

Outfitting a Born-Digital Archives Program

Ben Goldman

Penn State University

Archival repositories that intend to develop programmatic solutions for managing and preserving born-digital holdings will need to establish a dedicated computer workstation (and related devices) to support responsible capture, transfer, appraisal, and preservation steps. This brief examination provides baseline recommendations for implementing and equipping a dedicated Windows-based PC workstation.ⁱ

First Things First

Assessment is a critical first step before spending money to establish a workstation and supporting infrastructure. A survey of computer media in existing holdings will help identify media drives needed for a workstation. A comprehensive inventory of born-digital material will provide a baseline for quantifying immediate storage needs.ⁱⁱ

Partner with IT

Technical requirements should, to the degree possible, be coordinated with institutional information technology (IT) support. IT staff may be required to procure equipment, approve non-standard specifications (such as increased amounts of memory), locate obsolete hardware, install software, or establish dark archive storage.

Workstation as Quarantine

A born-digital workstation will preferably be exempted from other uses, and may even be isolated on the local network through firewall configurations, in order to minimize the risks from malware posed by both obsolete media and the Web. As OCLC notes, “the workstation serves the same purpose as the quarantine room many archives use for new acquisitions that have not yet been reviewed for mold, insects, etc.”ⁱⁱⁱ

Archivists can create a quarantine-like environment by forbidding web browsing on the workstation or manually disconnecting the ethernet cable connecting the PC to the network when working with archival material. Alternatively, IT staff can configure a ‘grey’ (not quite ‘dark’) workstation that connects to archival network storage and allows the machine to receive regular technical updates, but prevents all other access, including access to the Web. This second option is more secure but less convenient, limiting an archivist’s ability to quickly refer to resources or documentation, or blocking online registration of licensed software.

Computer Hardware Specifications

As the profession shifts from focusing on its legacy media backlog to acquiring gigabyte and terabyte-scale collections, archivists will appreciate having a workstation with some horsepower. Computers with multi-core processors and increased memory will enable better system performance when transferring or processing large deposits of data^{iv}. Four to eight gigabytes of RAM is considered minimal; 16 gigabytes of RAM would better suit archivists working with gigabyte- and terabyte-scale collections. Depending on the specifications chosen, a PC in this range might cost somewhere between \$500 and \$1000. Archives that expect to be working with significant amounts of system-intensive formats like digital video might consider top-of-the-line equipment. For setups that demand a significant number of peripheral disk drives, make sure the workstation has several USB ports, including front-facing ports for convenience.

Hardware Peripherals

Repositories with born-digital material on fixed legacy media will need disk drives in order to mount and read media. In certain cases, supporting devices will be needed to provide write-protection, or support interfacing between a modern workstation and an antiquated hard drive. The devices recommended below reflect the practical tools available based on the formats most frequently encountered by archivists; equipment should be selected based on local needs.

5.25" floppy disk drive: These can sometimes be found cheaply on Ebay (between \$50 and \$100), but archivists would be wise to query contacts in IT for retired hardware^v. Such drives are also notoriously unreliable, which may necessitate acquiring more than one.

Floppy disk controller card: Floppy controllers can help newer machines read media connected via obsolete disk drives^{vi}. One commonly used device is the FC5025 floppy controller^{vii} from Device Side Data, which costs around \$55. Documentation associated with the FC5025 recommends using it with a Teac FD-55GFR model 5.25" floppy disk drive.

3.5 inch disk drive: 3.5" floppy disk is one of the more ubiquitous formats encountered by archivists. USB-connected 3.5" floppy drives remain widely available at minimal cost (\$20), and can be purchased new or used.

Optical media drive: Most PCs currently come packaged with internal optical media drives for reading CDs and DVDs, but archivists should closely monitor the ongoing obsolescence of optical media drives by computer manufacturers^{viii}.

Memory card reader: USB Multi-format flash memory card readers can be purchased for as little as \$10, and support reading a variety of flash drive formats such as SD, CompactFlash, and MicroSD, found in digital cameras and other devices.

Write-blocker: These devices allow an archivist to write from but not to the connected media, protecting files (and their metadata) from unintended alteration. Hardware write-blockers, which are connected via USB between the media and the workstation, are most practical^{ix}. A typical hardware write-blocker, like the Tableau T8-R2^x, will accommodate most needs at a cost of \$300.

External hard drive: External hard drives are handy for transferring new acquisitions of born-digital material. A one terabyte hard drive can be purchased for around \$100. It should be noted that external hard drives formatted for Windows will present problems when seeking to transfer files from an Apple computer, which might necessitate purchasing multiple hard drives and formatting them for use with different file systems.



