume of data acquired, almost none of it has been processed into the Libraries' digital collections for long-term preservation and access. Hard drives and other digital storage media have instead been processed like paper material and placed in boxes, often with printed copies of file inventories or of the digital contents. This can only be a temporary solution as digital storage is susceptible to degradation. As born-digital donations grow, so will the need to protect this data in long-term digital storage.

In 2012, the University of Maryland Libraries and the Maryland Institute for Technology in the Humanities (MITH) recognized these risks and created the Born-Digital Working Group to address these issues. The working group consisted of staff from SCUA, Digital Systems and Stewardship (DSS), and MITH. The Libraries acquired a Forensic Recovery of Evidence Device, commonly referred to as FRED, in March 2013.

FRED is used to create verifiable, high quality copies of digital media without risking damage to or alteration of the original media. In 2014, having determined that the project would require sustained, dedicated attention, members of the working group assigned DSS and SCUA graduate assistants the task of testing FRED's abilities and making recommendations on how to integrate born-digital content into a new workflow.

The graduate assistants produced the FRED Guide, a systematic beginner's guide to using FRED, which explains how to identify and connect digital storage media, acquire disk images, and analyze those images. They also outlined a workflow in which several software systems are integrated for the management of digital objects and storage media. These systems include BitCurator, a software environment for digital forensics in Libraries; ArchivesSpace, a system for archival description and management; and Fedora, a digital repository system.

The University's digital collections are currently in the process of migrating from Fedora 2.2 to the most recent Fedora 4. The entire collection will be migrated to the new system and the workflow will be reconsidered to include new programs for description and accessioning. In the midst of these large-scale changes in digital collections, the Libraries decided it was the right time to implement the processing of born-digital objects. Beginning with a practical, detailed guide to making digital forensics hardware and software decisions, we have proposed and continue to refine a workflow that integrates several new systems while remaining flexible in anticipation of ongoing development. We present here the results of our experimentation and our proposed workflow as well as future steps for testing and implementation.

FRED

Digital Intelligence created FRED in order to help users create high quality copies of digital media; these bit-for-bit copies are called disk images. FRED is equipped with hardware write-blockers and ventilated shelves for imaging internal hard drives. FRED is partitioned to include Linux and Windows operating systems in order to run both imaging software applications we tested. [i] We focused on two programs for disk imaging, BitCurator and the Forensic Toolkit (FTK) Imager.

The first tool we tested was the FTK Imager, which is the free imaging tool that is part of the larger proprietary Forensic Toolkit. Although the Imager does not have the full functionality of FTK, it provides enough disk imaging tools to fulfill our needs. [ii]

The second tool we considered was BitCurator, an open-source Linux-based environment that provides a suite of tools for disk imaging, analyzing, and reporting. [iii] The BitCurator project was co-led by researchers at the School of Information and Library Science at the University of North Carolina, Chapel Hill (SILS), and at MITH. [iv] Professional development and support for software created by the original BitCurator project has transitioned to a consortium of 16 institutions in cultural heritage and higher education. [v] With funding from the Andrew W. Mellon Foundation, the next project phase (BitCurator Access) will see the development of tools for access to forensically packaged disk images. [vi]

Media Types

We focused on seven storage media types: 3.5-inch floppy disks, 5.25-inch floppy disks, CDs, DVDs, portable storage with USB connections, and internal (or bare) hard drives. FRED is equipped to connect to and image hard drives, CDs, DVDs, and USB connected storage, but requires external hardware to access legacy media [vii].

We researched and experimented with three tools to connect floppy disks, working exclusively with MS-DOS floppy disks, a type commonly found in SCUA’s collections. We first tested a 3.5-inch floppy drive with a USB connection. Once connected, disks’ file directories were read with a file manager, although file contents were not necessarily accessible. Next, we tested FC5025, which uses a controller card to connect a 5.25-inch floppy drive through a USB port. While the FC5025 graphic user interface (GUI) was accessible on Linux and Windows, we were more successful running it in Windows. We were also able to image a floppy disk with FC5025 and access its contents through FTK Imager [viii], although the output required additional tools for access and analysis.

