Born Digital Task Group Report
Submitted to the Collections Council of the Triangle Research Libraries Network (TRLN)
May 21, 2014

1. Introduction

The Triangle Research Libraries Network (TRLN) is seeing a growing presence of born digital materials in member library collections. To date, however, there has been minimal collaboration or sharing of information across TRLN libraries in this arena, with each institution devising its own tools, methods, and priorities for handling born digital resources. The heads of Special Collections at each institution met in the fall of 2013 and recommended the formation of the Born Digital Task Group, which was established by the Collections Council in January 2014 to increase information sharing, identify areas of need, collaborate on shared problems, and maintain a larger context and awareness of potential partnerships. Specifically, the task group was charged with providing updates and reports to the Collections Council that address the following areas:

- Explore needs in areas like collection development, acquisitions, ingesting, processing, and access to and long-term curation of collections
- Determine if there are areas/projects that would benefit from consortial action, such as training, equipment, and collaborative collections
- Provide regular updates on progress in managing born digital collections
- Create shared documentation

Members of the task group include Brian Dietz (Chair, NCSU), Jason Evans Groth (NCSU), Matthew Farrell (Duke), Meg Tuomala (UNC), and Yan Wang (NCCU). NCCU was involved in initial meetings and conversations but determined that its born digital activities are not yet substantive enough to warrant detailed contribution to this particular report. This report summarizes the discussions and findings of the group to date and presents its recommendations for action.

2. Current practices

Task group members extensively and thoroughly documented the processes and resources related to born digital resources at their respective institutions. That detailed information is provided in an appendix to this report. In this section we provide a condensed version of the findings.

Donor relations

Curatorial staff at Duke's David M. Rubenstein Rare Book & Manuscript Library (RL going forward) have created multiple gift and purchase agreements for different categories of acquisitions, most of which were revised in 2012 to address born digital components of potential

acquisitions. An example of the language surrounding born digital objects was included as an addendum to an ARL Spec Kit in 2012. The NCSU Libraries' Special Collection Research Center (SCRC) has both standard and alternative deeds of gift that are used when working with a donor to officially accept a donation. UNC addresses issues of ownership, transfer of rights, access, and privacy and confidentiality in its donor agreement forms but does not currently include language that specifies or singles out born digital materials as different from analog materials.

Whether and when the digital records archivist is involved in conversations with potential donors varies by institution, center, and curator. At Duke, the digital records archivist meets with donors, at the request of the curator, to discuss the specifics of transferring digital materials. These consultations tend to occur when the contents of a potential acquisition are heterogeneous and complex. At NCSU, born digital materials have been added as media to traditional accessions, and talking points and standards for acceptance of such material are still being developed by SCRC staff. The SCRC describes to the donor the process and its implications for preservation and access. Curators and collectors at the Wilson Special Collections Library at UNC are moderately comfortable addressing the issue of born digital materials with donors, and they are able to cover basic preservation and access issues. Usually, the collecting archivist and/or the curator is also present during this conversation. At each school, specific hardware and software are presented and demonstrated, as appropriate. Involvement of various staff in direct conversations with donors depends on the donor, the project, the center, and the curator.

Appraisal

Depending on the method of acquisition, some digital materials at Duke are inspected prior to accessioning. Network transfers are the most likely to be considered by curators or the digital records archivist. Materials received on removable media are rarely inspected prior to acquisition. An official appraisal policy for born digital materials is currently under discussion with curators and technical services. At NCSU, born digital collections are appraised based on the importance of the creator and what we can learn about the materials from the creator's description as well as through collection-level surveys using automated tools. Media objects must be readable by existing equipment in order to appraise files and resources fully. Wilson Library does not currently have an optimal solution for the appraisal of born digital materials after they are acquired but before they are processed; but various staff have started discussions on how workflows for the appraisal of born digital materials can be improved, perhaps using Curator's Workbench as part of the solution, and workflow priorities have been identified.

Accession

At Duke the curator creates a record in archival management software (AMS)[1]. For material on removable media, a photograph is taken of the medium. For each digital carrier, the digital

[1] All three institutions currently use Archivists' Toolkit. Duke will migrate to ArchivesSpace in August 2014. NCSU will soon charge an implementation team, with migration likely to follow over the summer. UNC will begin their investigation into implementation in the late fall.

records archivist assigns a unique identifier to the item, creates a disk image, performs sensitive information and virus scans, calculates fixity information, and logs this information. Accessioning work is performed on an electronic records workstation. Disk images and their photographs are stored in a semi-secure networked storage location. For materials acquired via network transfer, a record is created in AMS. Sensitive information and virus scans are performed and logged. Files are stored in a semi-secure networked storage location.

At NCSU, born digital resources, whether physical materials or materials received electronically, must be accessioned into a collection before they can go through the processing workflow. A born digital acquisition is attached to an accession number which is stored in Archivist's Toolkit and becomes part of the file naming convention for the ingest and storage process. Once born digital assets are flagged for processing, staff will image the digital media, perform virus scans, and extract metadata that may be helpful in describing the assets. All materials are bagged using the BagIt specification, and the bag is synced to networked storage.

UNC's accessions of born digital materials are handled primarily through the accessioning workflows already in place for each collecting unit of Wilson Library. While each unit handles accessions slightly differently, certain things happen across all collections. After an accession form is created, born digital materials are transferred from storage media to the library's digital storage using Curator's Workbench and/or BitCurator. A record for the material is then created in the AMS.

Arrangement and description

Initial efforts into determining a set of procedures for arrangement and description are underway at Duke. Currently, RL assigns all archival collections a processing level based on anticipated use, access restrictions, and other factors. Forays into arrangement and description use these processing levels as guidelines, mapping physical activities to digital counterparts as appropriate. A preservation copy of the born digital objects is maintained. RL will be moving to ArchivesSpace for description in the coming months. Examples of past efforts toward born digital description in RL collection guides include the Doris Duke papers and the Reynolds Price papers.

