



NEW YORK UNIVERSITY
LIBRARIES

National Digital Stewardship Residency | New York
Final Report
Julia Kim

Host: New York University Libraries

Access and Discovery of Born-Digital Archives

Access and Discovery of Born-Digital Collections

Despite the prevalence of born-digital collections, it is only in recent years that archives have been able to research and implement born-digital workflows that include researcher access.

My task was to research, develop, and implement access-driven foundational workflows for NYU's three archives. This encompasses not only the backlog of obsolete media materials such as floppy disks and zip disks; but also newly accessioned hybrid collections with content on hard drives and flash drives. All the steps in collection processing, with the necessary addition of pre-accessioning donor surveys and amendments to the original donor agreements, had to be reviewed and amended to compensate for born-digital materials. As the first individual to work full-time in the newly created Digital Forensics Lab, my fellowship mentor Donald Mennerich and I have worked closely to establish protocols. As part of the testing and implementation, I've trained archival staff and students to image and process collections. I lead a Born-Digital Access Working Group that is now tackling description, with the senior archivists from NYU's three archives participating. In the last portion of my fellowship, my primary task was to emulate and test of obsolete born-digital artwork. This is highly experimental and innovative. Like many of the processes I work on, there are no best practices; my best references are online gaming forums. This emulation, and several other accessible born-digital excerpts from collections, were made available to invited researchers to assess the documentation, manuals, and hybrid finding aids to improve our processes before these collections are open to all researchers.

There were a number of deliverables and outcomes noted in the original project plans, but I was most interested in piloting new ways for researchers to access complex born-digital collections. I was also interested in interviewing researchers to understand their values with born-digital emulation. Examples of marquee "test" collections I work on are the Jeremy Blake Papers and the Exit Art Collection. Jeremy Blake's "time-based paintings" relied on animated digital images; his work represents a unique intersection in technology and art making. The Blake Papers include over 340 pieces of external media. The Exit Art Collection is even more expansive. While the analog portions have been processed, I'm coordinating the preliminary arrangement of the institutional 2TB RAID that contains databases, exhibition documentation, email, and, for example, word processing files going back into the 1980s. Determining selection on such a large collections challenges traditional archival principles and MPLP approaches.

These collections represent the challenges besetting many institutions: complicated rights, unavoidable reliance on proprietary programs, privacy issues, obsolete interdependencies, and sheer quantity of files and data. While much has been said of the technical issues I am routinely challenged with, integrating the born-digital workflow at NYU is another, perhaps more difficult, administrative challenge to overcome.

Project Partners:

NYU's Digital Library and Technology Services Department and NYU's Fales Library and Special Collections and staff mentors Donald Mennerich (Digital Archivist) and Lisa Darms (Senior Archivist) supported the project with regularly (1-2 times a month) meetings. I also had approximately 6 meetings with senior archivists as part of the Born-Digital Access Working Group/Born-Digital Access Advisory Group. I work in a new Digital Forensics lab.

I was able to reach out to most of the peer institutions listed on the original proposal for meeting and discussion. Peer institutions included Yale University, Rockefeller Archives Center, Columbia University, Princeton University, New York Public Library, as well as other institutions. While on-site visits to peer institutions were limited, through my organizational and curatorial work for the one-day unconference, I was able to meet representative digital archivists and digital asset managers from many institutions.

While NDSR supported my ability to do much of my work, other institutions also stepped in to provide much needed supplementary funding and/or support. Thanks to these additional supporters, I shared my work at several additional conferences. My CurateGear presentation was supported thanks to my mentor Donald Mennerich's awarded funds. The Foundation of the American Institute for Conservation's George Stout grant generously funded my presentation at the American Institute for Conservation conference (May 2015). As an Emerging Archival Scholar, my participation in the Archival Research and Education Institute (July 2015) was also funded.



Figure 1. Removable Media in the Jeremy Blake Papers

Accomplishment of Project Goals:

Goal 1: To be trained in forensic disk imaging, forensic analysis techniques and archival processing methods for born-digital collections and to review, assess update existing documentation.

During the past 2 months, I've become fluent with multiple imaging techniques, hardware, software, and command-line tools to image removable media, internal hard disk drives, and experiment with the live-capture and exploration of write-blocked laptops. To a lesser degree, I've gained competency with forensics analysis techniques and archival processing methods that tackle the challenges inherent in born-digital, including complicated rights, scale issues, and technical obsolescence. I also created workflow documentation for different audiences including an imaging workflow document, a high-level workflow diagram, and several computer software and tool installation documents.

Goal 2: Gather requirements from staff and researchers to model potential workflows and access points for scholarly access.

While the original project documentation required my participation, I was tasked to chairing and co-establishing a new NYU Library-wide Born-Digital Access Working Group. This committee met every 1.5 months, providing an opportunity for discussion with representatives from NYU's three archives. The working group will dissolve with my departure, but the work of a born-digital advocacy group will continue with the formation of another committee and NYU's new department, Archival Collections Management (ACM).

