Challenges to Born-Digital Institutional Archiving: The Case of a New York Art Museum

Abstract

Purpose – The purpose of this paper is to highlight the challenges to born-digital institutional archiving, using a New York Archive Museum (NYAM) as a case.

Design/methodology/approach – The digital record keeping practices at NYAM were studied using three data sources: a) focus groups with staff, totaling 81 individuals, or approximately one-third of all staff, b) analysis of network file storage, and c) analysis of digital records in archival storage, or specifically removable media in acid-free archive boxes.

Findings - This case study indicates that the greatest challenges to born-digital institutional archiving are not necessarily technological but rather social and cultural. Or rather, the challenge is getting individuals to transfer material to a digital archive so that it can undergo the technological transformations needed to ensure its long-term availability. However, transfer is impeded by a variety of factors which can be addressed through education, infrastructure development and proactive appraisal for permanent retention.

Practical implications – This paper highlights the challenges to born-digital institutional archiving, yet notes that these challenges can be overcome by following a multi-pronged approach.

Original value – This paper outlines the challenges to born-digital institutional archiving, which is not often discussed in the literature outside of the context of higher education.

Introduction

In the field of archival studies, there has been growth in the literature related to born-digital archiving. Notable informational and educational resources include the "AIMS Born-Digital Collections" report, Society of American Archivists' Digital Archives Specialist (DAS) curriculum and the BitCurator project (AIMS Workgroup, 2012; SAA, 2014; Lee et al, 2012). One commonality shared by these projects is an emphasis on born-digital archiving for donation-based or acquisition-based collecting, where an archives or special collection needs to confront obsolete born-digital media or file formats included within a donation or acquisition. Although this problem too confronts institutional archives, it is only a small facet of the entire born digital archiving challenge. This paper will explore the challenges to born-digital institutional archiving through the case study of an art museum headquartered in New York, which will be referred to in this paper as NYAM for brevity.

Born-digital institutional archiving refers to the task of selecting, preserving, and providing access to the born-digital documentation created by an institution with historic and legal value. In this case study, the born-digital archiving practices of NYAM were studied from September 2013 through January 2014 with the end goal of helping NYAM plan for a born-digital archive. This paper aims to highlight the challenges to institutional born-digital archiving using NYAM as a case. The digital record keeping practices at this institution were studied using three data sources: a) focus groups with staff, totaling 81 individuals from 26 departments, or approximately one-third of all staff, b) analysis of network file storage, and c) analysis of removable media (e.g., floppy disks). Before the results will be revealed, relevant literature related to institutional born-digital archiving will be discussed, followed by a detailed description of the research methodologies. The paper will conclude with the challenges to born-digital institutional archiving, using NYAM as a case.

Literature Review

"Born-digital" refers to records that originate on computers, and there may (or may not) be an analog equivalent, such as a printout. Born-digital archiving is meant to highlight the difference from collections

created through digitization, which creates surrogates or access copies of materials that originate on paper, film, or some other analog medium. Because there is no analog equivalent, digital archiving also necessitates digital preservation, or "securing the long-term persistence of information in digital form" (Lavoie, 2004).

Particularly salient methods and models used for digital preservation include the use of the Open Archival Information System model (OAIS) and the Trusted Digital Repositories (TDR) framework. The OAIS model (Lavoie, 2004; CCSDS, 2012), has been adopted as a best practice in preserving digital information from a wide variety of institutions, including libraries and archives within colleges, universities and governments. The model necessitates the creation of SIPs (Submission Information Package), AIPs (Archival Information Package), and DIPs (Dissemination Information Package), among other elements (Lee and Tibbo, 2011). The SIP is the version of the information package that is transferred from the creator to the OAIS; the AIP is the version that is stored and preserved by the OAIS; the DIP is the version delivered to the researcher in response to an access request (Lavoie, 2004). The TDR framework, which builds upon the OAIS model by describing the elements needed to make a repository trustworthy (RLG & OCLC, 2002), includes aspects related to technology, resources and host organization (Cornell University Library, 2007). The criteria for trustworthiness are further articulated in the Trusted Repositories Audit and Criteria (TRAC) report, which make the elements needed to ensure trustworthiness even more explicit (OCLC & CRL, 2007).

A wide-variety of initiatives has looked to make digital archiving and preservation practical. In higher education, one such strategy is to create institutional repositories. DSpace, for example, is not only a library of the output of the institution, but also aims to act as "as a preservation archive, keeping this material accessible, and often immediately usable, far into the future" (Smith et al., 2003). Thus, systems such as DSpace provide functionality for researchers to deposit copies of their work for long-term preservation and access. Other institutional repository projects have taken a different approach by allowing users a high-degree of control over the archive, even allowing users to remove items from their archive at-will (Cocciolo, 2010).