NCSU does not change the arrangement of files or try to represent an arranged version of the objects. The current SCRC plan is to link a file manifest from the finding aid for the collection to which the born digital assets are attached. The original order is relied upon, and patrons can navigate it for their own purposes. The presence of electronic media is noted in AMS and the finding aid. NCSU is discussing a method for displaying directory structures of electronic accessions in finding aids, possibly using the <otherfindaid> element.

The arrangement and description of born digital materials is, for the most part, integrated into Wilson Library's analog processing workflows at UNC. Born digital materials are processed using Curator's Workbench which allows staff to re-arrange directories and files just as they

would in the analog environment, all the while preserving the original order as captured from the original media during accessioning. A submission information package (SIP) is then created within Curator's Workbench and placed in a library shared drive that is accessible to CDR staff. Technical services staff alerts CDR staff that the SIP is ready, and CDR staff proceeds to ingest the SIP into the CDR. Born digital materials can be re-processed, or re-arranged, once the SIP has been been ingested into the CDR.

Ingest

Duke's accessioned born digital materials reside on semi-secure networked storage drives managed by Library IT. These shares feature periodic fixity checks and are accessible by few staff members. A separate share, managed by university OIT serves as the secure server for material containing sensitive electronic information. This share is accessible only by the digital records archivist and members of OIT, and requires an ITS security provided VPN to connect. It also features periodic fixity checks. While these measures make both storage locations better suited for materials than a local computer or office share setup, neither location is a permanent preservation solution. Ingest, at this point, refers to the movement of accessioned objects from the workstation to one of the temporary, semi-secure shares.

At NCSU the bag that is created in the accession process described above is also the AIP put into networked storage. The bag's fixity is checked before ingest and then checked regularly while in storage.

At UNC, the first step in preparing born digital materials for ingest into UNC's digital repository happens during arrangement and description. A final step in archival technical services workflow for processing born digital materials is to create a SIP for ingest into the CDR.

Access

Requests for access to born digital materials at Duke are to this point infrequent. As such, access is facilitated on a single terminal in RL's reading room with a general user account that restricts access to external networks. Patrons can request reproductions of files as they would with physical collection reproductions. It is likely that this will remain an option for access onsite, particularly for RL's most sensitive content. Ongoing work is aimed at creating a virtual reading room in order to provide remote access in with some level of mediation.

At NCSU, patrons request electronic resources from SCRC just as they would physical materials. A disk image is downloaded from storage by SCRC staff and mounted on a dedicated laptop. Patrons can see the shared folder on the desktop and can access files inside of it but do not have the ability to turn on networking or USB capabilities. When the patron is finished the image is unmounted or the logical copy of files are removed from the shared folder and the computer is returned to the public services desk. The patron may place files of long-term interest

in a folder for curatorial staff to review after the patron has gone. Staff will then determine if copies of the requested files can be given to the patron (physically or electronically).

Web-level access to processed born digital materials at UNC is provided through the Carolina Digital Repository. Born digital materials are described at the file, or folder, level, and finding aids for collections that contain born digital materials include links to the "digital files" and/or objects in the CDR[2]. Embargoes, restrictions, and proxies that limit access to campus-only enable the CDR to control access to sensitive materials. UNC is also able to provide access to born digital materials that are described in finding aids but not yet ingested into the CDR. When these are requested by a researcher, an expedited workflow is initiated.

Long-term preservation

The Duke library became a Hydra partner in 2013, and Library IT has been developing a preservation repository since that time. The end goal is to have a repository suitable for the long-term storage of all of RL's collection material, even if many of the materials are too sensitive for web access. In the meantime, the storage locations described in the above Ingest section approximate a preservation space.

Born digital image files at NCSU are stored both on-campus and replicated in cloud or tape storage. Physical materials remain with the physical collections, but once appropriate images are made of the assets, they are used only for data extraction in extraordinary cases.

At UNC, born digital archival materials are preserved in the CDR. Digital objects and associated metadata files written to the CDR's storage infrastructure are automatically replicated across the storage grid and onto archival tape. One copy of every file is stored on a server administered by the Library, one copy is stored on a server administered by the campus' Information Technology Services, and the archival tape copy is stored off-site in a geographically separate location.

3. Proposed activities and support needed

The task group identified several potential activities where additional support may be needed.

BitCurator site visit - TRLN member institutions will use, or are exploring the use of, BitCurator, a forensics software package developed by UNC-Chapel Hill School of Information and Library Science and the Maryland Institute for Technology and the Humanities at the University of Maryland. The BitCurator team has issued a request for proposals to host a site visit. The Born Digital Task Group is likely to submit a request, and we would request the support of TRLN.

[2] For an example of how access to born digital archival materials is provided at UNC, see the finding aid for the John Kenyon Chapman Papers, series 12-16.

Communicating with IT departments and administrators - The group sees the need for strategies and protocols for communicating with respective IT departments and administrations. This is no way suggests that any group member perceives existing institutional support to be inadequate. However, all agree that the ability to communicate needs and goals more effectively with technologists and administrators could be improved.

Formalizing a relationship with UNC digital forensics laboratory - Group members have been invited to use the digital forensics laboratory at the UNC School of Information and Library Sciences (SILS), and we may seek to formalize the relationship with Professor Cal Lee and SILS. Based on our experience with them, we anticipate learning of software, hardware, and strategies that we may need for our own in-house use.

