



**Thoughts following a very small leap:
on a selective survey of born digital materials
held by the Houghton Library**

Melanie Wisner
Accessioning Archivist
Houghton Library
Harvard University

Preamble

As accessioner of manuscript and archival material for up to six curatorial departments for the last 13 years, I have watched a relatively slow trickle of born-digital materials into our still primarily paper-based collections. We are a small shop in terms of digital holdings for a good reason: not being able to manage these materials, we have chosen not to acquire them (when given the choice). The corollary to this principle is that we have, till now, taken the path of least resistance concerning the born digital material that *does* come in: keep everything but take no action. We can't and don't provide much description of these items.

Colleagues at several other repositories across the University have been more proactive in the born-digital realm, from choosing to stockpile antique equipment to obtaining state of the art digital forensics workstations, and, as of this writing, Cal Lee is slated to talk about digital forensics to our broad library community in a few days. But we have not yet acted proactively *collectively* across the University.

The Jump In Initiative seemed, therefore, the perfect low-barrier, do-no-harm action for one person in one repository to take. It would be a good opportunity to put one toe into the sea of policy and practice surrounding born-digital without committing funds, staff time, or born-digital media to irreversible actions—and without the burden of high expectation.

Hereafter, born-digital will be referred to as "BD".

Preparing the survey

I obtained permission to inventory BD media within 21 selected accessioned gift or purchase collections from the last six or seven years from three curatorial departments. (There were originally to be 20 items on the list, but an artist's book on a USB flash drive was too interesting to omit.) The 21 selected were known to have digital content because we started to track it informally via a 500 note in our MARC accession records about ten years ago. We are not yet equipped to begin to take any technical steps such as determining if the media are still readable.

The process of focusing on the attributes of disks assembled together has been interesting. I had to face my own, and, by extension, my repository's, low valuation of these items. We do many things to keep paper happy and strive not to alter its state (unless to undo damage). With digital media we have not applied similar finesse. I know from experience that the prevailing attitude is that a disk either works when inserted into a computer or not; there is little awareness of the possible shadings from "original integrity" through alterations to damage, much less awareness of our agency in the alterations, something that I plan to take more seriously going forward in order to protect media content.

This project has been a process of approaching the details carefully and watching for surprises. The artist's book is clearly the most thought-provoking item in the collections I surveyed. It is intended to be "read" in a certain environment with specific equipment—an environment I am fairly sure we are not equipped to create. To make matters weedier, the creation of its files is not well-documented. I could not help ruminating on this as a test case from the future of special collections librarianship. As has been noted by archivists globally, we will need to look to the museum world and many other domains for help and collaboration.

The mechanics of the survey

I devised an ID number system (with removable labels) for identifying and marking the media as they were removed from their home collections. (Hesitant at first to remove the media, I understood how much more efficient this rendered the project once I started surveying.) I labeled and photographed the media, placing a printed out image, with ID, in place of the removed media.

I established a project spreadsheet with multiple sheets: a list of the collections; the itemized survey; a content ranking system; timekeeping; project notes; and summary results.

The survey itself, disk after disk, went relatively quickly. The bulk of my project time was taken up thinking about procedures for identifying, tracking, and evaluating items. Not knowing what to track for what unforeseen purpose, I found the process to be iterative: examination of each new item suggested new points to consider and chart. With each variation in media, I consulted Wikipedia to refine OCLC's baseline storage capacity numbers (and expose myself to some electronic media history, hoping that I have finally found a way to remember which way the tab goes on a floppy disk that is write-protected).

I worked and reworked a content ranking system which I intended as an indicator of desirability of reformatting. It had to be simple, easy to remember and apply; but getting to simple was very, very hard. I came up with a ranking from 0 to 4 that is simple but deceptive: it requires more thought than I think would work in a production setting. I would definitely seek input from colleagues to improve it.

Results and leads

Within 21 collections, there were 149 individual items (108 of them unique) for a total of 348.5 GB. There were eight categories of media including floppy disks, various kinds of CDs and DVDs, Zip disks, and two USB flash drives. I spent roughly four work days on the project.

After the novelty wore off, the project resolved into the usual archival tasks: a small amount of preservation (new CD cases, etc.), arranging (type, commercial vs. not), and a lot of thinking about appraisal.

Seeing the number of commercial disks in a few collections, many still in original wrappers, points out to me that there are a few aspects of work with BD media that we can and should implement immediately—such as a guideline for disposition of published works (just as we do with print). We should also be able to add summary notes to our accession records and box lists for collections with BD material. I will think about preprinted labels for envelopes to house incoming BD media to warn staff and readers not to attempt to read the media...or some such action to prevent unwitting alterations.

In the end, I was left in each case with a familiar archival question: what is the meaning of this item within this collection? Based on that evaluation, what level of access should we try to provide? Is there any way to preserve an artist's multimedia book on a flash drive as it was intended to be presented (and is someone else better prepared to interpret it)?

I look forward to any and all discussions that derive from this project locally, within the Section, and in the literature. I thank the Houghton Library's leadership team as well as the Section's leadership for pushing us forward.