Within the context of institutional repositories in higher education, participation by community members has been an ongoing issue. Davis and Connolly (2007) studied community participation of the institutional repository at Cornell University, which they found the use as varying between low and "non-use." Further, a survey of 40 institutions using the DSpace platform as their institutional repository found that the faculty participation rate was 4.6% per archive with a median of 1.9% (Hong, 2008). A limitation with some of the earliest approaches to institutional repositories is that they focused on library goals (such as collecting and preserving scholarly work), and yet did not directly tap into the motivations of faculty members to contribute. One problem noted by Davis and Connolly is that faculty identify with an "international community of researchers working in a narrow discipline" more strongly than their own institution. They quote a Mellon-sponsored study of scholarly communication that found that "approaches that try to 'move' faculty and deeply embedded value systems toward new forms of archival, 'final' publication are destined largely to failure in the short term" (King et al., 2006). In sum, understanding and making full use of the motivations of users is essential to involve community participants. With respect to higher education, Cocciolo (2010) suggests that one strategy to engage faculty users to participate in institutional repositories is through their role as educators and teachers rather than researchers because teaching is usually conducted locally.

Outside of the context of institutional repositories in higher education, other initiatives have been developed to provide digital preservation functionality. This includes within the context of manuscript and special collection repositories, which accession the papers of a person, organization or family and can include born-digital material (Goldman, 2011). Goldman contends that while OAIS and TDR may be worthy long-term goals for manuscript and special collection repositories, the requirements may be too great and such organizations may be better served by starting small with something as simple as network file storage. The inability of small and medium-sized institutions to provide long-term preservation to digital assets is so great that it has spawned new research initiatives interested in finding ways to close this chasm,; for example, the U.S. IMLS-funded POWRR project (Preserving [Digital] Objects with Restricted Resources) (Rinehart and Prud'homme, 2014). Similarly, other initiatives have looked to simplify digital

preservation more generally, such as through the National Digital Stewardship Alliance's level of digital preservation, which provide a simplified criteria for ensuring the trustworthiness of a repository (NDSA, 2013). And lastly, several software has been developed to help create trustworthy repositories, such as Artefactual Systems' Archivematica, Tessela's Preservica, and OCLC's CONTENTdm.[1]

Within the museum archives community, relatively little published research exists that documents strategies for engaging community members (e.g. staff members) in institutional archiving. One strategy used in parts of Europe is to legislate the retention of archival records. For example, museum professionals in Croatia are legally obligated to deposit their digital archival records in a database, including what they refer to as primary documentation (collection information such as provenance and location of a painting) as well as secondary documentary (activities of the museum) (Nikolić Đerić, 2011). In the United States, where such matters usually do not involve the state, a legislative approach is highly unlikely and thus museums need to find ways to intrinsically motivate staff to participate. In conducting an electronic records management project for the Brooklyn Museum, O'Connell (2012) notes that staff are "eager to begin transferring permanent files to the Archives electronically"; however, she found that "electronic files at the Museum are copious, generally poorly organized, and at risk of deletion or obsolescence" (p. 5). Thus, the amount of work needed to transfer permanent digital records to the archives could dampen that enthusiasm.

Case Methodology

Case study context

NYAM collects contemporary and modern art, produces approximately a dozen exhibitions each year, has over 50,000 square feet of gallery space, and welcomes in the order of a million visitors annually. NYAM has maintained archives since the 1970s, and currently houses approximately 7,000 cubic feet of paper records. The records include exhibition files, artist files, as well as other historical records, and are open to researchers throughout the year on a request basis. In 2005, the department head of the archives established a records retention schedule for the institution that was approved by the Board of Trustees. The records schedule identified groups of records by department that had permanent as well as temporary retention periods, and highlighted vital records for the operation of the museum. This schedule was applied to paper records and accessions of such records continue to this day. However, because of the lack of infrastructures or workflows, digital records were not accessioned into the archives.

The Information Technology (IT) department maintains a networked file system that is interconnected via a wide area network across several locations in New York City. Although official IT policy does not state how long files are maintained, the strategy demonstrated by IT staff is to maintain all files on network storage until deleted by a user. When disk space limits have been met, then additional storage space is purchased.

This investigation occurred during the course of a small grant project to help NYAM plan for an institutional born-digital archives. Digital record keeping practices were studied using three data sources: a) focus groups with staff, b) analysis of network file storage, and c) analysis of digital records in archival storage, or specifically removable media in acid-free archive boxes. Details of each of one are discussed below.

