

## Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in Manuscript Repositories

---

A FEW YEARS AGO I attended a lecture by a well-known digital libraries researcher who, when discussing the issue of digital preservation, clicked forward to a PowerPoint slide that included Pablo Picasso's iconic ink illustration of Don Quixote, the placement of which seemed a little too conveniently symbolic to be accidental. Regardless of the presenter's intent, it is clear that the quest for sound and simple processes to support the management and preservation of electronic records—now often referred to as “born-digital” records—has remained a Quixotic one for many manuscripts curators and archivists, even more than two decades into the era of personal computers. While we continue to wait for that one perfect, affordable, all-encompassing solution for electronic records, digital materials already in our possession deteriorate, and the size of the digital universe grows ever larger.<sup>1</sup>

This circumstance is surprising when you consider how many projects, articles, and conference presentations (not to mention millions of dollars in funding) have been devoted to the issue of electronic records over the years. As evidence builds that our profession is quickly acquiring a broad range of born-digital materials without much thought given to next steps, perhaps what is missing from the literature is an articulation of the beginning actions that repositories can take to deal with the growing born-digital backlog in our holdings. Despite our profession's still-tentative relationship with technology, institutions must begin to take responsibility for born-digital material already entrusted to their care. Separating digital files from the impermanent media they were donated on is the most critical step, but we must also attempt to capture some essential information about the digital files acquired and begin to develop policies to guide us forward. This article will not aim to provide any across-the-board solutions to the problem of

---

1. The size of the digital universe (that is, all information that exists in digital form) is estimated by the International Data Corporation to be 1.8 zettabytes (or 1.8 million terabytes) in 2011, half of which will not find permanent storage. For more information, see: John F. Gantz et al., “The Diverse and Exploding Digital Universe,” Executive Summary, International Data Corporation, Mar. 2008. Available online at [www.emc.com/collateral/analyst-reports/diverse-exploding-idc-exec-summary.pdf](http://www.emc.com/collateral/analyst-reports/diverse-exploding-idc-exec-summary.pdf). [Accessed 6 April 2011].

born-digital materials but instead will seek to make some basic recommendations to begin.

### **Straddling the Fence: Where We Are Now**

Certainly there are a range of critical issues confronting special collections and archives in this age of digital production. But of most immediate importance, it would seem to me, is the active acquisition of born-digital material without formal plans for the ongoing management and preservation of these materials. A number of recent studies confirm that manuscript repositories are straddling this fence between stewardship and neglect.

A 2008 *American Archivist* article reported on the findings of a study that examined the state of electronic records in 125 collecting repositories from across the United States, with the majority (roughly 86%) of respondents being from academic institutions. More than 80 of the institutions surveyed had already acquired or would accept born-digital materials; but, of those institutions, only 30 had a policy governing the acquisition of such material, while a mere 15 had digital preservation policies in place, leading the author to conclude that institutions “are incorporating born-digital records and papers into their collections without necessarily altering existing policies to do so.”<sup>2</sup>

A similar survey by the Society of American Archivists’ Congressional Papers Roundtable (CPR) Electronic Records task force found that electronic records existed in the congressional collections of 36 of 46 institutions surveyed, with most reporting no policies in place governing the acquisition, management, or preservation of electronic records, most unable to access a significant portion of the records, and over half not even capable of estimating the extent of the electronic records in their collections.<sup>3</sup> This is especially troubling when you consider that congressional collections are some of the largest distinct collections in archival repositories.

The recently published survey of special collections and ARL research libraries by OCLC Research has only added to this chorus, with 79 percent of 169 respondents acknowledging the existence of born-digital material in their collections, a percentage that stood, as the authors note, “in stark contrast to the 35% who reported the size of their born-digital holdings.”<sup>4</sup> In summing up their findings

2. Susan E. Davis, “Electronic Records Planning in ‘Collecting’ Repositories,” *American Archivist* 71 (Spring/Summer 2008): 167–89.