Revising TRLN IP Rights Strategy - In 2011, TRLN issued its *Intellectual Property Rights Strategy for Digitization of Modern Manuscript Collections and Archival Record Groups*. This document is geared toward special collections materials that had been reformatted through collection-level, or large-scale, digitization. This group would like to see the statement extended to born digital collections, which share similar rights issues with mass digitized resources.

Using BitTorrent for transfer of remote files - Each institution is looking for better methods for archival-appropriate transfer of remote files. Known methods, including commercial tools like Dropbox and Drive, and non-commercial tools like SFTP and NC State's Velocity, are not sustainable, archival solutions for file transfer. The BDTG would like to pursue an investigation of BitTorrent as a potential solution. BitTorrent, a peer-to-peer file sharing protocol, is a powerful tool that has great potential for libraries and archives: BitTorrent automatically divides transferred files into pieces. This allows the user to download files non-sequentially, as well as allowing the transfer to be interrupted and restarted without requiring a complete restart. Additionally, each has a crypotgraphic hash calculated prior to transfer. The receiving user's BitTorrent client monitors the transfer, checking the hashes as it goes, to ensure an authentic file transfer. However, peer-to-peer sharing has negative connotations because of its use with illegal file sharing. The group would like to see TRLN issue a position paper on the use the BitTorrent protocol for libraries.

Emulation environments - Along these lines, group members are interested in pursuing emulation as a method for making accessible to patrons files from legacy software and operating systems. Creating emulation environments runs the risk of intellectual property infringement even though libraries would use legacy software for not-for-profit educational purposes; it is also uncertain what the commercial market for legacy software is. The group would like to propose that TRLN consider incorporating software into any position it takes on fair use and born digital resources.

4. Collaboration

The group has identified several areas for potential collaboration in order to support each others' efforts in managing born digital resources.

Continuation of group in current or more permanent form - The task group has served a valuable purpose in beginning to build relationships among those in TRLN immediately tasked with managing born digital resources. The task group has determined that there is considerable work left in order to firmly establish a foundation of collaboration in born digital resources. As such, the group would like to continue working together, at least in its current form if not as a more permanent TRLN committee. Continuation of the task group or establishment of a TRLN committee would, among other things, enable us to establish a pattern and expectation of inter-institutional communication on born digital issues. The group also sees an opportunity to work more closely with each other on specific problems. For example, institutions with the technical capacity to handle media that others cannot may offer a file retrieval service (e.g. disk imaging). A continued task group or committee will formalize our ability to approach one another for advice on unexpected and/or shared issues, to engage in periodic intensive problem solving sessions, and to advise on the formation of other related task groups or initiatives as needed.

Collaborative documentation site - Group members would like to develop a collaborative documentation site that may be based on the detailed current practices described in Section 2 and the appendix to this report. The task group anticipates several opportunities to promote and share its work, including co-authored articles and blog posts; if collaborative work were to continue further, the group could seek opportunities to present at state, regional, and national conferences.

Collaboration with affiliated campus units and community members - The task group has identified opportunities to explore working with affiliated campus units and community members, such as the Digital Forensics Lab at SILS (UNC-CH), computer science programs, eg, Digital Forensics at NC State, campus OITs, and computer enthusiasts.

TRLN seminars on management of born digital resources - Group members would like to pursue hosting a TRLN-led workshop or series of workshops. A combination of current task group members could provide born digital training sessions for TRLN staff as well as for others at local, regional, statewide, and, potentially, national conferences. These workshops may address any or all of the areas presented in Section 2 of this report.

Global born digital - The task group is aware that the TRLN Global Strategy Task Group has discussed the value of collecting global born digital materials (e.g. documenting diaspora, collecting digital materials related to specific global conflicts or natural disasters). The Born Digital Task Group will contribute to further discussions related to capturing this content.

Overall, the Born Digital Task Group has found its initial work to be productive and educational. Given the collective resources at our libraries, TRLN is positioned to contribute to local, national,

and global conversations and projects to help establish standards and practices for collecting born digital materials, at our individual libraries and consortially.

Appendix: Detailed Information on Current Practices

Donor Relations

Duke

Curatorial staff at the David M. Rubenstein Rare Book & Manuscript Library (RL going forward) have created multiple gift and purchase agreements for different categories of acquisitions, including manuscript collections from individuals and organizations, artificial collections, printed volumes, and university records transfers. The documents were revised in 2012 to address born digital components of potential acquisitions. An example of the language surrounding born digital objects was included as addendum to an ARL Spec Kit in 2012.

Curators have begun having conversations with donors about born digital materials, although practice varies by research center and curator. At the request of a curator, the digital records archivist meets with donors to discuss the specifics of transferring digital materials. These consultations tend to occur when the potential acquisition is heterogeneous and complex. For instance, an acquisition of digital materials that exist only on removable floppy or optical media may not garner a meeting with the digital records archivist. An acquisition that involves files stored in various places including mass storage devices, potential acquisition of whole computing environments, etc. is more likely to involve consultations.

Some curators circulate guidelines to donors regarding the types of materials suitable for preservation in the archives. Guidelines are tailored specifically to the research center. For instance, the Human Rights Archive's guidelines focus more on organizational records than personal papers. The guidelines are meant to be format agnostic, although discussion of born digital materials could be more explicit. Members of RL staff contributed to *born digital: Guidance for Donors, Dealers, and Archival Repositories*, a CLIR report released in Fall 2013 and current plans for RL general and research center-specific guidance will follow much of what was laid out in this document as well as the AIMS white paper and PARADIGM workbook.

Hardware & Software: Tableau Forensic Bridge Write Blocker and USB- 3.5" floppy drives are available for use in conversing with donors. More common are verbal descriptions of the systems and actions to be employed.

Staff: RL collection development staff includes seven professional positions. The university archivist has curatorial duties as well. The digital records archivist is consulted as needed.