Software

The variety of software that might be utilized in support of working with born-digital materials is too great to fully explore here^{xi}, but a typical setup would include virus protection, forensic software, file characterization/appraisal tools, data manipulation applications (for working with metadata), and perhaps even emulation software^{xii}. Because digital archive solutions are fluid, archivists should plan for the likelihood that new tools will constantly be installed, updated, assessed, added to the evolving workflow, or removed from it. Ideally, the archivist working with these materials will have the necessary workstation administrative privileges to support this level of exploration.

Because many of the applications likely to be used are open source, it's important to understand that system updates can affect the use of such tools. For example, DROID^{xiii}, a Java application developed by the National Archives to support batch appraisal and format identification, will not function with the latest release of Java, which might require a workstation to run concurrent versions of Java. Likewise 64-bit machines may need to be run as 32-bit in order to support certain emulation tools. Coordination with IT staff will help the archivist navigate these technical challenges.

Conclusion

A workstation built using the computer hardware and devices identified here should accommodate the needs of most born-digital archival programs, at a cost somewhere between \$1200 and \$1700.

A final caveat is in order: these recommendations have a shelf life. The theory underlying our work with born-digital material provides solid footing for the profession moving forward, but we should consider the tools and technologies we employ to be transient and in constant need of review and revision.

About the Author

Ben Goldman is the Digital Records Archivist in the Special Collections Library at Penn State University, where he is responsible for developing workflows around the management and preservation of born digital archival collections. Ben is the product owner for ArchiveSphere, a Hydra-based repository for managing and preserving born-digital archival collections.

ⁱ This topic has previously been explored in blog posts detailing institution- and technology-specific setups. See: <http://www.nypl.org/blog/2012/07/23/digital-archaeology-recovering-your-digital-history> and <http://www.bitcurator.net/2013/08/02/building-a-digital-curation-workstation-with-bitcurator-update/>. All-in-one forensic workstations, such as the Forensic Recovery of Evidence Device (FRED) are outside the scope of this piece. Those interested in such setups should explore documentation provided by Stanford University's Special Collections and University Archives (<http://library.stanford.edu/spc/more-about-us/born-digital-program/lab-equipment-and-software>), or Johns Hopkins University Archives (<http://blogs.library.jhu.edu/wordpress/2013/11/digital-forensics-in-the-archives/>).

-
- ⁱⁱ Detailed inventory steps have been provided by OCLC Research. See: Erway, Ricky. 2012. You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2012/2012-06.pdf>.
Templates and example inventories are documented in the Jump In Initiative, sponsored by the Society of American Archivists' Manuscript Repositories Section: <http://www2.archivists.org/groups/manuscript-repositories-section/jump-in-initiative>.
- ⁱⁱⁱ Barrera-Gomez, Julianna and Ricky Erway. 2013. Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house. Dublin, Ohio: OCLC Research.
<http://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf>
- ^{iv} To provide one example, the author currently uses a 64-bit Dell Optiplex 790, running Windows 7 Enterprise, Service Pack 1, with Intel Core i7-2600 CPU and 16 Gb RAM.
- ^v The author acquired a 5.25" floppy drives for free after emailing IT staff, which elicited the following response from IT: "How about something challenging like an 8" floppy? We've got a stack of the 5 1/4 ones."
- ^{vi} A detailed description of why and how to use a floppy controller is provided by the Maryland Institute for Technology in the Humanities: "Use Guide for the FC5025 Floppy Disk Controller" at: <http://mith.umd.edu/vintage-computers/fc5025-operation-instructions>.
- ^{vii} See: <http://www.deviceside.com/fc5025.html>
- ^{viii} See "Apple's plan to wipe out disc drives is nearly complete" at: <http://www.cnet.com/news/apples-plan-to-wipe-out-disc-drives-is-nearly-complete/>.
- ^{ix} Hardware write-blockers are not necessary for all types of fixed media; 3.5" inch floppy disks can be protected using write-tabs on the disks, and 5.25" inch floppies can be protected using strategically placed tape.
- ^x See: <http://www.guidancesoftware.com/products/Pages/tableau/products/forensic-bridges/t8-r2.aspx>

^{xi} One participant in a recent discussion on the Digital Curation Google Group listed 28 different software applications installed for the purpose of working with born-digital materials. See: <https://groups.google.com/forum/#!searchin/digital-curation/equipment/digital-curation/fYMBpQh6m6l/uNFEfQU7rO8J>

^{xii} The author's current primary software applications are ClamAV for virus scanning, FTK Imager to create disk images, BitCurator (running on VirtualBox) to process disk images, DROID and Quick View Plus for appraisal purposes, and Aid4Mail for email processing. All of these tools are freeware with the exception of QuickView Plus (\$50 license), and Aid4Mail (approximately \$100 annual license). This is not a comprehensive list of all software installed, however.

^{xiii} DROID: <http://digital-preservation.github.io/droid/>