The third tool we tested, SuperCard Pro (SCP), was designed to work with both 3.5- and 5.25-inch floppy disks. According to the manual, 8-inch floppy disks are also manageable using a SCP function we did not test. Like FC5025, SCP was also boxed with a tool for creating and accessing disk images. We were able to image one side of a 5.25-inch floppy disk with SCP but access was an issue: Images were only viewable as hex code in the SCP GUI, and we could not access or mount SCP disk images with other image mounting tools. We were able to image the majority of floppy disks through a combination of the USB-connected floppy drive and FC5025 so we decided not to use SCP [ix]. Ultimately, we recommend using the USB-connected drive for 3.5-inch floppy disks and FC5025 for 5.25-inch disks.

We discovered problems with our initial set-up in early attempts to image double-sided 5.25-inch floppy disks. Neither the Windows nor Ubuntu partitions recognized the connected drive. After testing the same batch of disks at MITH's digital forensics workstation, we determined that drive orientation was to blame. Although MITH's drive could read information from either side of a double-sided floppy disk, the drive in Special Collections only recognized information from one side of a disk, and only when the drive was oriented upright. We had placed the drive upside-down to avoid collecting dust. Going forward, imaging 5.25-inch floppy disks will require replacement equipment to identify and read double-sided disks.

Although modern carrier media posed fewer challenges, we faced some issues in disk imaging CDs and hard drives. CDs and DVDs have a variety of read/write settings so it is important that write-blockers are always in place before connecting the media. We discovered a bug in which FRED's CD/DVD tray did not open while BitCurator's software write-blocker was on. FRED is equipped with hardware write-blockers so we were able to turn off the software write-blocker and continue as normal. However, this could be a larger problem for other projects in which a hardware write-blocker is not involved.

In testing external hard drives, we discovered that BitCurator did not support the exFAT file system, which was designed for read-write interoperability between Windows and Mac OS X computers. Drives formatted as exFAT file systems–whether by manufacturers or users–were unrecognized by BitCurator and required imaging with FTK Imager.

Learning to image internal hard drives resulted in a different kind of challenge: the risk of damaging archival material before processing. Connecting an internal bare hard drive with a Serial ATA (SATA) connector was a matter of mechanics: plugging the drive into a plastic tray via power and data connectors, sliding the tray back into FRED, closing a latch, and turning a key to power up the drive. The reverse set of steps was followed to remove the drive. Timing the steps was as much of a challenge as following the correct order. In an early effort to image a SATA internal hard drive, damage to the drive resulted from unlocking and removing the drive tray before the drive had completely spun down. The resulting mechanical damage made data from the drive unrecoverable, and donors were contacted to replace the material. This experience emphasized the need for clear instructions about timing in FRED procedures; the reality of physical risks outside of media obsolescence; and the importance of transparency and communication between repositories and donors when it comes to nascent or experimental programs.

We anticipate working primarily with digital files on hard drives for the near future. Based on a preliminary survey of 85 collections (4,716 linear feet) within a relatively small collection area, existing born-digital material in SCUA includes legacy storage media like floppy disks, CDs, and even stacks of magnetic tape. [x] The Born-Digital Working Group produced a matrix of supported file formats and carrier media (see Appendix A) that will guide born-digital and mixed media appraisal decisions going forward: “Material in unsupported formats or media may not be accepted.”

Alternatively, imaging may be outsourced to vendors or, as recommended in Erway (2012) [xi], to a network of institutions with different digital forensics capabilities. This workflow may also evolve in response to new tools, new storage technology, and new capacity for working with born-digital objects. It is vital that we remain open to new possibilities.

Image Types

We plan to image all media in the aforementioned seven categories to preserve the contents via the Libraries' workflow. Based on characteristics of the storage media and circumstances of selection and transfer, there may be reason to acquire either a physical or a logical image. A physical (or forensic) image is a bit-for-bit copy of the medium, while a logical image acquires information in logical volume of media, at the file directory level. Kirschenbaum, Ovenden, and Redwine (2010) have made the definitive case for the value of digital forensics methods in documenting not only born-digital cultural heritage content but also the circumstances of its creation. [xii] Gathering forensic evidence is not always appropriate or possible when working with electronic records, or with born-digital content acquired by the Libraries before developing a clear work plan.