NC State

The NCSU Libraries' Special Collection Research Center (SCRC) has both standard and alternative deeds of gift that are used when working with a donor to officially accept a donation.

Born digital materials will be accessioned both as physical objects and objects accepted electronically. Currently they have been added as media to traditional accessions, and talking points and standards for acceptance of such material are still being decided upon by staff in the SCRC.

The materials may be brought to the NCSU Libraries, or an SCRC staff member may retrieve the materials, and in some cases may only be retrieving images of the physical materials rather than the physical objects themselves. The SCRC strives to create disk images of all accessioned born digital resources in order to preserve them as exactly as possible. The donor is made aware of what imaging means, including the fact that on certain media previously deleted files which were thought to be gone may still be part of the fabric of the object and viewable, by forensic means, by both the SCRC and potential researchers. The donor will be given the opportunity, as donors are given with all donated collections, to determine the level of access granted to researchers (not providing images, for example, but logical copies of directories should there be privacy concerns). The SCRC, in order to secure the best chance of preserving digital resources with some longevity, will suggest storing the information locally as full disk images.

Hardware and Software: We have "talking points" to explain the born digital process and the tools we use to extract data from the media. Donors will be given the opportunity to see the workstation to better understand what is done and how their information is stored and retrieved.

Staff: Collecting and curatorial staff are the point people in this situation, and are encouraged to bring in the born digital team for additional consultation at any point during the process.

UNC

Curators and collectors at the Wilson Special Collections Library feel moderately comfortable addressing the issue of born digital materials with donors. They certainly have been starting these conversations with donors, and are able to cover basic preservation and access issues with donors in regards to born digital materials.

The Electronic Records Archivist has been asked to consult with a handful of donors regarding born digital materials and uses the *Paradigm Workbook on Digital Private Papers* and the *AIMS Project White Paper* as a framework to guide these conversations.[3] Usually, the collecting archivist and/or the curator is also present during this conversation.

The issues of ownership, transfer of rights, access, and privacy and confidentiality are addressed in donor agreement forms however we do not currently include language that specifies or singles out born digital materials as different from analog materials.

Hardware and Software: We do give donors a brief overview of the Curator's Workbench and Carolina Digital Repository if and when appropriate.

[3] Specifically the Paradigm records survey (Appendix C) and the AIMS Donor Survey (Appendix F).

Staff: Collecting and curatorial staff lead conversations with donors on the topic of born digital materials, consulting with the Electronic Records Archivist as needed.

Appraisal

Duke

Depending on the method of acquisition, some digital materials are inspected prior to accessioning. Network transfers, whether via FTP, DropBox, or similar service, are the most likely to have materials glanced at by curators or the digital records archivist. For university records, this may be accomplished by the university Office of Information Technology (OIT) granted network access to the digital records archivist, who can work with the university archivist to identify the directories to capture. Materials received on removable media are rarely inspected prior to acquisition. During accessioning, obviously out of scope media are not imaged, including blank disks and most commercial software. An official appraisal policy for born digital materials is currently under discussion with curators and technical services. One potential component of an official policy may be the development of a questionnaire distributed to a potential donor about the nature of their born digital files.

Hardware & Software: Tableau forensic bridge and USB-3.5" floppy drive are available for inspecting and appraising media prior to its accession.

Staff: For materials inspected prior to acquisition, curatorial staff performs appraisal. For materials appraised after acquisition, digital records archivist consults with curatorial staff about materials that seem out of scope.

NC State

Born digital collections are appraised based on the importance of the creator and what we can learn about the materials from the creator's description as well as through collection-level surveys using automated tools. The media object or resources must be readable by our equipment in order to process it through the workflow.

Hardware and Software: AMS is updated with the information that born digital materials are part of the collection.

Staff: Curatorial and collections staff see the resources first. Should questions about the resources' chances of being processed arise, staff members from the born digital team are consulted.

UNC

Appraisal workflows for born digital materials are still being hammered out in the Wilson Library. Currently, the initial decision to acquire born digital materials is made by curators and collectors using information gleaned from conversations and meetings with creators and donors, as described in the Donor Relations section.

Wilson Library does not currently have a great solution for the appraisal of born digital materials after they are acquired but before they are processed. The Curator's Workbench, a born digital collections processing and ingest tool, can be used to see the directory structure/general organization of born digital materials, file types, and file names. For certain file types, a preview of the file can be viewed within the Workbench application or downloaded.

Curators and collectors, archival technical services staff, digital repository services staff, and the Electronic Records Archivist have started discussions on how workflows for the appraisal of born digital materials can be improved. We hope to add functionality to our existing workflows and tools, mainly the Curator's Workbench, to make them more useful in the born digital appraisal process. So far we have identified the following as priorities:

- Ability to "flag" files using Curator's Workbench (e.g. flag for privacy concerns, duplicates, discards, high-value, etc.)
- Sorting functionality (e.g. sort by flag type, content type, file type, etc.)
- Ability to attach notes and/or additional information to materials and save this as metadata
- Reporting features, including
    - File/content types
    - Size/extent
    - Identification of duplicates
- Ability to look at files "on the spot" with the creator/outside of the library

Hardware and Software: Currently, the Curator's Workbench can be used for minimal/basic appraisal. Additionally, detailed information related to appraisal is recorded in AMS and other internal documents.

Staff: The following staff from across Wilson Library are involved in the appraisal of born digital materials:

- Curators and collectors (NCC and NCCPA, SFC, SHC, UARMS)
- Electronic Records Archivist (UARMS)
- Administrative Assistant (SHC)

Additionally, archival technical services staff are a main stakeholder in discussions for improving our appraisal workflow.