3. Society of American Archivists Congressional Papers Roundtable, “Congressional Papers Roundtable News,” Spring/Summer 2010. Available online at [www.archivists.org/saagroups/cpr/newsletters/Spring%202010.pdf](http://www.archivists.org/saagroups/cpr/newsletters/Spring%202010.pdf). [Accessed 6 April 2011].

4. Jackie M. Dooley and Katherine Luce, “Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives,” *OCLC Research*, Oct. 2010. Available online at [www.oclc.org/research/publications/library/2010/2010-11.pdf](http://www.oclc.org/research/publications/library/2010/2010-11.pdf). [Accessed 6 April 2011].

prior to the release of the report, the authors called born-digital material “under-collected, undercounted, undermanaged, and inaccessible.”<sup>5</sup>

While it is encouraging to find so many institutions see the value of collecting born-digital materials, it is disconcerting to consider the state these records must be in if most institutions cannot even estimate how much of them are present in their collections, have given no thought to what kinds of records they acquire, and have not developed policies addressing how to manage and preserve them over time. Born-digital records *cannot* be managed using existing policies for analog materials, nor can they be neglected for very long. Procedures for responding to the deterioration of paper material often include long periods of inattention, which are afforded by the relative stability of the medium. This is not the case with digital materials. Elizabeth Dow sums up the distinction well in her excellent monograph, “Electronic Records in the Collecting Repository”:

[I]f five years go by before the repository initiates some form of intervention to rescue deteriorating newspapers or books, the information they contain won’t disappear. The curators can respond to the materials on the curators’ schedule. The exact opposite is true with digital materials. Five years can mean the difference between having the information ... and not having it. The curators must respond to the materials on the materials’ schedule.<sup>6</sup>

Why has our profession done so little with these materials? Many curators and archivists feel overwhelmed by the technology, while technical support and fluency are frequently mentioned as stumbling blocks in these studies. The CPR survey found that many archival units do not have a staff member dedicated to electronic records issues,<sup>7</sup> while the OCLC report cited lack of expertise as one of the three largest impediments.<sup>8</sup> The *American Archivist* article notes that having IT support “increases the chances that a repository accepts born-digital records,” but evidently it does not increase the chances of having established policies for managing or preserving these records once they arrive at the door.<sup>9</sup> Nonetheless, it is interesting to note in this article that 40 percent of the repositories surveyed report having technical resources available *within* the archival unit, with over two thirds having

---

5. Jackie M. Dooley, “Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives,” presentation at the 51st Annual ACRL Rare Books & Manuscripts Section (RBMS) Preconference, June 2010, Philadelphia, Pa.

6. Elizabeth H. Dow, *Electronic Records in the Manuscript Repository* (Lanham, Md.: Scarecrow Press, 2009).

7. Society of American Archivists Congressional Papers Roundtable, “Congressional Papers Roundtable News,” Spring/Summer 2010. Available online at [www.archivists.org/saagroups/cpr/newsletters/Spring%202010.pdf](http://www.archivists.org/saagroups/cpr/newsletters/Spring%202010.pdf). [Accessed 6 April 2011].

8. Dooley and Luce, “Taking Our Pulse,” *OCLC Research*, Oct. 2010.

9. Davis, “Electronic Records Planning in ‘Collecting’ Repositories.”

technical support at their disposal within the larger institution, which suggests that technical resources are more abundant than one would assume based on the dearth of organized activities in support of electronic records.<sup>10</sup> At minimum, these findings suggest that having technical resources has not been enough to get us where we need to be.

Lack of funding is noted in the OCLC study as the most crucial impediment to managing born-digital materials.<sup>11</sup> The assumption is that managing born-digital materials requires additional funding for hardware and software, ample data storage, and the employment and training of staff. There is no doubt this is true, though the degree to which it is true depends also on institutional will and the strategies pursued. Many institutions are using inexpensive external hard drives for storage, while free open-source software is bountiful. Funding for technology would seem to be more agreeable to institutions than funding for staff, since software and hardware tend to be more affordable than funding salaries and benefits. While all repositories, especially in this economic climate, are likely struggling to keep up with all areas of administration, if no additional funds for staffing are available then we must take steps to enlist existing personnel in tackling this issue. Lack of staffing cannot excuse the inability of repositories to manage born-digital materials they have already acquired and have therefore accepted responsibility for preserving and making accessible. It boggles the mind that 80 percent of the institutions surveyed in the OCLC study had already acquired born-digital records, when nearly half (45%) responded that responsibility for born-digital material was “not formally determined” or “not yet addressed.”<sup>12</sup>