In order to better understand and assess researcher access to born-digital and associated documentation and manuals, I also conducted five one-hour long researcher

interviews with a diverse cross-section of Fales researchers. These researcher interviews were not original to the project proposal. These sessions gave me the opportunity to enter in conversation with potential users of born-digital and hybrid collections. While the conversations were loosely structured and only a small sampling of archival researchers, it was nevertheless useful to hear what each researcher most valued in looking at portions of the Exit Art Collection and Jeremy Blake Papers, how they typically conducted research, and what they knew and assumed (sometimes incorrectly) about archives and archival research. Interviews were approximately one hour in length and the audio was recorded for further analysis. Researchers were encouraged to verbalize anything they saw and noted and the study, using the think-aloud-protocol. Results from these interviews will be shared at the upcoming Society for American Archivists Annual Meeting (August 2015).

Goals 3: Research and assess virtualization and emulation can be integrated in library infrastructure.

I was provided with a suite of computers, both contemporary and obsolete, to test potential emulation environments against native environments. As part of my research, I reviewed existing available research on emulation integration in workflows. In this phase, I gathered additional requirements (more obsolete software programs, for example) for selected emulation of key works from the model collections selected, as well from the Exit Art Collection and Jeremy Blake Papers. Additionally, I've been able to do some cursory research on networked emulation. While the technical issues are already complex, this is an area in which there are many security concerns as well. As noted above, I have been able to test emulation and other forms of born-digital access with researchers and library staff from

Goal 4: Participate in the prototyping of discovery interface.

In the first month's review of the fellowship goals, this prototyping goal was dropped. Any discovery interface would require work with the support of publishing departments. It was also agreed that the fellowship project included more than enough areas of research and this work was likely to be least sustainable and integrated by the library.

Other NDSR Goals

While the project goals were foremost in focus, the NDSR fellowship was also designed to create a cohort of resident and give them opportunities to network with key figures in the field. This was accomplished through our bi-weekly meetings, our participation and support of one other's professional events and enrichment sessions. The cohort also frequently supported one another through after-work events.

The NDSR-NY program also emphasized and supported individual professional development through funding and promoting workshops, some of which were mandatory and some of which were optional. METRO supported this by waiving all fees for METRO sponsored workshops.

NDSR instruction sessions:

- Assess, Describe, Store: Thinking like a Computer, Jacob Nadal (September 3, 2015)
- Digital Preservation Methods, Policies, and Practices, Sibyl Schaeffer (September 4, 2015)
- Project Management, Mitchell Brodsky (September 8, 2015)
- Trusted Digital Repository (TDR) workshop, with Kara Van Malssen and Seth Anderson (October 24, 2014)

NDSR cohort-led enrichment sessions:

- Archive-IT User Group (January 27 2015), Karl Blumenthal
- MoMA enrichment session (December 4, 2015), Peggy Griesinger
- American Museum of Natural History (April 17, 2015), Vicky Steeves

Examples of optional workshops I attended:

- Preservation Metadata with Rebecca Guenther (April 8-9, METRO)
- Introduction to Ruby on Rails and Project Blacklight workshop with Benjamin Armintor and Brian Hoffman (January 2015, METRO)
- Assessment in Focus: Designing and Implementing an Effective User Feedback Survey with Nisa Bakkalbasi (November 2014, METRO).
- Omeka Bootcamp (October 16, 2014, Columbia University Digital Humanities Center)
- PBCore: A How-To and Why-To Webinar (October 23, 2014)
- ePADD workshop with Peter Chan, Personal Digital Archiving Conference Workshop ((April 26, 2015)
- Digital Preservation for Public Broadcasting webinar with Rebecca Fraimow, WGBH (May 19, 2015)

Project Execution

"This (born-digital) is not covered in depth in (conservation) programs." (independent study graduate student)

I. Immersion in Digital Archives Technologies (Sept - Oct)

Activities:

Become fluent in digital forensics practices as they relate to archives through experimentation, hands-on work, research, documentation, and training of other staff.

1. Image fixed & removable media.
2. Ingest digital content and metadata into institutional repository.
3. Assist in arrangement and description.
4. Review, assess and update documentation regarding acquisition and preservation workflows.
5. Work with and assist graduate students learning imaging techniques in the Digital Forensics Lab. In addition to a once-a-week graduate student volunteer and a graduate student completing an independent study, graduate adjunct and staff also shared the Digital Forensics Lab space to learn to assess and process born-digital collections.

Deliverables:

1. Processing workflow document (and diagrams).
2. Emulation Report (directed towards repository heads).
Write and research a report on emulation as a sustainable practice for senior repository managers. Write a plan implementing and testing emulators.
3. Research and write an appendix to existing donor agreements taking into account born-digital materials and their concerns.
4. Co-write pre-acquisition questionnaire for donors to assess collections for born-digital material.
5. Explanatory optical media annotated bibliography for library media and preservation staff.

II. Analysis of Scholarly Research and Access Systems (Nov - Feb)

Activities:

1. Continue to process collections.
In addition to the Jeremy Blake Papers and semi-regularly monthly meetings with the graduate students and archival staff working on the Exit Art Collection. Image other born-digital collections.
2. Image metadata recommendations and research.
Research metadata standards for images and make recommendations.
3. Research, install, and configure emulators and dependent programs.
The installation of compatible emulators and dependent programs supporting the Jeremy Blake Papers was very experimental and took the majority of almost a full month due to disk imaging issues, host computer (Windows, Mac, and Linux) compatibility issues, and SheepShaver emulator variations.
4. Configure researcher access laptop.
Install emulators and dependent software and programs.

5. Install collection material on emulators and researcher laptop.

Deliverables

1. Born-Digital Access Working-Group Charge, co-write with members.
2. Refined documentation of workflows developed from Phase 1.
3. BitCurator installation documentation.
4. Reading list for hybrid and born-digital finding aids.