Accession

Duke

The curator creates a record in AMS. For material on removable media, a photograph is taken of the medium. For each digital carrier, the digital records archivist barcodes the item, creates a

disk image, performs sensitive information and virus scans, calculates fixity information, and logs this information on a SharePoint site (eventually AMS). Accessioning work is performed on an electronic records workstation. Upon completion of the tasks associated with accessioning, disk images and their photographs are stored in a semi-secure networked storage location.

For materials acquired via network transfer, a record in AMS is created. Sensitive information and virus scans are performed and logged. Upon completion of tasks associated with accessioning, files are stored in a semi-secure networked storage location.

Hardware & Software: In progress accessions are stored on an electronic records workstation, a Windows-based desktop PC which includes the following:

- Intel Core i7-2600 processor, 4 GB RAM
- Windows 7 Enterprise 64-bit
- two internal hard drives, 500 GB drive for application and operating system data and 1TB drive for in-process collection material.
- two DVD-R drives
- Kryoflux board for imaging floppy disks
- 3.5” Samsung floppy disk drive with common 34-pin connector
- 5.25” floppy disk drive with common ribbon connector
- Tableau T8-R2 USB forensic bridge write-blocker
    - recently discovered is incompatible with USB 3.0 connectors
- USB-based ZIP250 MB drive (backward compatible for 100 MB disks, but not compatible with 750 MB media)
    - Iomega stopped supporting these devices at Windows XP. So to accession ZIP media, we are looking into emulating an XP environment.

The following applications are currently in use, but will potentially be replaced by BitCurator:

- FTK Imager for imaging USB and optical media
- OSFMount to mount images
- IdentityFinder for SEI Scanng
- Symantec Endpoint Protection for virus scan.
- Jacksum for fixity calculation

Additional equipment available for accessioning electronic content:

- 3.5” floppy USB drive
- 1 TB USB external hard drive
- iMac in order to access and accession firewire-based media and peruse media employing HFS filesystems
- Desktop PC running Ubuntu to test BitCurator in a native environment and other Linux-based applications

Staff: Currently the digital records archivist performs all tasks related to accessioning apart from creating the record in AMS. Conversations are underway to assign certain tasks associated with accessioning to technical services staff.

NC State

Born digital resources must first be accessioned into a collection before it can go through the born digital processing workflow. Resources are attached to an accession number which is stored in AMS and also becomes part of the file naming convention for the ingest and storage process. Both physical materials (hard disks, floppy disks, USB drives, optical media, etc.) and materials received electronically (through NC State's Velocity tool, Google Drive, DropBox, or imaged/logical copied from the donor without receiving a physical object) will be assigned an accession number before being processed.

Once born digital resources is flagged for processing, the following steps are taken:

- A folder, named eadid_accessionNumber_yyyymmdd, which is shared between the desktop and the BitCurator environment, is created. Inside the folder are two additional folders: eadid_accessionNumber_yyyymmdd_image and eadid_accessionNumber_yyyymmdd_documentation (NOTE: "eadid" is our EAD ID number, "accessionNumber" is the accession number, and yyyymmdd is the datestamp that corresponds with the date of processing)
- Media is evaluated to determine which workflow is to be used
    - 3.5" disks, ZIP disks, CDs, DVDs, Blu-Rays, USB flash drives, and USB external drives have one workflow associated with our WiebeTech write blockers
    - 5.25" disks connect to the system via the FC5025
- Pictures are taken of the media object and combined into one file called eadid_accessionNumber_yyyymmdd_photos and placed in the eadid_accessionNumber_yyyymmdd_documentation folder
- A transcription of physical characteristics is placed within a structured CSV file called eadid_accessionNumber_yyyymmdd_transcription and placed within the eadid_accessionNumber_yyyymmdd_documentation folder
- The correct write blocker is connected, as is the device to be imaged (through the write blocker), and a virus scan using TrendMicro is run over the device, the results of which are screen captured into a file called eadid_accessionNumber_yyyymmdd_virusScanResults and placed in the eadid_accessionNumber_yyyymmdd_documentation folder.
- A RAW image file of the device is created.
    - For everything but 5.25" disks FTK imager is used. The file manifest and other reports created when using FTK imager are placed in the eadid_accessionNumber_yyyymmdd_documentation folder. The image file is placed in the eadid_accessionNumber_yyyymmdd_image folder.
    - For 5.25" disks an image is created by using the simple disk imager and browser that came packaged with the FC5025. FTK Imager is then used to create a file manifest and extract image metadata, and these reports are placed in the

eadid_accessionNumber_yyyymmdd_documentation folder while the image is placed in eadid_accessionNumber_yyyymmdd_image
- BitCurator is enabled, and bulk extractor is run on the the image file inside the eadid_accessionNumber_yyyymmdd_image folder. The reports it creates are generated inside a folder called bulkExtractorOutput within the eadid_accessionNumber_yyyymmdd_documentation folder
- The “run all” reports feature is enabled, running Fiwalk and the other BitCurator reporting tools across the image, the output of which is placed in a folder within eadid_accessionNumber_yyyymmdd_documentation called bitCuratorReports
- Bagger is used to create a bag file of the contents within the eadid_accessionNumber_yyyymmdd folder, and is called eadid_accessionNumber_yyyymmdd_bag
- The bag file is given to IT to be put into storage

Hardware and Software: NCSU Libraries uses multiple hardware and software tools to process their born digital materials. These tools include:

- Dedicated, non-networked Windows 7-based digital forensics workstation with 1TB of storage for local storage options
- WiebeTech write blocking devices with multiple interfaces for legacy media
- Digital camera
- 5.25” Teac Disk Drive
- FC5025 disk controller for the 5.25” drive
- 3.5” Teac USB Disk Drive
- USB CD/DVD/Blu-Ray drive
- USB ZIP drive
- TrendMicro Virus Protection
- Microsoft Excel
- FTK Imager
- The BitCurator environment and all of its integrated tools inside VirtualBox
- Disk Image Creator
- FITS
- Fixity
- Bagger

Staff: The born digital team has taken great strides to create a repeatable, useful workflow with extensive documentation. Ideally almost anyone in the SCRC would be able to run the processes necessary for accession. At the moment the born digital team is the only group that is processing these resources.