We must consider multiple factors when selecting an image type. Current accessioned but unprocessed digital material includes several internal and external hard drives purchased by donors and used exclusively as transfer media. The forensic value of such a device is negligible; producing a logical rather than physical (or forensic) image of the entire drive may be the right choice here. In our tests, one of the drives contained the entire desktop, dragged and dropped with little discrimination among files and folders. In this instance, it was appropriate to produce a logical image of specific files and folders or separate logical images of each desktop. However, logical images at the folder level must be created in a proprietary format and may be of limited use relative to forensic disk image formats no matter the depth of information required. The judgment of collection area archivists is essential in identifying areas of interest from a top-level inventory of the media contents, just as involving digital archivists in future appraisal discussions may help the Libraries select appropriate digital material to accession. Finally, donors may see access to deleted files as a privacy concern and may not wish their media to be forensically imaged.

Imaging Tools

We tested two disk imaging tools: Forensic Tool Kit (FTK) Imager, which is proprietary and runs in a Windows environment; and Guymager, an open-source Python script and GUI included with the BitCurator environment. [xiii]

Although both tools had similar forensic imaging capabilities, FTK Imager offered additional options that are useful in circumstances when a forensic image is not appropriate. Guymager could only image forensically or physically, while FTK Imager also acquired logical disk images and had the ability to acquire a disk image of an individual folder. Both tools included the option to segment disk images acquired from large volumes, for ease of transfer. The ability to write large disk images to shared storage is an ongoing challenge for this project. One potential advantage to Guymager was its up-front notification when a storage location had insufficient space.

FTK Imager offered a built-in option to create Custom Content Images: collections of material selected from multiple places in one or more evidence items. Possible uses for such a feature included the ability to transfer select material from a donor’s working storage media; combining multiple, related folders from across storage media into a single package; and other creative solutions for determining the scope of an accession. FTK Imager acquires logical and Custom Content Images in a proprietary file format (AD1) that provides limited metadata relative to true disk images. Use of this format has been deprecated for long-term digital preservation (Woods, 2014). The decision to image digital media in AD1 has downstream implications for how the contents of disk images are extracted for preservation.

The relative benefits of FTK Imager and Guymager were tested in imaging various media. FTK Imager proved to be faster at acquiring images of large storage media such as hard drives, by a matter of hours. Imaging 165 GB of data on a 1 TB external hard-drive with FTK Imager took 12 hours and 39 minutes for a forensic image and 10 hours and 39 minutes for a logical image. Acquiring a forensic disk image of the same drive with Guymager took 14 hours and 29 minutes. When the same test was repeated with 70 MB of data on a 2 GB flash drive, Guymager acquired a forensic disk image 14 seconds faster (1:42) than FTK Imager (1:56), which took the same amount of time to image the drive forensically and logically. Due to our anticipated imaging needs, the time difference involved in imaging large storage volumes was of more interest than the impact of scaling up imaging of multiple small storage volumes. For these reasons, we recommend creating forensic and logical disk images through FTK Imager rather than Guymager.

Disk Image Analysis

When making decisions about FRED, it was important to remember its context in a larger, in-development workflow. We know that FTK and BitCurator is part of that workflow, along with ArchivesSpace, Fedora 4, and Solr, other technologies currently employed by UMD Libraries. How this workflow is implemented affects what reporting tools are necessary and how the results are used.