Focus Groups

Staff in every department participated in a focus group, resulting in focus groups with 26 departments totaling 81 attendees, or approximately one-third of all staff. The goal of the focus groups was to discover how staff maintain their electronic records, especially those records that have permanent value, which can be identified by the institution's records retention schedule, as well as those records that might be at risk (e.g., obsolete file formats, on obsolete removable media, etc.).

Staff were invited to attend a focus group session for their department. In most cases, heads of the departments were invited, as well as those who may be familiar with the electronic records produced within

a department. Additionally, all staff having occupying key roles, such as Curators and Conservators, were invited. Meetings were held at respective department work addresses to make departmental participation as effortless as possible.

Questions, handouts and presentation materials for use during the sessions were assembled and the focus group protocol was piloted with the Library and Archives department. Updates to questions were made. A sample handout given to attendees is included in the Appendix. Focus groups were scheduled for one hour in duration and consent to audio record them was granted in all cases. Significant quotes were transcribed from each departmental meeting and notes were assembled.

Following the initial introductions and background on the project participants were then asked to turn their attention to the records they create with respect to their departmental and the institute-wide records retention schedules.

After discussing the records and record groups created within a department, the discussion turned to where these records are being stored. During the piloting of the focus group, it was decided that one way to talk about departmental network shared drives was in terms of "Mystery Folders," or records on network drives that may be at-risk because no one in the department is responsible for them. After this discussion, other locations of electronic records were discussed, such as specialized databases, removable media, cloud storage, among other potential locations. Lastly discussed were the file formats, with an emphasis on any formats that may be unusual, such as files that do not originate in MS Office, PDF, Adobe Creative Suite, or graphic files such as TIF or JPG.

Network File Storage Analysis

In addition to the focus groups, electronic records contained within networked file storage were studied. The objective was to uncover the extent to which network storage included inactive content with historical or legal value that would be well suited for transfer to the digital archives repository. A meeting was held with the Chief Information Officer was held to uncover the basic setup of the network file storage as well as the procedures associated with it, such as backup. In addition, read-only access to most of the departmental shared drives was granted.

Through discussions with the Director of Library and Archives as well as the Assistant Archivist, it was decided that investigating network content that was last modified over 10 years ago may be particularly effective, since this content may well be inactive, and some of it may have historic or legal value. Additionally, a portion of the records may be at risk because they originate in obsolete file formats (e.g., WordPerfect, QuarkXPress), or are no longer part of an active department with staff to monitor them and know about their contents.

Ideally, this analysis would take into account the file "date last accessed." Unfortunately, no file had a date last accessed before March of 2011. This piece of metadata is likely incorrect and resulted from a file migration from one shared drive or server to another. Fortunately, the "date last modified" is intact for all network files.

To study the network file shared drives, the software program TreeSize Professional v6.0 was used to export lists of files over 10 years and age, and produce graphics that illustrate the extent of older files and file types.[2]

Removable Media Research Methodology

To study the extent of removable media (CDs, DVDs, floppy disks, etc.) held at the institution, such content was sought out in two locations: a) held by staff within their work areas, and b) held within the institutional archives. To study removable media held by staff within their work areas, questions about removable media were asked during the focus groups. To study removable media held with the institutional archives, documentation relating to both processed and unprocessed archival collections was studied. The institution includes a box inventory of their unprocessed collections maintained in a Microsoft

Access database, which was used primarily for uncovering unprocessed collections with removable media. For processed collections, the online finding aids, combined with the use of the MS Access box inventory, were used.

Results

Staff in every department participated in a focus group. Additionally, 8.6 TB of network storage were analyzed using TreeSize Pro to identify the types of files used as well as the age of files stored on network storage. And lastly, processed and unprocessed archival collections (approximately 7,000 cubic feet) were studied for the extent to which they contained electronic records with permanent value using inventories and finding aids.

From studying these three sources of born-digital records, a number of challenges to institutional born-digital archiving were uncovered.

A major finding is that born-digital records are stored primarily on network file storage. Staff generally do not transfer electronic records to the archives because of lack of established procedures, workflows or infrastructures for doing so. Because electronic records do not occupy valuable physical workspace, staff have little motivation to identify records for permanent retention based on their departmental records retention schedule. For example, large caches of files from the 1990s can be found on the network. Although files greater than ten years old (based on date last modified) represent only 10.1% of all network files, this is still 276,028 files, which is a substantial number of files to appraise for permanent retention (see Figure 1).