UNC

Accessions of born digital materials are handled primarily through the accessioning workflows

already in place for each collecting unit of Wilson Library. While each unit handles accessions a little bit differently, there are certain things that happen across all collections. These are described here.

First, the curator or collector fills out both an accession form, and for accessions that contain a born digital component, they also fill out a digital materials survey form. For transfers of University records, University Archives staff and the creating unit fills out a transfer form. These forms contain the information needed for accessioning and later archival processing.

After the forms are filled out, born digital materials are handed over from the curators and collectors to the staff member who performs accessioning for the collecting unit. This staff member transfers the digital materials from storage media to the library's digital storage using the Curator's Workbench and/or BitCurator The Curator's Workbench is a born digital collections capture, processing, and ingest tool created for archival born digital materials at UNC. BitCurator is a suite of digital forensics tools brought together for collecting institutions.

Like analog accessions, born digital and/or hybrid accessions are entered into AMS. Within their AMS record, notes are made that indicate the born digital nature of the materials, and any specific formats. This information is gleaned from the forms cited above, and/or from the materials themselves. Other information in the accession record includes basic to establish the physical and intellectual content of the materials.

Additionally, for high priority collections or additions, "Curator's Netflix," a spreadsheet used by curators and collectors to prioritize and queue accessions for processing, is used. Together with information in AMS, archival technical services selects materials for processing, and

Transfers of born digital materials that contain legacy or rare formats or a very large quantity of digital files (in excess of 2000 discrete files) are accessioned by the Electronic Records Archivist as these require specialized equipment or knowledge, and extra time, to process.

Hardware and Software: born digital materials are captured from digital media or storage using the Curator's Workbench and BitCurator. Additionally, detailed accessioning information is recorded in AMS and other internal documents. Depending on the media type, the following hardware is used:

- Tableau write-blockers
- various USB storage devices
- USB 3.5" floppy drive

Staff: Staff across Wilson Library participate in the accessioning of born digital materials including:

- Curators and collectors (NCC and NCCPA, SFC, SHC, UARMS)

- Records Services Archivist (UARMS)
- Electronic Records Archivist (UARMS)
- Administrative Assistant (SHC)

Arrangement and Description

Duke

Initial efforts into determining a set of procedures for arrangement & description are under way. The digital records archivist meets weekly with a technical services archivist to discuss methods for processing born digital materials. RL assigns all archival collections a processing level based on anticipated use, access restrictions, and other factors. Forays into arrangement & description use these processing levels as guidelines, mapping physical activities to digital counterparts as appropriate. Materials receiving a minimal level of processing may be described at the container level and have no further arrangement other than the extraction of logical files from the disk image. Larger, more complex volumes, as well as materials at higher priorities may be described at a more granular level, and may have logical files extracted and further arranged. While tools that articulate such an arrangement in a standard metadata schema without touching the underlying files are desirable, at present the options available involve creating arrangements by interacting with copies in a workstation's filesystem. Because of this, maintaining a preservation copy of the born digital objects as received is vital.

RL currently creates finding aids by individually creating EAD-XML documents, but will be moving to ArchivesSpace for description in the coming months. Examples of past efforts toward born digital description in RL collection guides include the Doris Duke papers and the Reynolds Price papers.

Hardware & Software: The Duke Data Accessioner is used to extract logical copies from disk images. Further arrangement work is thus far performed on those extracted copies in Windows Explorer. Duke uses the AMS's spreadsheet import functions for creating EAD finding aids.

Staff: Work performed in this area has been exploratory and has involved the digital records archivist and the technical services archivist for University Archives. Earlier efforts were undertaken by the Doris Duke project archivist. Once practice has been settled upon, work will involve staff throughout the RL technical services department.

NC State

We rely on the arrangement of the files as they appear on the original object to dictate the arrangement that the patron sees and interacts with. By utilizing a home-grown tool which merges file manifest output from either BitCurator or FTK Imager and metadata extraction tools like ExifTool, we create a virtual Finder/Explorer-like environment that emulates the file system as it appears on the disk. In other words, one is able to explore the emulated disk image the same way they would be able to explore the actual disk. While donor agreement dictates whether this emulated environment would actually allow the user to remotely access particular files over the web, each file would show "description" when it is chosen or moused-over, that

being important metadata values that are extracted during the processing of the image (file type, file title, creator, date, etc.). Additionally, we plan to generate reports, harvested from this same metadata, which provide summaries of what types of files are available per disk, word lists, and potentially other text mining data which may help a patron zero in on the disk they wish to browse more easily. All of this will likely be added to the end of the current finding aid, with a note that indicates that the disk/object “may contain” files that the evidence harvested from the metadata points to.

We do nothing to change the arrangement of the files, or try to represent an arranged version of the objects. The current SCRC plan is to link a file manifest, generated by FTK imager or a BitCurator tool, from the finding aid for the collection to which the born digital materials are attached. We plan to rely on original order and allow patrons to navigate it for their own purposes. We note the presence of electronic media in the finding aid and in our AMS. We are currently discussing a method for displaying directory structures of electronic accessions in finding aids, possibly using the <otherfindaid> element.

UNC

The arrangement and description of born digital is, for the most part, integrated into Wilson Library’s analog processing workflows. As part of the appraisal and accessioning process accessions are prioritized for processing by the curator. Using this information, accessions are selected for processing by the Manuscripts Processing Coordinator and the University Archives Processing Coordinator, who then assigns processing projects to the appropriate archival Technical Services staff member(s). Analog and born digital processing often happens concurrently, and is performed by the assigned staff member(s).