Acquisition ID	Date Acquired	Partner	Collection ID	Collection Title	Accession ID	Collection Contents (media types and quantities)	Dispossession
2014.1	2/2/2014	Fales	MSS 402	I'm From Driftwood records	2014.402	1 Hard drive	imaged, need to check before giving back to donor
2014.2	2/5/2014	Fales	MSS 279	New York Women's Culinary Alliance records	2014.279	1 hard drive	imaged
2014.3	5/19/2014	Tamiment	TW 635	Brecht Forum records	2014	3 mac mini internal drives, 1 mac Timemachine	imaged
2014.4	8/6/2014	Fales	MSS 402	I'm From Driftwood records	2014.402	Files copied from hard drive	files transferred
2014.5	9/9/2014	University Archives	RG 6-5-1	Office of the Vice President		Files copied from computer	imaged
2014.6	10/1/2014	Fales	MSS 408	Heritage Radio Network records	2014.408	1 usb flash drive	imaged, ingested
2014.8	10/26/2014	Fales	MSS 402	I'm From Driftwood records	2014.402	1 Hard drive	
2014.9	11/4/2014	Fales	MSS 415	Fortnight Journal records		2 hard drives and 1 RAID	
2014.1	11/20/2014	Fales	MSS 419	Third Streaming Gallery records	2014.419	Files copied from server (500 GB)	files transferred, but should be re-transferred
	12/22/2014	Fales	MSS 402	I'm From Driftwood records	2014.402	Files copied from hard drive	files transferred 12/22

Figure 2: Accessioning log for internal tracking

III. Prototype and Testing (March - May)

If this type of artwork (born-digital) had been taught and available for research, I might have changed my course of study (interviewed researcher).

This was the best 1-day technical conference I have ever gone to (Jane Gorgevsky, Digital Assets Archivist, Columbia University).

Activities:

1. Coordinate and manage one-day “unconference” event, CURATEcamp.

The born-digital workflows CURATEcamp was conceived in early November by co-organizer Peter Chan (Digital Archivist, Stanford University). CURATECamp differed from the usual curate format. The morning session will consist of scheduled 15 minute presentations on born-digital workflows. Presenters spoke on CD-DA audio migration and error reporting, description, and metadata challenges with complex born-digital material, among other topics. After a “birds-of-a-feather” lunch break, the afternoon broke into smaller discussion groups and workshop sessions (see Appendix for event program). This unconference was affiliated with the international Personal Digital Archiving Conference hosted by NYU. While this unconference fulfilled the requirement for an enrichment session, the work required to organize the event took the majority of my time in the weeks leading up to the event.

- ~~2. Design and prototype a discovery system for born-digital archives.~~
3. Lead the design and execution on several researcher interviews to test the manuals and to born-digital access.

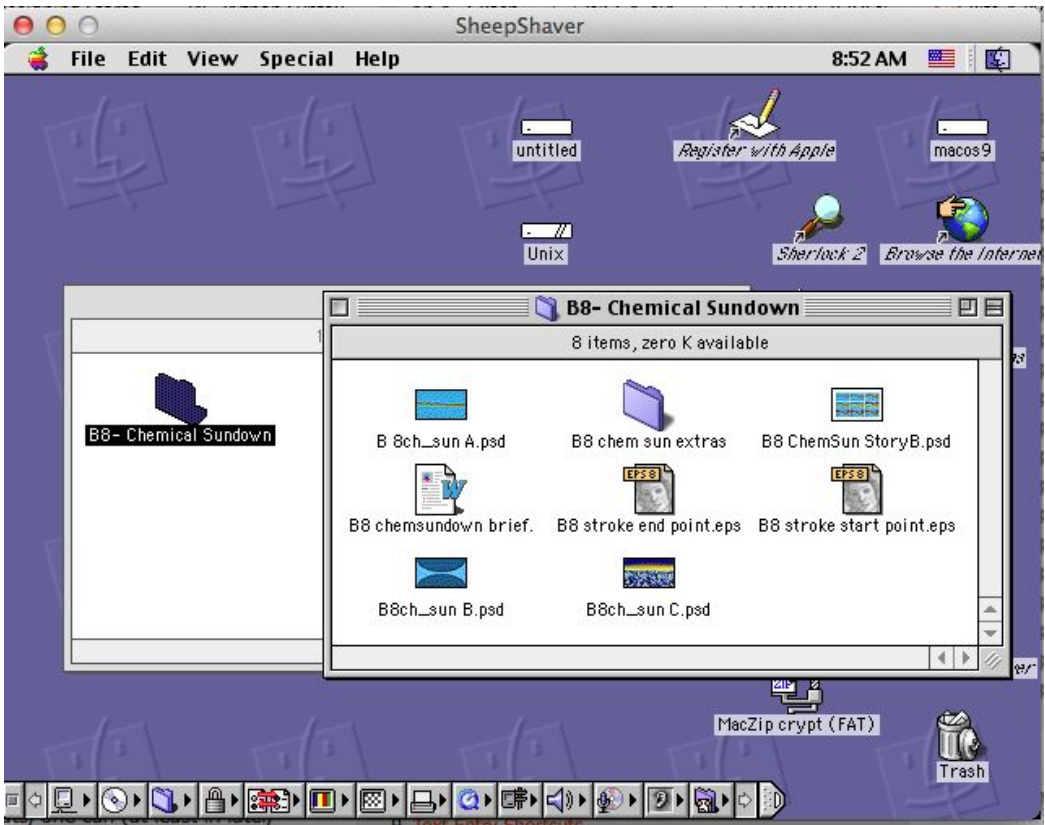


Figure 3. Emulated Environment



Figure 4: Classic Mac display of Blake's work

Deliverables:

