



You've Got to Walk Before You Can Run:

First Steps for Managing Born-Digital Content Received on Physical Media

By Ricky Erway

You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media

Ricky Erway

Senior Program Officer
OCLC Research



© 2012 OCLC Online Computer Library Center, Inc.

This work is licensed under a Creative Commons Attribution 3.0 Unported License.

<http://creativecommons.org/licenses/by/3.0/>



August 2012

OCLC Research

Dublin, Ohio 43017 USA

www.oclc.org

ISBN: 1-55653-508-2 (978-1-55653-508-6)

OCLC (WorldCat): 801595484

Please direct correspondence to:

Ricky Erway

Senior Program Officer

erwayr@oclc.org

Suggested citation:

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>.

Introduction

This is a set of initial steps for getting started with responsible management of born-digital materials that were received on physical media. The intended audience is those who already have or are currently acquiring such born-digital materials, but have not yet begun to manage them. These steps are intended as modest measures to allow those with unprocessed holdings to sleep peacefully, or at least nap without nagging doubt. It is possible to get started responsibly—albeit minimally—without expert staff, extra funding, or extensive training.

While many of the following principles and approaches can be applied to any born-digital content, this document emphasizes characteristics that are specific to born-digital content acquired on physical media. It focuses on identifying and stabilizing your holdings so that you'll be in a position to take additional steps as resources, expertise, and time permit. With relatively few steps and limited resources it is possible to know with more certainty what your repository holds and to make vulnerable data more secure.

You should begin by embracing these four essential principles:

- Do no harm (to the physical media or the content).
- Don't do anything that unnecessarily precludes future action and use.
- Don't let the first two principles be obstacles to action.
- Document what you do.

Survey and Inventory Materials in Your Current Holdings

In the typical scenario, there is a "backlog" of born-digital materials that have merely been stored along with other collection materials without having been copied, preserved, or made accessible.

1. Locate existing holdings.
 - a. Gather information about digital media already in your collections.
 - b. Do a collections inventory to locate computer media in any physical form.
2. Count and describe all identified media. Note that the suggested inventory process does NOT include mounting and viewing content on the media. For each collection:
 - a. Gather relevant information from donor files, acquisition records, the collection itself, and any other useful sources.

- b. Remove the media, but retain the order (if one exists) of the digital media and accompanying items by photographing the digital media and storing printouts of the photographs in the physical collection where each physical medium was originally located. (Alternative: place separator sheets that describe the media into the physical collection.)
 - c. Assign an appropriate inventory number and/or barcode to each physical piece.
 - d. Record the location, inventory number, type of physical medium, and any identifying information found on labels or media such as creator, title, description of contents, and dates. If no identifying information exists, indicate this.
 - e. Record anything that is known about the hardware, operating systems, and software used to create the files. Use consistent terms to help in future processing of the files.
 - f. Count the number of each media type, indicate the maximum capacity of each media type, calculate the total maximum amount of data stored in each medium, and then calculate the overall total for the collection. This will enable you to estimate storage needs, though keep in mind that the media are rarely full, so the estimate will likely be far in excess of the actual storage needed.
 - g. Return the physical media to suitable storage (if not with the rest of the collection, indicate its location in the collection documentation).
 - h. Add a summary description of the digital media to any existing accession record, collection-level record, or finding aid.
3. Prioritize collections for further treatment based on factors such as the following:
 - a. The value, importance, and needs of the collection as a whole—and the level of use (or anticipated use) of the collection
 - b. Whether there is danger of loss of content because of potential media degradation due to age or condition
 - c. Whether there appears to be significant digital content that is not replicated among the analog materials in the collection or elsewhere
 - d. Whether the use of digital content that is replicated in analog form would add measurably to users' ability to analyze or study the content
 - e. When there are just a few files that can be represented on a page; whether printouts might suffice
 4. Repeat these steps every time you receive new media.

Once you have gotten this far, you will have made very real progress in starting to manage your born-digital materials. Knowing what you have in terms of types of physical media, storage requirements, and general significance of the content is an important start.

Technical Steps for Readable Media

As possible, based on your established priorities and resources, begin the steps for born-digital content on media that can be read by available hardware.

1. Use a “clean” computer (a dedicated computer that is regularly scanned with up-to-date antivirus software and that is not used for online activities that may introduce viruses or used for other work that might be affected by viruses introduced when accessing media).
2. Use a write blocker on the clean computer and set any write-protect tabs on the media to prevent changes to the content.
3. Insert disk in the appropriate drive (or other medium in appropriate reader or attach other storage device to appropriate port). Do not attempt to open any files.
4. Create a directory on the clean machine for the current project, with a subdirectory for the data files.
5. Copy data from physical media to the subdirectory. Consider copying the data as a disk image, which is a single file that contains an exact, sector-by-sector bitstream copy of the disk’s content and ensures that various forms of essential metadata and technical dependencies will be retained. Alternatively, directly copy directories and files from the original medium to the subdirectory, but note that various forms of associated data and metadata may not be transferred. Moving the data directly from the original medium into a zip-compressed archive can help to preserve some file system metadata (e.g., timestamps, directory structures, and file permissions).
6. Generate a copy of the disk directory information (file names, sizes, extensions, and dates). Store a digital copy in the project directory and print out a copy to keep with the collection.
7. Generate and record a checksum (a unique value based on the contents of a file) on the disk image. Alternatively, if you copied the files instead of copying a disk image, generate and record a checksum on each file in the subdirectory.
8. Create a readme file containing pertinent information from the above steps, indicating the related analog materials and documenting each step taken. Store the file in the project directory and store a printout of the readme file with the physical collection materials.
9. Copy the project directory to trustworthy archival storage where it will receive regular back up, with a copy stored in another location.
10. Return the original physical media to storage. (This is optional, but it provides one more possibility for back up if something goes wrong. If you choose not to retain the media, discard them responsibly.)
11. Create or update an associated finding aid, collection level record, or accession record with information about the steps that were taken and the location of the files.

The guidance above will help you to take responsible, basic steps to preserve your digital media. You’ll want to do more when you can and should refer to guidance on managing and preserving born-digital material that is available elsewhere. Ultimately, your work will lead to making born-digital materials accessible for use. After all, the goal is not just to preserve the content on digital media, but to provide access to it.

For more information about our work regarding born digital materials, please visit:
<http://www.oclc.org/research/publications/library/born-digital-reports.html>



6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878

T: +1-614-764-6000

F: +1-614-764-6096

www.oclc.org/research

ISBN: 1-55653-508-2
978-1-55653-508-6
1208/215338, OCLC