Staff process born digital materials using our collections processing and ingest tool, Curator’s Workbench. The Workbench allows staff to re-arrange directories, sub-directories, and files just as they would in the analog environment, all the while preserving the original order as captured from the original media during accessioning. Directories and files can also be re-named in the Workbench, and again the original names are preserved as captured. Usually, staff does not rename individual files and just uses the creator supplied filename. It is more common that staff re-names directories and subdirectories in a style that conforms to Wilson Library’s established archival processing practices and/or the style of processing that has been completed on any analog portion of the collection. This can include using pre-established file naming conventions, normalization/standardization (e.g. removing special characters, spelling out abbreviations), and grammatical changes.

Once the arranging and re-naming of the born digital materials has been completed by technical services staff, a submission information package (SIP) is created within Curator’s Workbench and placed in a library shared drive that is accessible to CDR staff. Technical services staff alerts CDR staff that the SIP is ready, and CDR staff proceeds to ingest the SIP into the CDR.

Also at this time, technical services staff uses Curator’s Workbench to create and download a

.CSV file that contains the UUIDs for each “digital folder” of materials that are to be ingested into the CDR. They then copy and paste these links into the collection’s finding aid that has been created or updated to include the born digital materials. At this time, archival description is only found in the finding aid. The CDR does not currently link back to the finding aid.

Additionally, born digital materials can be re-processed, or re-arranged, once the SIP has been been ingested into the CDR. The CDR administrative interface also allows technical services staff to restrict or place embargoes on materials at any level (directory, sub-directory, and/or individual files), and to mark materials for permanent destruction.

Hardware and Software: born digital materials are arranged and described using Curator’s Workbench, a collections processing and ingest tool developed at UNC specifically for these tasks. Since the data have already been acquired from the original media source, no specialized hardware is required for arrangement and description. Additionally, born digital materials can be arranged and described within the administrative interface of the CDR, however this is more time consuming. When materials need to be reprocessed after ingest, this must happen within the CDR.

Staff: The bulk of the arrangement and description of born digital materials is performed by Wilson Library’s Archival Technical Services unit, and is integrated into analog processing workflows.

Additionally, Archival TS staff devote 1.5 hours per week to participate in “born digital Case Studies,” where staff working with born digital materials meet to discuss special projects and issues, and actually perform accessioning and arrangement and description tasks. Attendance at the case studies is optional, but the core group managing the case studies includes:

- Electronic Records Archivist (UARMS)
- Manuscripts Processing Coordinator (TS)
- Visual Materials Archivist (TS)
- Head, Digital Repository Services (LIT)
- Lead Developer, Digital Repository Services (LIT)

Ingest

Duke

Accessioned born digital materials reside on a semi-secure networked storage drives managed by Library IT. These shares feature periodic fixity checks and are accessible by few staff members. A separate share, managed by university OIT serves as the secure server for material containing sensitive electronic information. This share is only accessible by the digital records archivist and members of OIT, and requires an ITS security provided VPN to connect. It also features periodic fixity checks. While these measures make both storage locations better suited for materials than a local computer or office share setup, neither location is a permanent

preservation solution. Ingest, at this point, refers to the movement of accessioned objects from the workstation to one of the temporary, semi-secure shares.

Hardware & Software: Teracopy used to securely move accessioned files from the electronic records workstation to secure storage server.

Staff: Due to the access restrictions on secure storage locations, the digital records archivist performs transfer.

NC State

The bag that is created in the process described above is also the AIP that will be given to IT to place into network storage. The bag's fixity is checked before ingest and then checked regularly while in storage. The bag will be rsynced to storage.

Hardware & Software: We use a command line utility, rsync, to move files to secure networked storage.

Staff: This work is performed by members of the born digital team.

UNC

As described above, the first step in preparing born digital materials for ingest into UNC's digital repository, the CDR, happens during arrangement and description. A final step in archival technical services workflow for processing born digital materials is to create a SIP for ingest into the CDR.

Using Curator's Workbench, archival technical services staff exports the SIP (a METS XML file), which is saved to a shared drive where CDR staff access and upload it to the repository. The SIP is essentially a manifest that references the locations of the digital objects (the AIP) that have already been staged to preservation storage, in the accession workflow described above.

Hardware and Software: Curator's Workbench, Carolina Digital Repository administrative interface.

Staff: Preparation and creation of the SIP can be performed by any archival technical services staff member working on the project. Actual ingest of the SIP into the repository is performed by digital repository services (i.e. CDR) staff.

- Archival Technical Services staff
- Digital Repository Services staff

Access

Duke

Requests for access to born digital materials are infrequent. As such, access is facilitated on a single terminal in RL's reading room. When a request arrives, the digital records archivist or member of the research services department loads a copy of the requested materials onto the workstation. The general user account restricts means of accessing external networks. Patrons can request reproductions of files as they would with physical collection reproductions. It is likely that this will remain an option for access onsite, particularly for RL's most sensitive content.

Ongoing work is aimed at creating a virtual reading room in order to provide remote access in with some level of mediation. University OIT manages the Virtual Computing Lab (VCL). RL has created a base VM running Windows XP. For each collection, the digital records archivist creates an instance of the base VM loaded with the digital objects and any software needed to view and interact with the objects and will update the collection's catalog record to reflect the VM as a requestable object. A patron requesting a VM will be routed through RL's request system. Because the user could still create a screen capture with their host machine, this system is unsuitable for the most sensitive materials. A final option for future access will be the access layer of the in-development Hydra repository, which may be suitable for materials with little to no need for mediated access.