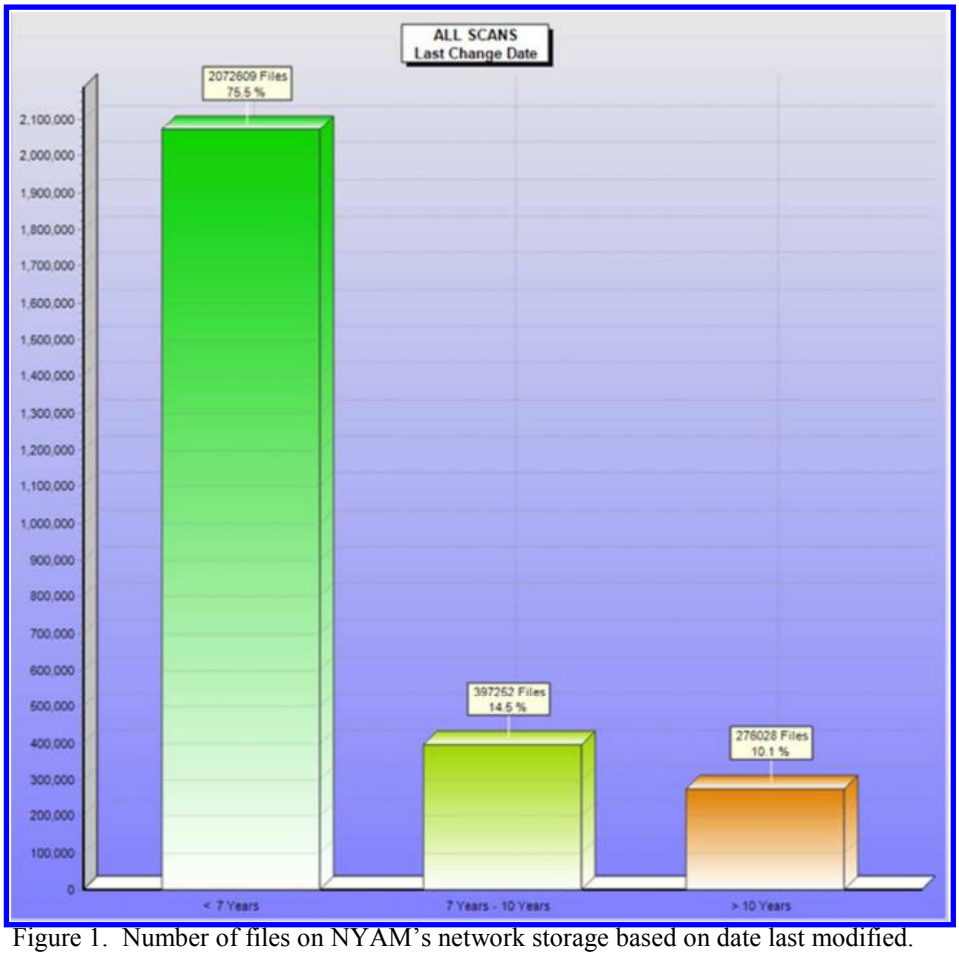

Figure 1. Number of files on NYAM's network storage based on date last modified.

Further, there is variability in how departments treat their records. Some departments with extensive collections of older records view these collections as "archives" of their department's work, and exhibit a somewhat proprietary attachment to them. They prefer to maintain their records on their departmental network shared drive. However, they are not sure if the files can be accessed because they may originate in obsolete file formats. Despite this trend, departments with a long-standing tradition of transferring inactive paper records—such as the Curatorial department—exhibit less of a proprietary attachment to their records and a greater willingness to transfer electronic records. Thus, practices in the physical paper-based world impact staff members' interest in and willingness to participate in born-digital institutional archiving.

Specialized database software, such as The Museum System (TMS) and MediaBeacon, increasingly store vital records to the institution. TMS is used to store information about artworks in the collection, and includes vital information such as the location of a piece of artwork.[3] MediaBeacon, a digital asset management system (DAM), was initially used to store the digital work of the Photography Department, but has grown to include work from other departments and has a staff member dedicated to managing it.[4] Although staff are generally satisfied with these software systems, research indicates that staff trust in these systems—both the software itself and its maintenance—varies by specific software, department and individual staff member. For example, mentions were made by staff registrars that they continue to print-out records from TMS because their trust in the system and its maintenance is not one-hundred percent complete. One staff member made sure to mention that the DAM had failed at some point in the past and all its content was lost.

Absent from networked file storage and most DAMs are file format obsolescence monitoring tools. Ideally, files ought to be migrated to formats well suited for long-term preservation, such as those that are openly documented or originate in open formats such as XML (Todd, 2009). These files are often not based on open formats, such as Unicode text or XML, but might have a proprietary binary structure, which is an arrangement of zeros and ones that is not known outside of the company who created the software. Without knowing how to decode the data (e.g., first byte represents this, the second byte represents that, etc.), users are reliant on the software manufacturer to continue to support the format, or attempt to reverse engineer the format. If open formats do not exist for a particular file format, or the migration to an open format is not entirely successful, then best practice is that the original software should be preserved so that files can be opened. This preserved software could be run in emulated environments (e.g., a modern computer emulating Windows 3.1) (Granger, 2000).