Perhaps, as well, there remains in our profession some reluctance to consider any form of digital preservation as true preservation. An unavoidable truth is that born-digital records necessarily narrow our concept of preservation. Used to thinking of preservation in terms of centuries, curators and archivists are now struggling to ensure that electronic records will be accessible mere decades from now, a reality that might make printing to analog seem the most desirable solution. While printing should remain one option to consider, especially, perhaps, for our most valuable “four corner” documents, it can hardly be employed universally to all born-digital records. There is not enough paper in the world to print, en masse, all the electronic records we have acquired (and will likely acquire in the future), nor would the solution even be appropriate for more complex types of digital files, such as databases, Web sites, or multimedia.

---

10. *Ibid.*

11. *Ibid.*

12. Dooley and Luce, “Taking Our Pulse,” *OCLC Research*, Oct. 2010.

Still, there is some merit to the notion that technology alone cannot solve problems associated with digital preservation, as evidenced by the recent Blue Ribbon Task Force report on sustainable digital preservation and access. Just as important, if not more, is a sustained effort at mobilizing a diverse array of stakeholders.<sup>13</sup> This means getting our own archives and library staff to the table for discussions, and it includes attempting to develop decision and policy frameworks for at least examining solutions. And the solutions need not be complex, which should make them more easily achievable with limited resources. Arguably, waiting for high-end solutions has only led to neglect. Instead of waiting on the right personnel, enough money, and comprehensive technical solutions, practical action should be taken now to bridge this gap between neglect and stewardship.

Knowing that institutions are fully engaged in the ad hoc collecting of born-digital material begs the question: what is happening with all this material after it is acquired? A survey of our own institutions might answer this question. My own repository had no defined policies or procedures for electronic records; yet we, like the 70 percent of institutions surveyed in the *American Archivist* article, have been actively acquiring electronic records for years. In a very few cases—as in the acquisition of the university president’s e-mail, or the archiving of topical Web sites—we have done so in ways that aligned with accepted standards, though no program or process had been formalized for wider collecting. We were, as the previously mentioned studies suggest, making decisions on a case-by-case basis.

In most situations, the bulk of born-digital material had come to us on removable media such as floppy disks, CDs, and DVDs, which were deposited in boxes and left in the stacks, untouched, unmanaged, bit-rotting slowly over many years (and perhaps for years before being acquired). This is what some in our profession colloquially refer to as the “disk-in-a-box” problem, and I suspect that, for most of the institutions collecting born-digital material without any devoted policies or procedures, this situation is quite commonplace.

### **A Way Forward: Narrowing the Lens**

When my institution, the American Heritage Center at the University of Wyoming, began internal discussions on the topic of born-digital collections, we found the scope and complexity of recommended solutions overwhelming. Consuming enough of it, a curator’s first question is unavoidably, *now where do I actually start?* We began by discussing, for example, how best to acquire electronic records, how to

---

13. Blue Ribbon Task Force on Sustainable Digital Preservation and Access, *Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information*. La Jolla, Calif.: Blue Ribbon Task Force on Sustainable Digital Preservation and Access (Francine Berman and Brian Lavoie, co-chairs), 2010. Available online at [http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf). [Accessed 6 April 2011].

preserve and process them, and how, ultimately, to make them available to researchers. Besides the fact that some of these questions put the proverbial cart before the horse, there was also the problem that such questions were highly speculative, attempting to address in the abstract donors, collections, and scenarios that had not yet materialized. Simpler questions to answer were: what electronic records do we have already; what state are they presently in; and what should we do with them?