1. Finalized system designs, wireframes and mockups of discovery interface for born-digital archival collection.
2. NDSR Final Report summarizing research and work during the fellowship.
3. CURATEcamp wiki.
4. CURATEcamp born-digital workflows institutional survey on practices (see Appendix)
5. Researcher agreement and interview protocols.
6. IS&T Conference paper co-written with Donald Mennerich, "Preserving Jeremy Blake's Digital Art and Archives," Archiving Conference, Volume 2015, Number 1, May 2015, pp. 49-50(2).
7. AMIA emulation panel proposal (summer project).
8. SAA presentation, outline (summer project).
9. AIC paper for publication in the Electronic Media Review (summer project).

Outreach and Dissemination

I'm contacting you because I recently read about your work on Jeremy Blake's Papers (on your blog). I am currently writing a research grant...for a two month residency in New York with a project about the archive of Jeremy Blake's Papers. For a few years already I have wanted to view the digital archive left by Blake. When I found out that you were working on the archive, I was thrilled to see this opportunity becoming closer. (international researcher and artist, March 2015 email correspondence)

While the project goals, activities, and deliverables were significant, the NDSR fellowship also encouraged public outreach through supporting conference presentations and use of social media. This portion of the fellowship has far exceeded expectations; I've been very fortunate to be able to promote the projects I'm working on at a number of very diverse national-level conferences. One unanticipated result from this is highlighted in the above excerpt from an email correspondence: researchers interested in working with the collections learned about the availability of the collection through my outreach efforts rather than a finding aid. Through learning about collections this way, they also were able to learn about some of the hurdles with archival born-digital processing.

I've also been fortunate that my mentor, Donald Mennerich, has shown me encouragement and support through our co-presentations and publications on our work together. In addition to the conference presentations, Don, Peter Chan, and I also co-organized a born-digital workflows CURATEcamp. The CURATEcamp put me in touch with leading professionals around the country. In addition to CURATEcamp is a list of some of the public events I was able to participate and/or present work specific to this project at.

Public Events:

- Association of Moving Image Archivists (AMIA), October 8-11, 2014, Savannah, GA. Chaired and organized a digital forensics with personal archives panel, with speakers from Emory University (Digital Archivist Elizabeth Roke) and Stanford University (Digital Archivist Peter Chan).
- CurateGear, January 7, 2015, Chapel Hill, NC.
Project presentation with mentor Donald Mennerich.
- BitCurator Access meeting, January 2015, Chapel Hill, NC
- METRO Annual Conference, January 15, 2015, Baruch College, New York City, NY
NDSR project presentation.
- Tour and question and answer for NYU MIAP "Complex Media" course, February 2015
- Art Libraries Society of America meeting (ARLIS), March 30th 2015, MoMA, NY
NDSR project presentation.
- CURATECamp, April 23, 2015, Brooklyn Historical Society, NY.

co-organizer, curator, and co-sponsor of the Personal Digital Archiving Conference.

- NYU Library Preservation week, library staff presentation, April 29, 2015.
- American Institute for Conservation (AIC), May 13-16, 2015, Miami, FL presentation and publication.
- Archival Education and Research Institute (AERI), July 13-17, College Park, MD.
- Society of American Archivists (SAA), August 16-22, 2015, Cleveland, OH presentation

In addition to the listed events, NDSR also supported multiple online publications to the Library of Congress Blog, "The Signal."

- "Creating Workflows for for Born-Digital Collections: An NDSR Project Update."
<http://blogs.loc.gov/digitalpreservation/2015/03/creating-workflows-for-born-digital-collections-an-ndsr-project-update/>

Publications and Pending Publications

- Imaging Science & Technology (IS&T) conference proceeding, "Preserving Jeremy Blake's Digital Art and Archives," Archiving Conference, Volume 2015, Number 1, May 2015, pp. 49-50(2), co-written with Donald Mennerich.
- Electronic Media Group journal, (AIC) (pending Fall publication), co-written with Donald Mennerich.
- "Digital Preservation for Libraries, Archives, and Museums, Edward M. Corrado and Heather Lea Moulaison, book review, Metropolitan Archivist (v21:1).
<http://www.scribd.com/doc/254647679/Metropolitan-Archivist-Volume-21-No-1-Winter-2015#scribd>

Conclusions:

In terms of my own professional development and growth, the NDSR fellowship has been an immensely enriching experience. While ostensibly I was to be treated as an entry-level employee at my host institution, the grand scope of the project, its collections, and their possible impact on the field gave me an invaluable professional experience that would far exceed most entry-level employees. I was given the opportunity to quickly learn from leaders in the field while also conducting my own work and research. Additionally, I was highly encouraged to promote this research.

While I had very little prior conference presentation experience previously, the NDSR fellowship also quickly habituated me to learning to effectively present to different audience levels and types. I did not anticipate this and my initial discomfort in this area has diminished greatly. This area of growth cannot be overstressed.

