

Swatting the Long Tail of Digital Media: A Call for Collaboration

Ricky Erway

Senior Program Officer
OCLC Research



A publication of OCLC Research

Swatting the Long Tail of Digital Media: A Call for Collaboration
Ricky Erway, for OCLC Research

© 2012 OCLC Online Computer Library Center, Inc.
Reuse of this document is permitted as long as it is consistent with the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 (USA) license (CC-BY-NC-SA):
<http://creativecommons.org/licenses/by-nc-sa/3.0/>.
September 2012

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 1-55653-509-0 (978-1-55653-509-3)
OCLC (WorldCat): 806235133

Please direct correspondence to:
Ricky Erway
Senior Program Officer
erwayr@oclc.org

Suggested citation:
Erway, Ricky. 2012. *Swatting the Long Tail of Digital Media: A Call for Collaboration*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2012/2012-08.pdf>.

Introduction

Archiving born-digital content on a wide range of physical media types requires specialized knowledge, expertise, and equipment. As professionals in archives and special collections face the challenge of managing born-digital materials, they may be daunted by the prospect of having to read and preserve the content from all possible physical media, ranging from punched cards to flash drives. This truly is an inhibiting viewpoint, and it contributes to lack of action on the part of many.

It is difficult to do much with digital media unless you can read its content and transfer that content to more stable media. Few institutions can be expected to manage all media types. In order to make real progress in preserving and providing access to born-digital content, libraries and archives need to leverage specialized resources and expertise across the community. In this paper I posit the need for SWAT (software and workstations for antiquated technology) sites: organizations or institutions that are willing to put their expertise to use for the benefit of the broader community by providing specialized services to institutions with limited resources.

NOTE: The archiving of time-based media, such as audio and video, is not addressed in this brief consideration.

The Daunting Variety of Physical Media

In general, transferring content from a particular physical medium requires not only a compatible computer with a device to read the data in the format that is stored on the medium, but also the appropriate ports, cables, and drivers needed to connect to and run the device.

A variety of older media exists and its related hardware and software can be difficult to acquire and maintain. Examples of early set-ups include:

- Commodore PET 4000 with a Commodore 1530 Datasette model C2N and a PCB edge connector and cable to read cassette tapes
- IBM System/3 computer with an IBM 5496 Data Recorder to read 96-column punched cards
- DEC PDP-10 and DECTape drive to read 5 inch 10 track data tape reels

There is also a lot of variety in current media. Examples of contemporary set-ups include:

- Computer with a DVD drive and ability to read DVDs
- Computer with an SDHC (Secure Digital—High Capacity) card slot or a USB SDHC card reader adapter
- Computer with a SCSI interface and drivers for a Sun/Oracle StorageTek T10000C tape drive to read StorageTek T10000 T2 5 TB tape cartridges

For purposes of this discussion, I'll refer to media such as punched cards, data cassettes, and tape reels as *obsolete* media, and tape cartridges, SD cards, and DVDs, as *current* media. [This said, I can't cleanly lump all media into the obsolete or current categories. Media such as 3½ inch diskettes are no longer current, but it's not complicated to maintain a computer with a 3½ inch disk drive (though over time it will become more challenging). On the other hand, managing the varying types of current back-up tape cartridges may be beyond the capabilities of many repositories, even though they are not yet obsolete.]

Because of the rapid growth of digital information—and because current media formats tend to be in use for a longer period of time than those of the past—at some point, archives will acquire far more current (and future) media than obsolete media. It would make sense, then, for most individual archives to focus on keeping up with current and future media and look to others for help with media they cannot read.

Envisioning SWAT Sites

A community-based approach would use SWAT sites wherein a few self-selected institutions acquire and maintain the gear and expertise to read data and transfer content from particular types of obsolete media. The SWAT sites would provide transfer services for institutions that don't have the capacity to read a particular medium (or the SWAT sites might become the likely places to deposit particular types of media).

A SWAT site might have a wide array of computing platforms, including the software and drives to read many types of media. Alternatively, a SWAT site could specialize in one particular challenge, such as punched cards, early word-processing disks, or Apple II media. Services could be offered to other archives (and, conceivably, to businesses, law enforcement, and individuals), perhaps on a cost-recovery basis—or the service could become a modest revenue center for the institution.