Also absent from network file storage and most specialized database software are checks for hardware or software failures that could render a file inaccessible or inaccurate, often referred to as bit-rot or bit-flipping (Rosenthal, 2010). This is usually addressed through running a file fixity check, such as applying the MD5 checksum generator to a file upon ingestion, and verifying the checksum remains the same through time (Barrera-Gomez and Erway, 2013). If the file changes somehow (a bit is flipped), the file can be restored from backup.

At NYAM, staff do not necessarily see planning for long-term preservation and access, such as through format migration and integrity checks, as a prerequisite for considering something a digital archive. Systems designed for access, such as item-level digital libraries or DAMs, as well as networked file systems, will be referred to as "archives." There was a high degree of variability around what is known about the contents of departmental network shared drives. For example, some departments know exactly what everything is, and others have many "mystery folders." Although this corresponds to some extent with the stability of the staffing within the department, knowledge of the past three of more years is generally vague even for staff with extended tenures. Many departments have records in obsolete file formats (e.g., WordPerfect, Lotus 1-2-3, Quark XPress). However, since these files are rarely if ever referenced and thus not barriers to completing work, this issue is not a priority for any department. For example, on network storage there were 11,694 WordPerfect files, 5,194 QuarkXPress files and 1,430 Lotus 1-2-3 files.

Certain file types were expected, and the focus groups and network scans confirmed their salience; e.g. MS Office files, Adobe PDF and Creative Suite files as well as graphic files (JPG, TIFF). Figure 2 shows the breakdown of file types spread across all of network storage. This figure notably illustrates that there are

certain files that were in abundance (e.g., .jpg, doc), but dozens of file types that may only have had one or two files of that type.

Although not initially expected was the importance of two-dimensional drawing (2D) and three-dimensional (3D) modeling file formats (VectorWorks, AutoCAD, and Rhino). For example, there are 19,660 VectorWorks documents (2D drawings), 7,290 AutoCAD files (2D drawings), and 6,246 Rhino files (3D drawings). These software are used for designing and building exhibitions. Research indicates that there is a not an open format for 3D models; thus, the software itself would need to be preserved (Smith, 2008). Further, VectorWorks is a proprietary format that requires a costly piece of software to run, and thus such files should be migrated to AutoCAD. Although AutoCAD is also proprietary and costly, Autodesk has made free software available for viewing AutoCAD drawings,[5] and there is some open documentation.[6]

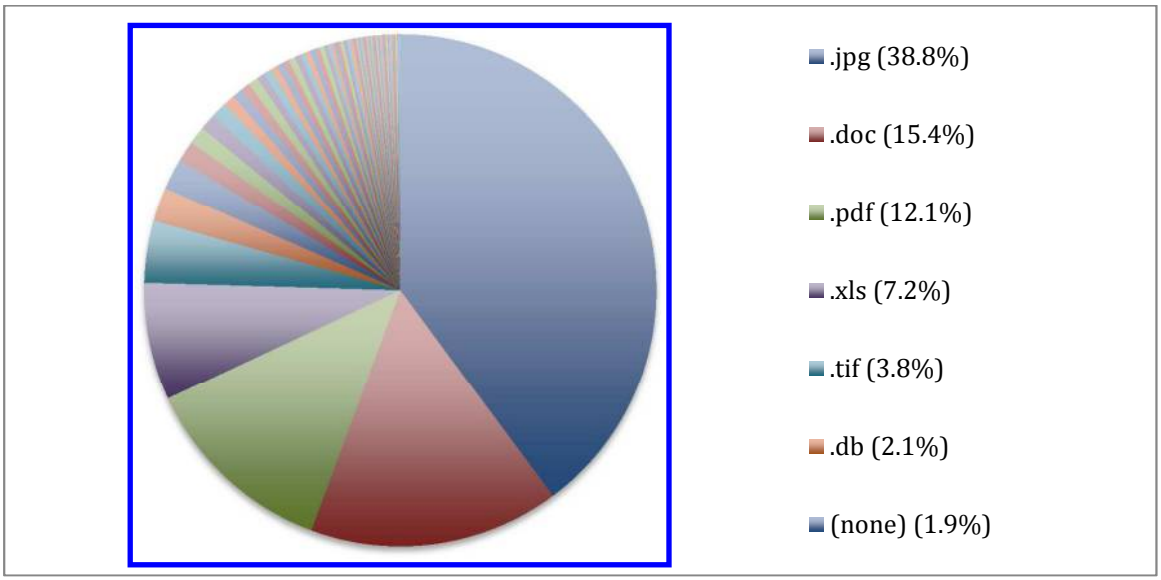

Figure 2. Distribution of file types, with file types representing less than 2% of all files included in the endnotes.[7]