Hardware & Software: In the reading room, a workstation has been set up to facilitate access to audio-visual and born digital materials. It includes the following hardware/software:

- iMac, Intel Core Duo processor, with the following modifications made
    - applications limited to Safari, OpenOffice, VLC, TextEdit, and Preview
    - web limited to localhost
    - printing disabled
    - AirPort disabled
    - FireWire and USB ports disabled\

The VCL's base image for RL materials is a Windows XP virtual machine with the following features:

- Limited software availability:
    - OpenOffice, IRfan View, Notepad++, VLC, Windows Media Player, WinDirStat, HxD, SQLiteStudio, XML Copy Editor, GIMP 2.4, grepWin, DROID
    - Depending on the materials to be served, other applications are located and installed, e.g. a collection including Outlook Archive files has PST Viewer installed.
- Strict Group Policies

- Data redirection restrictions including clipboard, COM port, client printer, and drive
- Disabling of network connections

Staff: For onsite requests, research services notifies the digital records archivist, who loads the materials onto the reading room terminal. For VCL requests, because the system is undergoing development, work is performed by the digital records archivist with assistance from Library IT. Eventually, RL research services will be integrated into the workflow.

NC State

Access to born digital resources are mediated by the SCRC. The file manifest created by FTK Imager or BitCurator is attached to the collection's finding aid. A patron requests electronic resources as they would physical materials. That disk image is downloaded from storage by SCRC staff and mounted on a dedicated SCRC reading room laptop (Macintosh). The mounted image is virus scanned, and the PPI data extracted from BitCurator on accession and the donor agreement are reviewed. Network and USB capability are disabled on the laptop, and the patron is given access through a patron account. The SCRC staff, with administrator privileges, either mounts the image and provides access to it through a shared folder or copies the requested files from the mounted image to the shared folder. When the patron is logged in, they can see the shared folder on the desktop and can access files inside of it, but do not have the ability to turn on networking or USB capabilities without the administrator password. Problems with this scenario will be approached on a case-by-case basis. The patron is only given access to what the donor has allowed per the donor agreement, which, in some cases, may mean a simple logical copy (retaining file paths) rather than a mounted image with hidden files. The image will be indexed by OSX to allow for text searches inside of files which will also help the user navigate what may be poorly or inconsistently named files. When the patron is finished the image is unmounted or the logical copy of files are removed from the shared folder and the computer is returned to the public services desk. The patron will be given the chance to place files of long-term interest in a folder for curatorial staff to review after the patron has gone. Staff will then determine if copies of the requested files can be given to the patron (physically or electronically).

Hardware & Software:
MacBook Pro
Spotlight (built-in)
Various search applications
Homegrown file directory emulation utility
Disk Utility for disk mounting

Staff:
SCRC Reading Room Staff
Born digital team

UNC

Web-level access to processed born digital materials is provided through UNC Libraries' digital repository, the Carolina Digital Repository. Finding aids for collections that contain born digital materials include links to the "digital files" and/or objects in the CDR.[4] Embargoes, restrictions, and proxies that limit access to campus-only allow us to control access to sensitive materials.

Additionally, UNC is able to provide access to born digital materials that are described in finding aids but not yet ingested into the CDR when these are requested by a researcher. When such a request is made the processing of these materials is prioritized, and an expedited version of the workflows described in the Arrangement and Description and Ingest sections above happens.

Hardware and Software:

- Carolina Digital Repository/Fedora (web application)
- oXygen XML Editor (used to create/update EAD finding aids)
- Curator's Workbench
- Carolina Digital Repository administrative interface (used to apply and manage embargoes, restrictions, and proxies)

Staff:

- Archival Technical Services staff
- Digital Repository Services staff

Long-term Preservation

Duke

The university library became a Hydra partner in 2013, and Library IT has been developing a preservation repository since that time. The digital records archivist and the archivist for metadata and encoding meet weekly with repository developers to discuss RL's preservation needs. The end goal is to have a repository suitable for the long-term storage of all of RL's collection material, even if some (or many) of the materials are too sensitive for web access. The project maintains a GitHub page. In the meantime, the storage locations described in the above Ingest section approximate a preservation space.

NC State

Born digital image files will be stored both locally (on-campus) and replicated in cloud or tape storage. Fixity checks will be performed on the images regularly. Physical materials will remain with the physical collections, but once appropriate images are made of the resources, they will only be used for data extraction in extraordinary cases.

[4] For a more detailed description of this see a https://drive.google.com/file/d/0B8-PNIzbs0eAOG1GNmk0N0EzTjQ/edit?usp=sharing

born digital archival materials are preserved in the CDR, which uses iRODS (integrated Rule Oriented Data System)[5] to implement its preservation storage environment. Digital objects and associated metadata files written to the CDR's storage infrastructure are automatically replicated across our storage grid and onto archival tape. One copy of every file is stored on a server administered by the Library, one copy is stored on a server administered by the campus' Information Technology Services, and the archival tape copy is stored off-site in a geographically separate location.

We also automate file characterization, checksum generation, virus scans, thumbnail creation, and access copy generation. Replication and fixity checks are performed on a quarterly basis for every file in the repository, and results of these checks are logged in the repository. Preservation events for each object are recorded in PREMIS.[6]

Hardware and Software:

- iRODS
- Carolina Digital Repository/Fedora
- Iron Mountain (tape storage)
- StorHouse (phasing out)
- Carolina Digital Repository administrative interface

Staff:

- Digital Repository Services staff
- Electronic Records Archivist (policy and workflow development)
- Metadata and Special Formats Cataloger (planned/future)

[5] https://www.irods.org

[6] A more detailed technology overview, including software, standards, and protocols used can be found at https://cdr.lib.unc.edu/scontent/aboutPages/techOverview.xml