For example, a university with a digital forensics lab and experienced technicians might develop SWAT services. A computer museum might take advantage of the fact that it houses a lot of old equipment and become a SWAT site to generate revenue to support the museum. A hobbyist with a passion for preserving early computer games may have a garage full of old gear and might offer transfer services based on early personal computing equipment. A commercial service might be used to transfer content from tape cartridges at a reasonable price. Members of the community can be on the lookout for the various equipment needed to read unsupported media and get it to the appropriate SWAT site that can put it to use.

Once a SWAT site has transferred the content to more stable storage, the owning archive should be able to manage it from that point forward. The existence of SWAT sites would allow other institutions to focus on the more common current media and keep up with future media.

Considerations when Managing Data Stored on Physical Media:

- Is the information on the media likely to have enduring value?
- Is the medium viable or has it been compromised in any way)?
- Do I have the hardware/software to be able to read the data?
- Do I have the particular operating systems and application software (and sometimes specific versions) needed to understand the file formats and make sense of the data?
- Can I export the data into a non-proprietary form and/or copy it to more stable media?

Developing Your Strategy

As the above discussion suggests, not every repository needs to become expert in converting and managing content from all types of physical media. If an archive collects materials from the late 20th century and beyond, receiving born-digital content will be fairly commonplace,

and it should develop the capacity to process current media routinely. When a repository receives a type of media it does not have the means to read and transfer, it would be advantageous to be able to outsource it to a SWAT site that handles that particular media type. Other options may be worthy of consideration, as well:

- If your collections predominantly consist of analog papers but include the occasional diskette, it may be reasonable to simply have the SWAT site print the content for you so you can store the printouts with the other analog materials.
- Prospective outsourcing of all born-digital content may be prohibitively expensive, but transferring the data on an as-needed basis might be more manageable, depending on user interest and the volatility of the medium.
- For digital materials in collections that would not otherwise receive high priority, and when the opportunity arises, consider working with researchers who have the gear, expertise, and motivation needed to read and transfer the data.
- Consider whether there might be a more appropriate archive for deposit of the material.

As long as libraries and archives remain stymied by the plethora of issues involved in archiving born-digital content and the daunting prospect of having to “do it all,” their progress will continue to be slow. Inaction may impede the course of research and contribute to the loss of important content. By leveraging specialized resources and expertise across the archival and library communities, individual archives and libraries are far more likely to be able to keep up with the onrush of born-digital content and actively further the course of research.

Questions to Consider for Using, Investigating, and Establishing SWAT Sites

How might your institution engage with SWAT sites?

- Are you willing to pay for born digital transfer services that you need, but are unable to provide?
- Do you know of a facility that might provide SWAT services?
- Do you have equipment that you could donate to a SWAT site?
- Could your institution potentially host a SWAT site?

What are the characteristics of a trusted SWAT site?

- What kind of assurances do we need as to who will have access to the data, protection of confidentiality, and secure deletion of any copies created on servers during the process?
- How should personally identifiable information, encrypted information, and viruses be handled?
- Could a standard agreement be devised to cover these details?
- What deliverables might be desired beyond transfer to a different medium? Full disk images with cryptographic hashes? Provenance metadata associated with the data extraction process? Migration to other file formats? Emulation (where an operating environment recreates that of a certain time period, so materials of that vintage can be run without maintaining the original hardware and software)?

How would SWAT sites make their services known?

- How might SWAT sites cover their costs?
- How could we jumpstart some SWAT sites so they could perform these functions?
- Might donors help with start-up costs?
- Are there institutions that already have the gear and expertise, but need to figure out how to offer the service to others and achieve cost recovery?
- Could cost recovery fees similar to those in many libraries' photoduplication services suffice?
- Might SWAT sites be able to charge enough not just to keep the service running, but to update and expand it?
- Are there other incentives and models that could be considered?

If you are interested in pursuing these topics, please contact Ricky Erway at erway@oclc.org or +1-650-287-2125.