While BitCurator came “boxed with” a large array of tools for forensic analysis, we were primarily interested in extracting directory structures and identifying personally identifiable information (PII). This was accomplished using Fiwalk, a tool for mapping the file directory of a disk image in DFXML (digital forensics XML); the Bulk Extractor GUI tool (or bulk_extractor.py script); and Annotated Features Reports, which triangulated the outputs of Bulk Extractor and Fiwalk to identify specific files containing privacy-sensitive information. Bulk Extractor scanned files for identifying information like social security numbers or email addresses. [xiv]

Both FTK and BitCurator supplied a directory of the files that were copied in the disk image. BitCurator used Fiwalk to create an XML file that included the filename, path, checksum in MD5 and SHA1. FTK created a CSV file that included the file name, path, size, and dates of creation, modification, and access. Although there were redundancies in keeping both directory listings, there were also advantages to having this information in more than one format. The CSV file was more easily read by people and could serve as a manifest for archivists, while the XML file created by Fiwalk provided more metadata when imported to ArchivesSpace. We recommend including both the CSV and XML files in the reports associated with a disk image.

Bulk Extractor's graphic interface arrived pre-loaded with filters for detecting social security numbers, email addresses, and phone numbers, among other types of sensitive information. Depending on the type of record, and based on conversations with donors, archivists may choose to create custom filters that screen for and capture sensitive information in other formats. Searching for specific names or URLs may also help determine what is appropriate to preserve and make accessible.

BitCurator's reporting tool also provided file format information in both a table and a bar graph, so archivists can understand the number and type of files at a glance. We discovered a bug that occurs when the reporting tool is run twice on the same disk image: Although the file types are consistent, the number of files changes each time this command is executed. We are in communication with the BitCurator team to determine the cause of the problem. Until this bug is resolved, we will rely on the Fiwalk XML for file format information.

Workflow Recommendations

Our workflow uses ArchivesSpace to track digital objects from the original donation to providing access copies. The following section demonstrates our proposed workflow and identifies each action performed in ArchivesSpace alongside the processing steps. A visual representation of our workflow is available in Appendix B.

These recommendations are based upon currently employed systems and technologies or proposed systems at UMD Libraries.

Work with donors

The combined expertise of collection area and digital projects archivists will help identify born-digital material of archival value and potential research interest. The reports of the AIMS Born-Digital Collections project provides guidance on procedures and checklists [xv], which has inspired SCUA staff to begin work on donor interview checklists of our own. When finalized, these checklists will be appended to existing templates for donor agreements.

Format-based decisions

We have developed a workflow specific to the digital storage media types manageable by SCUA given the systems available to the department. Upon identifying digital material of value, archivists will determine whether the storage medium is one of seven that can be processed by SCUA. Digital Conversion and Media Reformatting (DCMR), an in-house digitization lab working with SCUA and Special Collections in Performing Arts, is capable of processing a much wider variety of digital audiovisual media on legacy storage formats. DCMR may be enlisted as an alternative site for processing born-digital material if they further develop current services for the Libraries. Contracting work to outside vendors, with DCMR as a go-between, is a third option that archivists may consider, using existing agreements for digitization projects as a guide. Separate workflows exist or may be developed for processing born-digital material via DCMR or a vendor. If none of the above options are viable—for example, if prices quoted by vendors are too high—we recommend that SCUA not accept the digital media without additional donor support.

Discuss privacy issues with donor

In addition to interviewing the donor about his or her digital material and the context of creation of digital records, archivists will educate the donor about the scope of possibility that digital forensics introduces to archival processing. They will clearly and simply explain that deleted and hidden files are viewable under certain circumstances, and be ready to discuss resulting privacy concerns in an accessible way that inspires neither panic nor mistrust. Outcomes of this conversation will include identification of material for accession: what SCUA may accept for potential long-term preservation, and agreements regarding access restrictions for sensitive information.

Receive digital objects on storage media

Archivists will discuss transfer methods with the donor and choose an appropriate method based on the content, current formats, and information gleaned through interviews and site visits. Upon receiving digital material on storage media, archivists will create a new Accession record in ArchivesSpace as well as logging a Custody Transfer event.

Determine type of image

Conversation with the donor about privacy and access issues, as well as characteristics of the storage media, will inform the type of image it is appropriate to acquire.

Connect and image media

Archivists will follow instructions in the FRED Guide to connect and image media, writing disk images to shared storage on the Libraries' local access network. A .info file generated and stored alongside the image includes image checksums (MD5 and SHA1) and metadata for the imaging process. Capture and validation events will be logged in ArchivesSpace.