Once we narrowed our lens, the way forward quickly came into focus, with our guiding principle being that, *at the very least*, we must separate electronic files from the physical media it arrived on. The other questions are irrelevant unless and until we can deal with this disk-in-a-box problem: we cannot manage or preserve born-digital material, let alone adequately accomplish basic administrative tasks such as appraisal, arrangement, and description, nor even provide adequate access to it, unless we can separate the disks from the intellectual content they contained.

As a profession that lacks the full resources and expertise to address this issue, perhaps our short-range efforts are better spent addressing this problem in more manageable chunks, in moving forward with practical and achievable steps, and in developing the institutional framework to tackle the issue with greater complexity at some future point. Let us not focus, then, on finding the path from point A (no electronic records program) to point Z (a fully realized and implemented solution). Instead, let us focus on making our way responsibly to point B. Let us begin the necessary dialogue both internally and with necessary collaborators, and we can use what we learn from this process to lay the groundwork for future policy-making.

In support of this more pragmatic philosophy, we might consider the following general steps, which I think most organizations, even those with limited resources, can pursue to varying degrees:

- inventory existing born-digital material and estimate the number of bytes;
- implement a storage solution, however imperfect, with the help or input of *whatever* IT resources are available;
- transfer records from disks to storage, while capturing authenticity information and documenting your activities; and
- begin formulating policies for future acquisition and preservation activities.

### ***Inventory Existing Born-Digital Material and Estimate the Number of Bytes***

Archives and libraries cannot begin to plan for the storage of born-digital material already acquired unless they know how much is there, so an essential first step is producing an inventory. Despite some recent examples of institutions acquiring

whole PCs, it is likely that most born-digital material still arrives on impermanent, removable media such as CDs, DVDs, and floppy disks. For disk media it is exceedingly simple to estimate the maximum extent (in bytes) of digital files and requires nothing more technical than simple arithmetic. I say *maximum* because, to begin with, institutions should not attempt to inventory the actual files on disks. Before inserting disks into computers, it is best to have some idea of what is meant by authenticity and the various steps that must be taken to avoid compromising it (which will be discussed below). Instead, the extent can be calculated by adding up the different disk formats present and multiplying by the maximum size of the disk. Common sizes include: 1.2 megabytes for a 5.25-inch floppy disk; 1.44 megabytes for a 3.5-floppy disk; and 700 megabytes for compact discs.<sup>14</sup> Of course, adding up the number of these formats will not provide an accurate number of bytes, since it is likely not all disk space on the media has been fully used. The idea, though, is to get a rough estimate on institutional digital storage requirements.

### *Implement a Storage Solution, However Imperfect*

With an estimated number of bytes in hand, repositories can begin to explore storage options. The corpus on electronic records tends, when discussing storage, to focus on trusted digital repositories (TDRs) and applications that follow the Reference Model for an Open Archival Information System (OAIS model). While these are useful models to learn and follow, I think they should be considered long-term goals. What many institutions need are short-term strategies that might ultimately serve as a bridge to those more developed future strategies. So, while an implementation of digital repository software like Fedora or DSpace might be more representative of the OAIS model, for some institutions simple network file storage might be a better start; for others, local file storage on external media might be an adequate short-term solution until more resources are available.<sup>15</sup>

It is essential to understand that, at minimum, two separate instances of storage are required: one for archival masters, and one for access copies. Copies of every electronic file should be transferred from the original media to these two separate storage locations. The first copy is considered the archival "master" and should be completely restricted, never accessed or opened by anyone, not even repository staff. To support the eventual use by staff and researchers, a second copy—the "use" copy—is required. From this use copy, repositories might produce additional

---

14. These are estimates based on the most common disk formats. It's impossible to provide here a full accounting of the various types of media likely present in repositories. In most cases, it should be simple to identify the format and perform a quick Google search for the disk capacity. Variations in disk space also exist within specific disk formats. For additional information on various floppy disk capacities, see: [http://en.wikipedia.org/wiki/Floppy\\_disk](http://en.wikipedia.org/wiki/Floppy_disk). [Accessed 6 April 2011].