In a very short period, the fellowship fostered highly compressed activity in new areas of research. Most of the project aims were achieved within the 9 month timespan. My mentors allowed me to also interpret the project proposal and shape it to my research interests in emulation and access. Instead of making emulation recommendations, I was able to change the focus of the last third of my fellowship to configure emulators and offer researchers guided access to portions of collections as part of my researcher interviews and studies. This adaptation better fit my specific interests in researcher access. It should be noted that this was also easier to manage as it did not rely on departmental consensus making or the modification of any pre-existing workflows.

Program Recommendations

- Extend the timeframe to 12 months for similar projects.
In addition to just having more time to finish the project, I would have found it helpful to have more time for intensive technical training
- Scale projects goals, taking into account fellow conference travel and preparation, research, and workshops.
- Ensure institutional support.
Have institutions match funds to increase the investment in the fellow? Provide business cards, local travel, and set number of vacation and sick day, for example.

Project Highlights

Emulated access to Blake and access to portions of the Exit Art Collection.

There are very few research institutions that have instituted emulated access to born-digital materials. NYU joins a select but quickly growing group of institutions that include Emory University, New York Public Library, Cornell University, and Yale University. Additionally, researchers have contacted NYU for access to these collections due to the NDSR outreach and social media activities.

Emulation of Jeremy Blake, however, is a medium-term solution. The emulation created for the Blake collection should be further experimented with to better stabilize the environment and create further support for Blake's (and other collections) works. Further dedicated support, funding, and knowledge sharing in emulation is necessary for it to be sustained and further implemented at NYU. We are relying on older technology for the emulation and there are unresolved intellectual property issues that must be addressed to reach a large-scale and sustainable solution for more than a single marquee collection. As of now and for the foreseeable future, there is no one at NYU whose job it is to work in this area specifically.

The work that I've done will be taken up by both my supervisor, Donald Mennerich, and the support staff hired in the newly created Archival Collections Management (ACM) department. Staff from ACM will accession and arrange and describe collections and continue to refine any existing workflows. The emulation and access work will be sustained by the Reference Archivists and Public Services department, but maintenance and further necessary research will not be sustained.

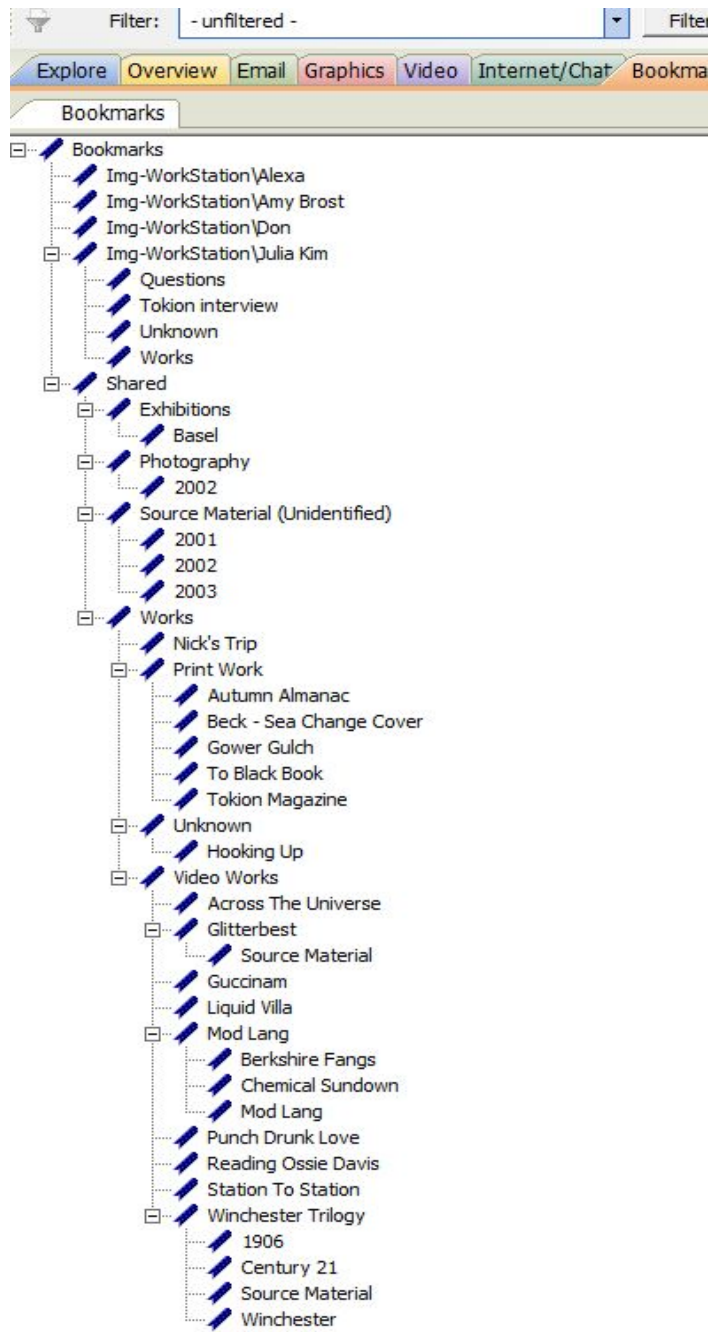


Figure 5: Jeremy Blake Papers - potential arrangement

Sample Products (Appendix)

- CURATEcamp born-digital workflows wiki
- CURATEcamp born-digital workflows institutional survey
- “Preserving Jeremy Blake’s Digital Art and Archives,” Archiving Conference, Volume 2015, Number 1, May 2015, pp. 49-50(2).