Virus check

Upon accessing the disk image in the BitCurator environment, archivists will first initiate a virus check using ClamAV, an open-source tool pre-installed in BitCurator. A virus check event will be logged in ArchivesSpace. Should any threats be identified, the workflow will be abandoned and the donation further assessed.

BitCurator analysis

If a virus check shows the disk image to be uncompromised, archivists will next analyze the image using BitCurator reporting tools. Of BitCurator's arsenal of forensic analysis tools, we recommend using the Bulk Extractor tool to screen for personally identifiable information (PII); producing DFXML (digital forensics XML) with Fiwalk, a file structure analysis tool; and running Annotated Features reports to connect PII to its location in the image's file contents using Bulk Extractor and Fiwalk results. Outputs from the reporting process will include a DFXML file indicating directory structure and contents, PDF reports on file formats and PII, and text files with location information for PII found through Bulk Extractor.

Generate Submission Information Package (SIP)

Archivists will package each disk image with its checksum and BitCurator report outputs using the BagIt specification. Each bag will constitute a Submission Information Package (SIP) as described in the Open Archival Information System (OAIS) model, with metadata in "sidecar storage."

Trusted storage

Once packaged, each SIP will be transferred to intermediate storage and an ingestion event will be logged in ArchivesSpace. The transition between SIP transfer and additional processing of the accession is a point at which significant time lapse or backlog is likely to occur. Academic Preservation Trust (APTrust), a higher education consortium for collaborative digital preservation services of which the University of

Maryland is a member, is one possible candidate to provide this storage location. In addition to providing necessary preservation services, including bit-level preservation, checksums, preservation action logging, standardized metadata, and security by geographic diversity [xvi], APTrust provides easy access by members to stored material. These factors have influenced our recommendation to store SIPs in APTrust until further processing can take place. Acknowledging that a backlog of minimally processed born-digital material is likely, we prefer the security of APTrust to that of a working folder on the library network.

Assign preservation disk image UUID

Bagged SIPs will be moved back into a staging area on the library LAN for further processing. Each disk image will be assigned a universally unique identifier (UUID). A new Resource record in ArchivesSpace will also record this UUID as the item's preservation copy identifier. Fedora 2.2 creates sequentially generated persistent identifiers (PIDs) for digital objects. Each object is assigned a unique PID formatted according to the convention umd:#####, while its pre-ingest file name is recorded as administrative metadata. A transition from the use of PIDs to the use of UUIDs is planned for the migration to Fedora 4.

Sensitive information and access restrictions

Examining BitCurator reports for personally identifiable information will pertain primarily to setting and refining conditions for access. If unexpected sensitive material is revealed, a follow-up conversation with the donor may be necessary to refine access limitations. Revising the scope of what SCUA may preserve over the long term will be another possibility at this point, albeit a less than ideal solution that would needlessly tax library resources.

Identify material for access

Identifying material for access will require archivists to answer a number of questions: What access limitations are encoded in the donor agreement? How does the discovery of additional PII affect those restrictions? What access mechanisms are currently available to University of Maryland library users, and what mechanisms do we anticipate employing in the future? Should access to born-digital objects be provided item by item, or in aggregate? The answers to these questions will vary by collection and format, and will likely change over time.

Enhance metadata

Upon identifying material for preservation and access, archivists will evaluate and enhance metadata embedded in or stored alongside each disk image. Acquiring disk images of storage media will generate DFXML recording information about the imaging process. Additional description will be necessary to fully incorporate born-digital material into ArchivesSpace and Fedora for discovery and access.

Two "homegrown" metadata schemas, first released as part of a Best Practice Guidelines for Digital Collections document in 2007 [xvii], have been used to describe digital collections over the past several years. The University of Maryland Descriptive Metadata (UMDM) Tag Library documents required and optional descriptive metadata for digital objects, including appropriate content standards, while the University of Maryland Administrative Metadata (UMAM) Tag Library documents administrative and technical metadata requirements. Metadata policies planned in connection with the upcoming migration to Fedora 4 include implementing a resource description framework and employing industry-standard metadata such as MODS and PREMIS.