15. For an excellent examination of TDRs and whether or not institutions should implement one, I recommend chapter 5 of Elizabeth H. Dow's book, *Electronic Records in the Manuscript Repository*.

copies on demand for researchers or perform preservation activities such as migration to newer formats in years to come. A separate access storage environment is also essential so the born-digital files may be available for common administrative steps—for example, arrangement, description, or disposal. All of these activities could produce unwanted alteration or corruption, which is why separate “master” and “access” storage locations are warranted: should inadvertent alteration occur to the access copy, it can be restored from the master copy. And since digital files are so easily susceptible to corruption or hardware failure, backup of these storage environments is highly recommended; at the *very* least, the master storage should be backed up to a separate location to prevent any loss of data (though a backup of the access copies is also desirable).

For those keeping track, this dual storage framework, with the attendant backup considerations, means a doubling, tripling, or possibly even quadrupling of the number of bytes counted during the inventory phase.<sup>16</sup>

There are a number of ways to set up storage to support this framework of master and access copies. Networked storage would be the best option, since it can provide a higher level of security and redundancy needed for electronic records, and it is typically managed by IT professionals. For most academic repositories, I suspect these resources will be housed within the larger institutional setting (such as the academic library system) or perhaps available at the universitywide level. Since most IT data centers have established backup procedures, those institutions with the luxury of organizing storage through a formal IT department may only require from them two separate, equal-sized storage areas, with backup managed through other means and at a much higher level in the information architecture.

If technical support from such departments is insufficient or too expensive, then more affordable, localized options may need to be considered. Institutions with reasonably small byte-counts (in the single terabyte range) can feasibly support storage of existing born-digital materials in the short term by purchasing external hard drives and treating them as rudimentary storage environments. Decent one-terabyte hard drives can be purchased for as little as \$100; and, of course, an institution would need at least three of these to support the master-access-backup framework outlined above. I acknowledge that external hard drives are neither foolproof nor ideal, but they are, undoubtedly, a considerable step up from leaving files on floppy disks.<sup>17</sup>

---

16. It should be noted that, while any amount of redundancy is desirable, remote backup is ideal. Most IT data centers should have remote backup procedures in place to guard against loss from regionalized natural disaster. If using external hard drives, remote backup will, obviously, not be an option. But again, consider such recommendations short-term solutions until such time as better options become available.

17. External hard drives should be regularly booted up—at least once a month—to ensure continued sound functionality.

With an inventory in hand and storage in place, an institution must next ensure that they can play the media found in their collections. Most newer office PCs do not include hardware such as floppy drives, but such machines still do exist and are usually not getting much use. Retired legacy machines and hardware can be found in many libraries and archives—and especially in universities. A recent discussion of this topic on a prominent digital preservation listserv revealed that many archives have access to hardware like floppy drives and zip drives, and that such hardware is sometimes obtainable by locating and querying staff responsible for supporting (and retiring) the hardware.<sup>18</sup>

It is entirely possible that nothing will come of any consultations with technical departments or staff regarding storage and hardware. Nonetheless, it will have been good to have broached these topics with technical staff that will ultimately need to play a role in preserving digital information (whether they realize it or not).

### *Transfer Records from Disks to Storage, While Capturing Authenticity Information and Documenting Your Activities*

In an archival context, authenticity may be loosely defined as digital files “being what they are purported to be.”<sup>19</sup> To prove that files are what they claim to be, creators and curators alike might attempt to capture important provenance and file integrity information, and ideally this information would be bundled together using a standardized metadata schema, such as PREMIS (short for Preservation Metadata: Implementation Strategies, a standard for capturing digital preservation metadata).<sup>20</sup> Many of the documented requirements for authenticity, such as those elucidated by the long-running InterPARES project (International Research on Permanent Authentic Records in Electronic Systems),<sup>21</sup> are long and perhaps already beyond the means of any repository to document. In the short term, I think streamlining what we capture will make documentation more feasible. For the majority of institutions just trying to get started, we might characterize this as complying with the spirit, not the letter, of the law.