Appendix

Born-Digital Workflows CURATEcamp, April 23 at Brooklyn Historical Society

http://wiki.curatecamp.org/index.php/Born-Digital_Workflows_CURATEcamp,_April_23rd_at_Brooklyn_Historical_Society

A one-day CURATEcamp will precede Personal Digital Archiving Conference 2015, which will be held at New York University in New York City. This iteration of CURATEcamp will focus on born-digital workflows. You can read more about the topic below. The facilitators include Julia Kim, National Digital Stewardship Resident at New York University Libraries, Donald Mennerich, Digital Archivist at New York University, and Peter Chan, Digital Archivist at Stanford University.

Logistics

- WHEN: April 23, 2015, 9:00 am–5:00 pm (+ post-event drinks)
- WHERE: [Brooklyn Historical Society](#), 128 Pierrepont Street, Brooklyn, NY
- COST: Free, there is no cost to register for the event. However, registration is limited to 50 people and [advanced registration is required](#). Please note that this is a highly participatory event!
- LOGISTICS: Brooklyn Historical Society is a short distance from both the Court Street N and R subway stop and the Borough Hall 2 and 3 subway stop.
- DISCUSSION: [#curatecamp on Twitter](#) and #curatecamp on irc.freenode.net. Notes and documentation for each session will be made available.

Morning Talks

- **Automating Disk Imaging Processes - Euan Cochrane** (Yale University, Digital Preservation Manager)

The Yale library preservation department has been working its way through imaging its large collection of digital material found in its general collections in an attempt ensure the preservation of these materials for future generations. The volume and diversity of the floppy disks and optical media found in the general collections, along with a lack of skilled staff to process them has necessitated an investigation of ways in which the disk imaging process can be streamlined, and where feasible, automated. This talk will focus on the workflows developed to meet these challenges and, time permitting, may delve into issues of developing workflows to automatically emulate the outputs of the imaging process. [slides available here](#)

- **Arrangement & Description for Born Digital - Hillel Arnold and Bonnie Gordon** (Rockefeller Archive Center, Lead Digital Archivist and Assistant Digital Archivist)

At the RAC, processing archivists arrange and describe born-digital materials in FTK; this description is transformed to EAD and presented in our online finding aids. This talk will provide an overview of the workflows and systems in place to get description out of FTK and accessible to researchers. [slides available here](#)

- **Maximizing Description to Enhance Access to Born-Digital Archival Collections - Jarrett Drake and Rossy Mendez** (Princeton University, Digital Archivist and Public Services Project Archivist)

In this presentation, a public services archivist and a technical services archivist from Princeton University's Mudd Manuscript Library will consider how the description of born-digital archival collections impacts the access to these materials. After they explain the results of a recent review of their finding aids that describe born-digital records, they will advocate that user needs should inform the description of born-digital collections as well as demonstrate how their workflow for generating description is evolving to leverage existing metadata bound in born-digital records, thereby enabling processing archivists to create richer, more precise descriptive data. [slides available here](#)

- **ePADD - Peter Chan** (Stanford University Libraries, Digital Archivist)

ePADD - Special Collections Department at Stanford University Libraries received a grant from the National Historical Publications and Records Commission to develop a software program to allow repositories and individuals to interact with email archives before and after they have been transferred to a repository. It consists of four modules, each based on a different functional activity: Processing (arrangement and description), Appraisal (collection development), Discovery (online via the web), and Delivery (access). [slides available here](#)

- **Mass Migration: Building a Bulk Hard Drive-to-LTO Workflow From Scratch - Rebecca Fraimow** (WGBH Media, Library and Archives, NDSR Resident)

How do you successfully transfer 300 TB of material from hard disk to LTO over the course of a nine-month project? This presentation will detail the process of constructing a workflow for digital migration of large amounts of audiovisual data—and then adapting it, and adapting it again, to deal with the various roadblocks hit along the way. [slides available here](#)

- **Open Source QC Tools for Compact Disc Digital Audio (CD-DA) - John Passmore** (WNYC, Archives Manager)

John will briefly survey available free and/or open source tools archivists can use to assess the quality of their CD-DAs. The talk will include a demonstration of the QPXTTool (<http://qpxtool.sourceforge.net/>) command line interface and some instruction on how to interrupt the quality of your audio CDs based on BLER/E22/E32/FBE/Jitter measurements. [slides available here](#)

- **New Media Art: Preservation, Technical and Descriptive Metadata - Jason Kovari** (Cornell University, Head of Metadata Services and Web Archivist)

Overview of the metadata environment in PAFDAO (Preservation and Access Framework for Digital Art Objects), an NEH-funded project at Cornell University Library's Rose Goldsen Collection to preserve interactive born-digital New Media Art. [slides available here](#)

- **The Challenges Access Demands for an Established Accessioning Workflow - Gabriela Redwine** (Beinecke Rare Book and Manuscript Library, Yale University, Digital Archivist)

Since 2011, the Beinecke has followed an established accessioning workflow for creating disk images of physical media, extracting metadata, and moving the Bagged images and metadata into storage. Often, this process is initiated by a researcher's request for access to the digital media in a collection. I will share real-life examples to demonstrate some of the pitfalls of disrupting established workflows and the potential implications for digital preservation.