Generate Archival Information Package (AIP)

Archivists will create a BagIt Bag containing a preservation disk image; its administrative, technical, structural, and descriptive metadata; its checksum; and its

UUID. This Bag will be submitted to archival storage as an Archival Information Package (AIP).

Digital object extraction

When disk images contain multiple, diverse digital objects, archivists must determine whether to extract image contents as individual digital objects or to proceed through the workflow using a copy of the disk image as an access object. Extraction may mean separating an image into folders or even separate files. Decisions about object extraction or non-extraction will be based entirely on access concerns, including answering questions about access detailed above.

Extraction will involve mounting a disk image as a drive using BitCurator's disk image mounting tool, extracting files and folders in desired groupings, and generating a checksum for each grouping. A Digital Object record will be created in ArchivesSpace and a UUID assigned for each extracted object or disk image intended for access. Extracted objects may require normalization for long-term preservation, guided by a matrix of acceptable file formats developed by the Libraries' Born-Digital Working Group. A normalization event will be logged in ArchivesSpace if this process takes place.

Generate Dissemination Information Package (DIP)

Each digital object for access will be packaged—again using the BagIt specification—with its technical, structural, and descriptive metadata; its checksum; and its UUID. A "copied for access" event will be logged in ArchivesSpace.

Ingest DIP to Fedora

This Dissemination Information Package (DIP) is ingested to the Fedora digital repository system. Future preservation actions—that will also be logged as events in ArchivesSpace—will include periodic fixity checks and, if necessary, format migration.

Workstation Recommendations

BitCurator required a full re-installation for each update, the first sign that a separate workstation may be required for BitCurator tasks. The need for regular re-installation made it impossible to rely on local storage in the Linux system. FRED is partitioned in order to run FTK through Windows and BitCurator through Linux; users cannot run both simultaneously and must restart the machine to switch between the two systems. Hardware write-blockers, which are essential for creating a disk image without altering the original media, are built into FRED. Creating disk images through FTK Imager alone eliminates any need for have BitCurator to be installed on FRED. We therefore propose moving BitCurator from FRED to a separate workstation dedicated to reports and analysis. [xviii] This will alleviate the need for frequent workstation restarts and the problem of losing stored items or programs with every BitCurator update. We do not anticipate problems with mounting or analyzing disk images on a dedicated BitCurator station.

Challenges & Concerns

Connecting FRED

It is vital that the workstation be clean and virus-free when creating disk images [xix]. Competing theories exist on what makes a workstation safe from external risk. Erway (2012) suggests that the workstation remain “dark,” or non-networked in order to reduce vulnerability. By connecting FRED, we can automate virus scans and system updates. Connectivity will also be required for file transfers once a disk image is created. Ultimately we decided that the value of simplified transfers and automated virus scans outweighed the risks associated with connecting the workstation.

FRED is only equipped with 4GB of storage, so large disk images must be written to a larger storage area. Some of the drives in SCUA hold over 1 TB of data, which could not be imaged and stored on FRED alone. Disk images will therefore be exported to a working drive on the Libraries' local area network (LAN). This will allow disk images to be accessed from the BitCurator workstation immediately. File directories exported via FTK Imager are saved to the same location (such as a subfolder) as their corresponding images. We propose using this drive as a staging area for initial analysis of disk images using BitCurator's forensic tools.

We encountered difficulties in connecting and writing to the LAN through the Ubuntu partition, which appeared to be permissions-related. We anticipate that moving BitCurator to a Linux workstation will provide additional feedback towards resolving this issue. The LAN is still accessible through the Windows partition, so we will continue to write images to the LAN through FTK Imager.