The letter of the law might, for example, dictate following the InterPARES

---

18. E-mail exchange on Digital Preservation Google Group, Apr. 2010.

19. Heather MacNeil and Bonnie Mak, “Constructions of Authenticity,” *Library Trends* 56, no. 1 (Summer 2007): 26–52. [Accessed 17 November 2008 from ProjectMUSE database].

20. For more information, see [www.loc.gov/standards/premis/](http://www.loc.gov/standards/premis/). [Accessed 6 April 2011].

21. InterPARES is the International Research on Permanent Authentic Records in Electronic Systems whose mission is aimed, in their own words, “at developing the knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and providing the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity.” For more information, see [www.interpares.org/](http://www.interpares.org/). [Accessed 6 April 2011].

benchmark requirements for authenticity,<sup>22</sup> or the capture of authenticity documentation using a formal, structured metadata language. In reality, though, it is unlikely InterPARES requirements could be met for disks that have sat in boxes for years and probably have less than adequate accompanying chain-of-custody documentation. The time and resources required to manually create XML-encoded metadata for all the born-digital files presently found in archives and libraries simply do not exist. Complying with the spirit, in my opinion, means taking steps toward documenting authenticity through use of checksums, and it includes documenting, in whatever form most practical, all archival actions taken on born-digital materials over time.

We must become familiar with the term “checksums,” which are captured as part of the documentation on authenticity of electronic records.<sup>23</sup> Very simply, a checksum is a digital signature (akin to a fingerprint) in the form of a string of characters that can be generated for any digital file. If the file changes in any way, either due to malicious tampering or just simple deterioration, that digital signature will change, indicating the alteration. A checksum should be captured for all files at the point of transfer from the impermanent media to the established storage environment, and record of these checksums for all files should be retained and secured. In an ideal world, institutions would have a mechanism in place for periodically producing a checksum and comparing the value to the checksum produced at the time of accession into storage. Such a mechanism could then alert archivists or curators to any changes. In reality, though, this is likely to be far off for most institutions. By capturing checksums at the point of accession, repositories will at least have taken a step toward ensuring authenticity of the born-digital materials in their collections and can then work toward more sound practices for ongoing verification later.

A simple way to document some key authenticity information would be to create a spreadsheet with columns for the file name, its checksum value, a date captured, and the name of the person who captured it. Additional information might be captured such as the collection the file came from and the disk it was originally found on. Obviously, manually recording this metadata could be immensely time-consuming and will not scale well over time, but it is likely more achievable to those institutions with limited resources than coding this information into standardized XML files. Repositories might also consider a variety of

---

22. InterPARES 2 Project, “Requirements for Assessing and Maintaining the Authenticity of Electronic Records,” *International Research on Permanent Authentic Records in Electronic Systems*, Mar. 2002. Available online at [www.interpares.org/book/interpares\\_book\\_k\\_app02.pdf](http://www.interpares.org/book/interpares_book_k_app02.pdf). [Accessed 6 April 2011].

23. According to much of the literature, archives and libraries must capture much more information than this to document authenticity. The InterPARES benchmark requirements are a good place to start understanding this issue; however, as I have repeatedly emphasized, reducing these requirements to a manageable core is essential to moving forward, with more comprehensive implementations considered later.

freely available open source applications that can help automate creation of such metadata.<sup>24</sup>

Checksum information is just one part of the metadata that institutions should capture for born-digital materials. In general, any actions taken, from the basic steps outlined here to future preservation activities, should be recorded by the repository. Over time, a simple spreadsheet or word processing file might document ongoing actions that repositories take on digital files, such as migration to other formats for access or preservation (documenting this will serve as a record between two instances of the same record).