Email was an important repository of significant correspondence across many units, and further work is needed to plan for its transfer and preservation. For example, there are nearly 400 MS Outlook PST files on network storage, which does not include active staff accounts. There are also 3,357 MS Outlook MSG files on network storage, 1,552 MS Outlook Express EML files, and 4,307 Novell GroupWise files. Using the work of Prom (2011), appraisal needs to be performed on these email messages and mailboxes and then the files migrated to MBOX and EML, text-based formats well suited for long term preservation.

Removable media, such as CDs, DVDs and floppy disks, exist in both the paper-based archives and within departmental files and work areas. Knowledge of what is contained on these media was generally vague in almost all cases since they are either rarely referenced or insufficiently described. Using the inventories and finding aids available, it was difficult to discern the hardware format of removable media, since earlier archivists would often simply refer to a “disk” or “disc” or “floppy.” Thus, the only way to know for certain would be to recall the boxes and look inside, a time consuming and costly activity since the archives are stored offsite.

Institution websites have grown to become substantial and noteworthy publications. However, web archives of these websites do not exist outside of the Internet Archive. Thus, creating web archives that capture these continually evolving publications is needed to document this increasingly important activity. Staff do not store digital records with permanent value using third party cloud-based services. For

example, Dropbox.com and Google Drive are used for working documents that eventually get transferred to network storage. However, there may be some value to preserving select NYAM social media and wiki sites because of the curation of existing assets and public engagement. Additionally, some social media sites may have unique content not available elsewhere (e.g., streamed video content).

Conclusion and Implications

This case study indicates that the greatest challenges to born-digital institutional archiving are not necessarily technological but rather social and cultural. Or rather, the challenge is getting individuals to transfer material to a digital archive so that it can undergo the technological transformations needed to ensure its long-term availability. However, transfer is impeded by a variety of factors, which include: 1) lack of motivation to transfer files because digital files do not take up valuable physical space, 2) lack of attention to or awareness of past digital work products that may be well suited for long-term preservation, and 3) although individuals are somewhat aware of challenges stemming from technological obsolescence, there is a lack of action to respond to this because the methods available for addressing this issue are not known nor is it viewed as a high priority by most staff.

Individual behaviors are shaped not only by organizational culture and practices but also by experiences with digital technology from various aspects of life. For example, many staff assume that curating a digital collection in an item-level digital asset management system means that the asset will always be available much in the same way that individuals assume that uploading a photo to Facebook means it "will always be there." In both cases, this view does not take into account the organizational and technological resources needed to ensure this continuation, as well as factors from the external environment that could impede such continuation.

A multi-staged approach is required to address the vast amounts of digital data that need to be appraised for permanent retention. The first stage is infrastructural: developing the capacities to accession, preserve and make accessible the born-digital documentation with historic and legal value, as well as designing detailed workflows for completing this work. The second stage is educational, or ensuring that staff are aware of the capacities for preserving and making accessible born-digital documentation as well as developing confidence in the process. This educational work should dispel views that digital preservation is automatic, not needed or completely impossible, all of which are pervasive views sometimes held paradoxically by the same person. And the last stage is to perform selection and accession work on the vast backlog of digital files that have permanent value, as well as the destruction of files that do not hold such value. This is "proactive appraisal," or rather appraising digital files for permanent retention in consultation with authoring department, and performing the transfer.

Thus, for archives that have already accessioned born-digital material such as floppy disks into their collections, the challenge is primarily technological: finding hardware and software to render the media intelligible while preserving the context such as through disk imaging. However, for institutional archives, this is only one relatively small facet of the entire born-digital archiving challenge. For institutional archives like those at NYAM, the most time and resource intensive work is the educational work needed to address the social and cultural attitudes as well as the proactive appraisal of vast digital backlogs. However, through dedicated effort, preserving born digital documentation with historic value is an obtainable goal.

References

AIMS Working Group. "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," available at: http://www.digitalcurationservices.org/aims/white-paper/ (accessed 18 March 2014).

Barrera-Gomez, J. & Erway, R. (2013), *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read in-house*, Dublin, OH: OCLC Research, available at: http://oclc.org/content/dam/research/publications/library/2013/2013-02.pdf (accessed 1 April 2014).