Appraisal and Redaction

Redaction is an important but complicated issue in this project. When possible, the archivist performing born-digital work will discuss access restriction needs with the collection area archivist and the donor during the appraisal and acquisition process. The need to restrict access to files containing personally identifiable information may be part of image type decision making—some files and folders may prove too sensitive to acquire—but most restriction will happen after a disk image is acquired and placed in intermediate storage. For our current born-digital holdings, screening will take place after media have been imaged, packaged as a SIP, and submitted to APTrust, ready for processing into archival and dissemination packages. Given the large backlog of unprocessed born-digital items, it will be unrealistic to screen disk images at the rate they are created. Doing so would greatly impede the process of imaging and risk additional deterioration of unprocessed media.

Our proposed workflow includes several mechanisms to identify sensitive information, including conversations with donors about digital forensics and what the imaging process can reveal; creating logical disk images to avoid acquiring sensitive material that donors would not like to be preserved; and Bulk Extractor reports to identify any additional information. Access restrictions may not be necessary for some acquisitions. For example, carrier media used exclusively as a transfer mechanism will not require any redaction or access restrictions to mask deleted or hidden files. We expect that as SCUA's born-digital program evolves and archivists gain experience discussing these processes with donors, decisions about preserving sensitive information will be addressed entirely and exclusively during the appraisal process.

System Interoperability

The Libraries are implementing ArchivesSpace for the management of special collections across the Libraries as a way to standardize accessioning and description workflows. Describing born-digital material in ArchivesSpace will involve importing DFXML (produced by the Fiwalk tool in BitCurator) to digital object records, enhancing metadata in ArchivesSpace, and exporting enhanced metadata as XML. A Bag containing each digital object and its checksum, UUID, and enhanced metadata will be created and moved to long-term storage. This transfer constitutes the Archival Information Package (AIP) for a given digital object or group of objects.

Access to digital collections is managed through Fedora, and forthcoming enhancements with the migration to Fedora 4 will take into account ways to provide access to born-digital as well as digitized material. Until additional details of the upgrade are finalized, we propose that the final step of processing born-digital material will involve creating access copies from digital objects stored in the "grey" working area; generating a checksum for each access copy or group of copies; and packaging each access copy and checksum with the appropriate metadata and UUID for upload to Fedora. Understanding and fully using the capabilities of Fedora 4 is a major next step in this project. Each of the above steps will be traceable in ArchivesSpace, making it a useful tool for managing the entire born-digital workflow and providing a bridge between accession/transfer and access environments.

Future of the Project

The future of this project depends on the implementation of a larger workflow in Digital Collections. Questions still surround the migration to Fedora 4, which will include a migration of the entirety of Digital Collections. Existing metadata will be changed from University of Maryland's schema to a standard metadata schema. Developers are working to improve the system for batch loading into the new repository. All of these issues will affect the ingest of born-digital content. Flexibility will be necessary as the larger workflow is defined and implemented. The reports included and the processes recommended here may change based on the requirements of Fedora 4 as well as Solr and ArchivesSpace.

Concrete next steps for born-digital processing will include building upon the preliminary inventory and processing a 1TB hard drive of born-digital records of the National Labor College. The hard drive is part of the George Meany Memorial AFL-CIO Archive [xx], an extensive collection of documents, publications, photographs, film, sound, and other records related to the history of labor in the United States. Originally maintained at the National Labor College in Silver Spring, MD, the majority of this material was acquired by SCUA in October 2013 [xxi]. The National Labor College hard drive will provide an appropriate test case for the workflow proposed above because it contains complex storage media and a wide range of record types and file formats.

We will continue to explore new processes and technologies as we focus on the future of the born-digital initiative at the Libraries. The BitCurator Access project, for one, is expanding its capacity to support web-based disk image access services and enable users to export file systems and metadata. Digital forensic tools were created in the context of law enforcement and are still new to archival processing, but projects like BitCurator and FTK are working to bridge the gap in archival usability. As we continue to develop this process, we must remain flexible and open to incorporating new tools and workflows.

About the Authors

Alice Prael holds a BA in English from the University of Missouri in Columbia. She currently works as the graduate assistant for Digital Programs and Initiatives at University of Maryland Libraries, providing support and research for digital projects.