Demos & Workshops

- **Peter Chan** - Stanford University Libraries, Digital Archivist
 - AccessData FTK@Stanford

Special Collections Department at Stanford University Libraries started to explore the use of AccessData FTK for archival processing of born-digital materials since 2010. I would like to share how Stanford use AccessData FTK for archival accessioning, processing, and delivery with colleagues in the archival world.

- **Ben Fino-Radin** - Museum of Modern Art, Digital Repository Manager
Ben will demo the software MoMA built w/ Artefactual to create its Digital Repository Museum Collections – Management Application (DRMC-MA).

- **Cal Lee and Kam Woods** - BitCurator Access
 - I Imaged the Disk, Now What? Enabling and Mediating Access to Born-Digital Data

The BitCurator Access project, funded by the Andrew W. Mellon Foundation, is developing open-source software that supports the provision of access to data and metadata from disk images. We'll demonstrate software developed to date and discuss future development plans, including tools to redact files, file system metadata, and targeted bitstreams within disks or directories.

Potential Session Topics

Please list potential topics you're interested in learning more about [here](#), before the unconference.

CURATEcamp schedule

The morning session will consist of brief, scheduled presentations on born-digital workflows. After a “birds-of-a-feather” lunch break, the afternoon will break out [[CURATE-style](#)] into multiple sections to discuss the morning’s talks as well as any ideas for conversation that participants have. Alongside these streams, we will also have pre-planned workshop demonstrations.

We will provide a catered breakfast and coffee.

Click on the links below for session descriptions, notes, references, and documentation.

Time	Classroom	Great Hall (A)	Great Hall (B)	Great Hall (C)
9:00–9:30	Registration, Breakfast, Surveys			
9:30–11:45 /noon	Scheduled presentations			
12:00–1:00	"Birds-of-a-Feather" Lunch			
1:00–2:00	BitCurator Access Demo	acquisition and file transfer in the field	emulation	Unconference C1
2:10–3:10	Archivematica/MoM A Demo	FTK user group	hybrid collections	web archiving

3:20–4:20

[AccessData FTK @ Stanford](#) [born-digital workflows](#) [cloud acquisition](#) [Unconference C3](#)

4:30–5:00

Wrap-Up

Imaging Science and Technology, Archiving 2015 May 19-22, 2015

Preserving Jeremy Blake's Digital Archives

Donald Mennerich, Julia Kim; NYU Libraries; New York, New York/United States

Abstract

Jeremy Blake (1971-2007) was an American digital artist most well known for his animated video installations, "time-based paintings" and large-scale digital C-prints. Blake's evocative work combined 8mm film, vector graphics, and hand-painted imagery to create a distinctive aesthetic: color-drenched, atmospheric and even hallucinatory. His acclaimed work was a defining example of new media art.

The artist worked primarily with Adobe's Photoshop software, saving his work in its native PSD format. Working from an archival perspective in collaboration with technologists, curators, and individuals familiar with Blake's artistic process, this paper highlights the challenges, as well as the opportunities, for preserving and creating access to complex born-digital formats like PSD.

Overview

Born in 1971, Jeremy Blake came to prominence in the late 1990s through the exhibition of his large format digital C-prints. These prints combined elements of both photography and painting, but were generated through digital means. During the course of the subsequent decade, Blake worked on numerous high profile projects: he created the cover art for Beck's album *Sea Change* (2002), produced the animation sequences for Paul Thomas Anderson's film *Punch Drunk Love* (2002), participated in three consecutive Whitney Biennials (2000, 2002, 2004) and exhibited his work internationally in over a dozen major museums. Shortly after Jeremy's Blake's tragic passing in 2007, New York University's Fales Library and Special Collections acquired the pioneering digital artist's archives. Blake's archive consist of a variety of physical formats: over three hundred optical discs, three external hard drives, six Zip disks, ten Digital Linear Tapes and thousands of additional files transferred from a fourth hard drive and Blake's laptop.

By far, the predominant file format contained in Blake's archive is Adobe's native format, the PSD file, for its photo editing software Photoshop. Blake worked extensively in Photoshop during all years covered in his archives, circa 1999 to 2007, during which he generated upwards of three thousand unique Photoshop files generated on several different versions of the software. These files encompass the various projects Blake was involved in, e.g., video installations, gallery prints, commercial work, etc.

Blake is best known for his "time-based paintings," which were the result of his complex working methodology that involved the dense layering of myriad source material: hand-rendered images, photographic elements, 8mm film and various scanned objects. Blake would collaborate with animators and sound artists, providing them with Photoshop files containing his source material that they would further modify to generate the finalized moving image. The discrete layers in Blake's Photoshop files could be used as independent elements in a production and were often animated to

evolve from one to another. This resulted in Blake's unique, vivid and often hallucinatory aesthetic. For example, a single file from his work on *Punch Drunk Love* was found to contain all elements used to generate the animation seen in the movie's opening sequence. Additionally, Blake would produce "briefs" of works in separate Photoshop files, which documented key transitions points with overlaid textual notes. These files, and supplementary Microsoft Word files communicated the overall desired look and feel of a work to his collaborators.