Consultative Committee for Space Data Systems (2012), Reference model for an Open Archival Information System (OAIS), available at: http://public.ccsds.org/publications/archive/650x0m2.pdf (accessed 1 April 2014).

Cocciolo, A. (2010), "Can Web 2.0 Enhance Community Participation in an Institutional Repository? The Case of PocketKnowledge at Teachers College, Columbia University," *Journal of Academic Librarianship,* Vol. 36 No. 4, pp. 304-312.

Cornell University Library (2007), *Digital Preservation Management: Implementing Short-term Strategies for Long-term problems,* available at: http://www.dpworkshop.org/ (accessed 1 April 2014).

Davis, P. M. and Connolly, M. L. J. (2007), "Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace," *D-Lib Magazine*, Vol. 13 No. 3/4, available at: http://www.dlib.org/dlib/march07/davis/03davis.html (accessed 25 September 2014).

Goldman, B. (2011), "Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in Manuscript Repositories," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage," RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage*, Vol. 12 No. 1, pp. 11-24.

Granger, S. (2000), Emulation as a Digital Preservation Strategy, *D-Lib Magazine,* Vol. 6 No. 10, available at: http://www.dlib.org/dlib/october00/granger/10granger.html (accessed 1 April 2014).

King, C.J., Harley, D., Earl-Novell, S., Arter, J., Lawrence, S., Perciali, I. (2006), *Scholarly Communication: Academic Values and Sustainable Models*, Center for Studies in Higher Education, University of California, Berkeley, available at: http://www.cshe.berkeley.edu/sites/default/files/shared/publications/docs/scholarlycomm_report.pdf (accessed 25 September 2014).

Lavoie, B. F. (2004), *Technology Watch Report: The Open Archival Information System Reference Model: Introductory Guide*. OCLC and Digital Preservation Coalition, available at: http://www.dpconline.org/docs/lavoie_OAIS.pdf (accessed 18 March 2014).

Lee, C. A., Kirschenbaum, M., Chassanoff, A., Olsen, P., Woods, K. (2012), "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions," *D-Lib Magazine*, Vol. 18 No. 5/6, available at: http://mirror.dlib.org/dlib/may12/lee/05lee.html (accessed 18 March 2014).

Lee, C. A. and Tibbo, H. (2011), "Where's the Archivists in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills," *Archivaria,* Vol. 72 Fall, pp. 123-168.

National Digital Stewardship Alliance (2013), *NDSA Levels of Preservation*, available at: http://www.digitalpreservation.gov/ndsa/activities/levels.html (accessed 1 April 2014).

Nikolić Đerić, T. (2011), "Importance of Secondary Documentation in Ethnographic Museums," *Ethnological Researches*, Vol. 1 No. 16, pp. 91-101.

O'Connell, K. (2012), "Survey Says: How a Museum-wide survey of electronic resources is forming the basis of ERM at the Brooklyn Museum," *Museum Archivist,* Vol. 22 No. 2, pp. 1, 5-8, available at: http://files.archivists.org/groups/museum/newsletter/current/pdfs/MAS_newsletter_Summer_2012.pdf (accessed 25 September 2014).

Prom, C. (2011), DPC Technology Watch Report: Preserving Email, Digital Preservation Coalition, available at: http://www.dpconline.org/component/docman/doc_download/739-dpctw11-01.pdf (accessed 1 April 2014).

RLG & OCLC (2002), *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA:

RLG, available at: http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf (accessed 1 April 2014).

Rosenthal, D.S.H. (2010), Keeping Bits Safe: How Hard Can It Be?, *Communications of the ACM*, Vol. 52 No. 11, pp. 47-55.

Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., Tansley, R., Walker, J. H. (2003), DSpace: An Open Source Dynamic Digital Repository, *D-Lib Magazine*, Vol. 9 No. 1, available at: http://www.dlib.org/dlib/january03/smith/01smith.html (accessed 1 April 2014).

OCLC & CRL (2007), *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, Chicago, IL: CRL and Dublin, OH: OCLC, available at: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (accessed 1 April 2014).

Rinehart, A. and Prud'homee, P. (2014), "Overwhelmed to action: digital preservation challenges at the under-resourced institution," *OCLC Systems & Services,* Vol. 30 No. 1, pp. 28-42.

Smith, M. (2008), Curating Architectural 3D CAD Models, *International Journal of Digital Curation*, Vol. 4 No. 1, pp. 98-106.

Society of American Archivists (2014), "Digital Archives Specialist", available at: http://www2.archivists.org/prof-education/das (accessed 1 April 2014).