Amy Wickner holds an AB in Architecture and Urban Planning from Princeton University. She is currently the graduate assistant for Special Collections and University Archives at University of Maryland Libraries. She has previous worked as Research and Development Associate at Editorial Projects in Education.

Notes:

[i] Digital Intelligence. “FRED.” Accessed April 14, 2015. https://www.digitalintelligence.com/products/fred/

[ii] AccessData. “FTK Imager User Guide.” March 12, 2012. Accessed April 14, 2015. https://ad-pdf.s3.amazonaws.com/Imager3_1_4_UG.pdf

[iii] BitCurator. “BitCurator.” Accessed April 10, 2015. http://www.bitcurator.net/

[iv] BitCurator. “BitCurator: About the Project.” Accessed April 10, 2015. http://www.bitcurator.net/bitcurator/

[v] BitCurator. “BitCurator Consortium.” Accessed April 10, 2015. http://www.bitcurator.net/bitcurator-consortium/

[vi] BitCurator. “BitCurator Access.” Accessed April 10, 2015. http://www.bitcurator.net/bitcurator-access/

[vii] Digital Intelligence. “FRED.” Accessed April 14, 2015. https://www.digitalintelligence.com/products/fred/

[viii] Device Side Data. “FC5025 USB 5.25” Floppy Controller.” Accessed April 14, 2015. http://www.deviceside.com/fc5025.html

[ix] Drew, Jim. “Setup and Usage Manual.” SuperCard Pro. April 3, 2014. Accessed April 14, 2015. http://www.cbmstuff.com/downloads/scp/scp_manual.pdf.

[x] Lauren M. Cahill, “Surveying for Born-Digital: The Search for Special Formats” (poster, University of Maryland College of Information Studies Spring 2015 Experiential Learning Expo, College Park, MD, May 11, 2015).

[xi] Erway, Ricky. Swatting the Long Tail of Digital Media: A Call for Collaboration. Dublin, OH: OCLC, 2012. http://www.oclc.org/research/publications/library/2012/2012-08.pdf.

[xii] Kirschenbaum, Matthew G., and Richard Ovenden. Digital Forensics and Born-digital Content in Cultural Heritage Collections. Washington, D.C.: Council on Library and Information Resources, 2010.

[xiii] “Guymager homepage.” Accessed April 14, 2015. http://guymager.sourceforge.net/

[xiv] Forensics Wiki. "Fiwalk." Last modified September 13, 2013. http://www.forensicswiki.org/wiki/Fiwalk; Bradley, Jessica R., and Simson L. Garfinkel. "bulk extractor 1.4 User Manual." March 23, 2015. http://digitalcorpora.org/downloads/bulk_extractor/BEUsersManual.pdf; BitCurator. "Generating an Annotated Features Report." Last modified March 9, 2014. http://wiki.bitcurator.net/index.php?title=Generating_an_Annotated_Features_Report.

[xv] 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

[xvi] Academic Preservation Trust. "Core Preservation Services." Accessed December 16, 2014. https://sites.google.com/a/aptrust.org/aptrust-wiki/home/content-advisory-group/basi

[xvii] Yvonne Carignan et al., Best Practice Guideliens for Digital Collections at University of Maryland Libraries. College Park, MD: Office of Digital Collections and Research, University of Maryland, College Park, 2007.

[xviii] Plans for a dedicated digital forensics workstation built around BitCurator would use the following as a starting point: Olsen, Porter, "Building a Digital Curation Workstation with BitCurator (update)," BitCurator Blog, August 2, 2013, http://www.bitcurator.net/building-a-digital-curation-workstation-with-bitcurator-update/

[xix] Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-House. Dublin, OH: OCLC, 2013. http://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf

[xx] University of Maryland Libraries, George Meany Memorial AFL-CIO Archive, accessed May 1, 2015. http://www.lib.umd.edu/special/collections/afl-cio

[xxi] Hottle, Jenny, "AFL-CIO archive donation largest in University of Maryland library history," The Diamondback, October 8, 2013, accessed May 1,

2015. http://www.diamondbackonline.com/news/campus/article_1be96904-2fd3-11e3-a58a-0019bb30f31a.html