### *Begin Formulating Policies for Future Acquisition and Preservation Activities*

Taking action on the born-digital materials already in our collections is an excellent first step, not only because those materials likely are in dire need of attention, but also because implementing a process for these records, however remedial, can inform future collecting activities. Following an examination of the born-digital materials already in our collections, institutions may be able to: identify which media it can and cannot receive digital files on; have a better understanding of its available storage and, therefore, how many additional bytes it can take on through future acquisitions; have a better understanding of the software and formats used to create common types of born-digital records, and whether or not the formats are problematic for the repository; have a better understanding of how easy or difficult it will be to arrange and describe the material. In embarking on a more formal understanding of institutional capabilities and challenges, archivists and curators might also be in a position to influence the state of born-digital material that comes to them in donations. As with analog materials, it would be nice to only acquire what is of documentary value rather than a comprehensive data dump of a donor's computer; it would be equally agreeable to acquire born-digital material that has some amount of order imposed by the creator.

At the American Heritage Center, we are deliberating whether we want to accession born-digital files on impermanent media at all. To the extent possible, it would be preferable to acquire material by transferring from the donor's computer using

---

24. At the AHC we have elected to use an open source tool created by Seth Shaw at the Duke University Archives. The Duke DataAccessioner is a very simple, straightforward interface that most archivists, I believe, would feel comfortable using. With just a couple of clicks, the tool will copy the entire contents of a disk to a storage environment without actually touching the files while also automating the creation of a basic XML document that lists file names and formats of the disk contents, documents the folder structure of the disk, and even captures checksum values for all files. To learn more or download the application, see <http://library.duke.edu/uarchives/about/tools/data-accessioner.html>. [Accessed 6 April 2011]. For a review of the tool at Chris Prom's Practical E-Records blog, see <http://e-records.chrisprom.com/?p=1809>. [Accessed 6 April 2011].

our own external media, such as USB flash drives or external hard drives, meaning that we would acquire no physical media at all, only the digital files themselves. This preference became concrete for us upon examining the electronic records in a recently acquired congressional collection. The collection contained over one hundred disks of various formats (mostly DVD), which we have estimated to contain as much as 350 gigabytes of data. This congressional representative was active from the end of the 1980s to the mid-2000s, which is evident by the prevalence of analog records in the collection. But we know from talking to the staff of a current U.S. Senator that the volume of electronic records being produced on Capitol Hill is growing fast, and we do not want to receive terabytes of data on thousands of disks.

In the case of this active congressional representative, we would also like to receive the born-digital material in regular accessions. As Dow pointed out, the further we get from the point of creation, the harder digital files are to preserve. A process of ongoing consultation and iterative acquisition seems to be in the best interest of the digital materials, but this is a sea change in the relationship between records creators and archival institutions, both of whom are more used to having such discussions toward the end of the creator's career or, in the case of politicians, at the end of perhaps decades of public service. It is difficult and time-consuming navigating these waters of more active engagement with records creators much earlier in the lifecycle of records creation, but under the old regime we risk acquiring digital files that are already beyond our ability to preserve them. Institutions with well-articulated collecting priorities should emphasize these dilemmas when reaching out to potential collection donors. For institutions with established relationships with donors, the time to ask about their digital files is now.

It is not just the medium of transfer we would like to influence, but also the state of the collection when it comes to us. The disks found in the aforementioned congressional collection exhibit an almost complete lack of organization of files and folders. Many of the disks appear to be data dumps, and in many cases file and folder names are not very helpful in determining the contents of the records. As we do with donations of paper collections, we would like to perform some preacquisition appraisal with born-digital material and provide some direction on preparing the materials for transfer to the archive. This has led us to begin drafting a brief document for records creators that outlines steps they can take with their electronic records to prepare them for transfer to an archive. Examples of such documents abound, but an excellent example of such a document is Yale's "Author's Guidelines for Digital Preservation,"<sup>25</sup> which recommends such actions as "save old media and

25. Beinecke Rare Book & Manuscript Library, "Authors' Guidelines for Digital Preservation." Available online at [www.library.yale.edu/~nkuhl/AuthorsGuidelines.pdf](http://www.library.yale.edu/~nkuhl/AuthorsGuidelines.pdf). [Accessed 6 April 2011].

files,” and “name your files consistently.” An explanation of the recommendation “organize your files” reads:

The management of your digital materials can be enhanced if you handle them in groups and organize them in a logical manner. This structure should be consistent with the organization of any paper records you have, or records in other media, so that all records related to the same activity or subject, or of the same type, can be identified as part of one conceptual grouping.