Todd, M. (2009), *Technology Watch Report: File formats for preservation*, Digital Preservation Coalition, available at: www.dpconline.org/component/docman/doc_download/375-file-formats-for-preservation.pdf (accessed 1 April 2014).

Appendix – Handout for the Focus Groups

ELECTRONIC RECORDS MANAGEMENT PROJECT
MEETING WITH THE CURATORIAL DEPARTMENT
DECEMBER 2, 2013

Facilitator: {removed for review purposes}

1. Welcome, Introduction and General Purpose (15 minutes)

- Introduce records project, with some background on archiving.
- Introductions around the table (Your name and role at NYAM)

2. Discussion of your department's electronic records, with consideration of relevant Records Retention Schedule(s) (45 minutes)

Question to consider:

- Does your department produce any records that fall (or still fall) into these categories? Are there new types of documents that you are creating that are not captured here that may have legal and historical value?
- *Mystery Folders* - Do you have records on shared drives or personal drives that you think are at-risk (e.g., no one in your department knows what some folders are, and they could possibly have historical or legal value)?
- For each record group destined for the archives, where are you storing these digitally-created files: Media Beacon, TMS, Raiser's Edge, Great Plains, Departmental Shared Drive (I drive or S drive), All Staff Q Drive, My Documents (F Drive), External Hard Drive or USB Drive, Burned CD/DVD/Blu-ray, Workstation-installed software (e.g., Lightroom, iPhoto), Personal device (iPhone, Android), Hard drive on Workstation (Macintosh HD, Desktop, C Drive)

    - Are any of these original (or master) records stored exclusively using Cloud-based services (e.g., Dropbox, Google Drive, YouTube, Issu, Facebook, Sound Cloud, uStream, Pinterest, Instagram, Flickr, Brightcove)
- For each record group destined for the archives, what are the formats that you make use of (MS Office, PDF, Google Docs, JPGs, Email, Websites, Smartphone apps, Adobe Creative Suite products).
- Do you have possession of any physical digital media covered by the record groups discussed that should be transferred to the archives (e.g., floppy disks, CD/DVDs, tapes, Zip drives, etc.)? Have you inherited any such media from your predecessor or departed colleagues?
    - Please note that any physical media could be considered at risk of data loss (e.g., CDs, DVDs, floppy disks, Zip drives)
- How would you like the electronic records transfer to the archives process to work in the future?

Next Steps: Come back to your department to review the initial plan for archiving your department's electronic records.

Please feel free to reach out to me if you have any further feedback related to our discussion today. You can reach me at {email removed for review purposes}.

[1] http://www.digitalpreservation.gov/ndsa/activities/levels.html; http://preservica.com/; http://www.oclc.org/contentdm.en.html
[2] https://www.jam-software.com/treesize/
[3] http://www.gallerysystems.com/tms
[4] http://www.mediabeacon.com/
[5] http://usa.autodesk.com/adsk/servlet/pc/index?id=6703438&siteID=123112
[6] http://opendesign.com/files/guestdownloads/OpenDesign_Specification_for_.dwg_files.pdf
[7] (no extension) (1.9%), .jp2 (1.3%), .tiff (1.0%), .png (1.0%), .gif (0.7%), .indd (0.6%), .html (0.6%), .htm (0.6%), .psd (0.5%), .eps (0.5%), .vwx (0.5%) [VectorWorks], .mcd (0.4%) [VectorWorks], .bmp (0.4%), .txt (0.4%), .wpd (0.4%) [WordPerfect], .zip (0.3%), .ppt (0.3%), .tmp (0.3%), .dwg (0.3%) [AutoCAD], .docx (0.3%), .3dm (0.2%) [Rhino], .qxd (0.2%) [Quark XPress], .rtf (0.2%), .jpeg (0.2%), .csv (0.2%), .otf (0.2%) [Font], .000 (0.2%), .xml (0.2%), .mlm (0.2%) [Novel Groupwise Email], .mov (0.1%), .msg, (0.1%) [Outlook email message], .as (0.1%) [Text-based ActionScript], .pdz (0.1%) [ProntoDoc for MS Word plugin], .xlsx (0.1%), .frm (0.1%) [WordPerfect Form], .frd (0.1%) [additional research needed], .wav (0.1%), .car (0.1%) [additional research needed], mp3 (0.1%), .nef (0.1%) [Raw image format], .dng (0.1%) [Digital negative], .ai (0.1%), .dv (0.1%) [Digital Video], .dat (0.1%) [additional research needed], .lay (0.1%) [low-resolution EPS files used for print layouts], .afm (0.1%) [Adobe Font Metrics], .wk4 [Lotus 123] (0.1%), and dozens more file formats with less than and equal to 0.1% of total files.