Blake's PSD files presented numerous preservation problems that test the boundaries of current digital preservation methodologies. Adobe's Photoshop format is proprietary, under documented and not well understood, all of which contribute to it being undesirable for digital preservation. To address these problems, NYU Libraries developed a multi-pronged test of potential strategies, which included maintaining original hardware, emulating the original environment, and migration of file formats to support long-term preservation of Blake's archive. Initial work was done to create forensic disk images of all media carriers. This was done to stabilize, refresh and provide a baseline, byte-level, preservation service for all of Blake's works. Files from these disk images would be used in all subsequent preservation work.

To test potential file migration workflows, a sample of Blake's Photoshop files were migrated to the current version of the PSD file format using an up-to-date version of Photoshop (Creative Cloud 2014.) Through the course of our testing we found that virtually all of Blake's Photoshop files, regardless of the version of Photoshop they were created in, could be opened, viewed, manipulated and saved in the current PSD format using the current version of Photoshop. Post migration, a number of analysis tools: FFmpeg, FFProbe, ImageMagick and ExifTool, were all tested to determine to what degree they supported the PSD format and what functionality could be expected of them. With ImageMagick's *identify* command we were able to count and compare technical characteristics for each layer present in both the original and migrated Photoshop files. Using this procedure, we were able to determine, with some certitude, that all encapsulated layers within the file persisted through the migration process.

Secondly, we also tested migration from PSD to multi-layered TIFF files, a relatively recent feature supported by Photoshop. Our research has indicated that neither the core TIFF specification [2] nor its official extensions include support for layered images. No TIFF-viewing packages tested were capable of displaying the multi-layered files generated using Photoshop. As such, conversion to TIFF format through Photoshop will likely only ensure rendering in Photoshop itself. At present, we are unable to find standardized practices to encode layered images in TIFF files [3].

Ensuring the authenticity of the files and his art-making process has been a key concern. Given that the majority of Blake's work was created in now-legacy environments, i.e. pre OS X versions of the Mac operating system and Photoshop, emulation was explored to provide more authentic access to the files in Blake's archive. SheepShaver [4], an open-source PowerPC emulator, was installed on numerous Mac, Windows and Linux platforms to test if it was capable of running the correct versions of the Mac OS (7.0-9.0) and Photoshop (versions 3.5 through Creative Suite 2.) While we

were successful in demonstrating that now obsolete versions of Photoshop could be run and Blake's files could be accessed using SheepShaver, only Ubuntu 10.04 worked consistently. Other versions of Windows, Macintosh and Linux operating systems were able to run various versions of Photoshop, but Blake's Photoshop files would often be rendered with major glitches, visual artifacts and distortions that made the images unrecognizable. Even simple, non-layered PSD files would render as crude blocks of color in many emulated environments. However, we also discovered that the same PSD file across all platforms could open correctly in older versions of Photoshop (3.5), latter versions (such as 6.0) would not, indicating compatibility issues across platforms as well as software versions.

Ultimately, while it was promising to see Blake's files run on the correct version of the software in which they were created, the emulator itself proved to be highly unstable; SheepShaver frequently froze and crashed. The platform that we found to be the most stable, Ubuntu 10.04, itself was released in in 2010 and has already reached its end-of-life date from its parent company[5]. We expect further complications with host-emulator interaction, as the difference between hardware and software dependencies needed to run the emulator will only increase. Additionally, SheepShaver was initially released in 1998 and was last in active development more than 5 years ago. While there is an active user community and new builds are released regularly, development of the core codebase has essentially ceased. Documentation on the dependencies and performance of SheepShaver, especially with complex and proprietary software like Photoshop, is similarly lacking.

Based on ongoing research, NYU Libraries has created and is currently executing preservation workflows for layered PSD files based on migration to the current version of Photoshop for all legacy materials in addition to retaining the originals. We advocate for comprehensive study of the effects of migration of legacy PSD to up-to-date versions of the format using Photoshop, the effect of migrating from PSD to multi-layered TIFF and for further study and support for emulation as a long-term preservation and access solution.

References

- [1] "Adobe Photoshop File Formats Specification" Accessed December 7, 2014. <http://www.adobe.com/devnet-apps/photoshop/fileformats.html/>
- [2] "TIFF, Revision 6.0" Accessed December 7, 2014. <https://partners.adobe.com/public/developer/en/tiff>
- [3] "TIFF, Tag Image File Format, FAQ." TIFF File Format FAQ. Accessed December 8, 2014. <http://www.awaresystems.be/imaging/tiff/faq.html>.
- [4] "The Official SheepShaver Homepage" Accessed April 17, 2015. <http://SheepShaver.cebix.net/>
- [5] "Releases - Ubuntu Wiki" Accessed April 17, 2015. <https://wiki.ubuntu.com/Releases>

Author Biographies

Donald Mennerich is Digital Archivist at NYU Libraries, where he works primarily with the application of digital forensics tools to manage born-digital archival collections.

Mennerich currently serves on the ArchivesSpace Technical Advisory Council, the BitCurator Access Advisory Board, and is a mentor in the 2014 National Digital Stewardship Residency program. Mennerich holds an MS in Information Systems from Pace University and an MLS from Simmons College.

Julia Kim is a National Digital Stewardship Fellowship Resident at NYU Libraries, where she collaborates on the creation of access-driven workflows in the context of complex born-digital media. Julia graduated from New York University's Moving Image Archiving and Preservation Program with a thesis on born-digital forensics applications and adaptations in archives. Julia has worked with audiovisual materials at the Dance Heritage Coalition, Anthology Film Archives, and the Fundacion Patrimonio Filmico in Bogota, Colombia.