This will not only assist the creator in managing his or her own records, but it will also ease the management of these materials by archivists and curators later.

It may not be possible to avoid acquiring documents in unknown (or even proprietary formats), but in transferring digital files from impermanent media to storage, we can begin to recognize common formats and software used to produce common types of records. As our planning continues, we hope to eventually develop a list of preferred formats for common documentary types and then share that list with potential donors. It can even be elucidated in disposition lists that are shared with donors. Such a list could then form the basis for digital preservation and access plans, a crucial component of any program to manage born-digital materials.<sup>26</sup> At the very least, it would be preferable to learn of any unusual formats or software used by records creators prior to donation.

We should also adjust donor forms to accommodate born-digital materials specifically. Our archive has added a sentence to our Deed of Gift form that authorizes us to take necessary action with born-digital materials in support of preservation and access.

I give consent to the Center to digitally reformat the collection or migrate existing digital content to new technical environments as appropriate for preservation and/or access purposes.<sup>27</sup>

Given that preservation of born-digital content may ultimately require collaboration with IT departments or regional digital preservation consortia, it may be necessary to amend this text to explicitly provide consent for the involvement of external organizations.

---

26. For an immensely helpful example of such a plan, see Chris Prom's Practical E-Records blog, available online at [http://e-records.chrisprom.com/?page\\_id=581](http://e-records.chrisprom.com/?page_id=581). [Accessed 23 March 2011].

27. American Heritage Center, “Deed of Gift” form.

## If We Do Nothing, Failure Is Guaranteed

In a seminal article on appraisal from 1992, Timothy L. Ericson defined the “rim of creative dissatisfaction” as the boundary where two competing worlds collided and presented new difficult problems, a place “where creativity flourished, and gave [people] the ability to adopt new ideas, and solve old problems in new ways.”<sup>28</sup> By any measure, it seems we have inhabited such a rim, where digital technologies challenge our previously held professional assumptions, values, and practices, for almost two decades without sufficient progress toward satisfying solutions. The studies mentioned earlier clearly demonstrate dissatisfaction with where we stand now. What progress has been made is perhaps negated by continued exponential growth in the creation of digital content. We may never catch up. But we will continue to try. We must at least be good stewards of the born-digital material we acquire.

I have purposefully avoided any significant discussion of preservation strategies. There is ample literature elsewhere on this topic, literature that I hope archivists and curators of born-digital material will make time for and consider within the context of their own collections and institutions. What I hope to accomplish with this article is a compartmentalization of the born-digital dilemma suitable for inspiring some action, where even resources are limited. What progress may be found here is modest, but it is progress; and, in the spirit of former Society of American Archivists President Richard Pearce-Moses, I would suggest that modest progress is exactly what we require. In his 2006 presidential address, Pearce-Moses stated:

We need the initiative and drive ... to dive in and begin working with digital materials ... We cannot wait until we have everything figured out. I didn't want to start working with electronic records because I knew there was a real chance of failure. I am enormously grateful to my friend ... who counseled me early on: “Whatever we do, we may fail; but if we do nothing, failure is guaranteed.”<sup>29</sup>

Leaving unattended the born-digital materials already entrusted to us—leaving disks in boxes—is guaranteed failure. If we were going to ignore the material, we might as well not have acquired it. I do not propose that the approach outlined here will be sufficient in the long run; we will clearly need to amplify it, considerably, over time. But we must start somewhere if we hope to avoid losing forever the born-digital material we have already acquired—even if, in getting started, we put off harder decisions for another day.

---

28. Timothy L. Ericson, “‘At the Rim of Creative Dissatisfaction’: Archivists and Acquisition Development,” *Archivaria* 33 (1991–92): 66–77.

29. Richard Pearce-Moses, “Janus in Cyberspace: Archives on the Threshold of a Digital Era,” 2006. Available online at [www.archivists.org/governance/presidential/pearce-moses.asp](http://www.archivists.org/governance/presidential/pearce-moses.asp). [Accessed 6 April 2011]