Managing locally-produced born-digital A/V Content

Howard Besser, NYU

Managing locally-produced born-digital A/V Content-

• General problems of born-digital content
• What type of Content are we discussing today?
• Why is this different than other problems we face?
• 3 Examples

General problems of born-digital personal content-

In the analog world

• Traditionally, we have come to understand the work of writers and scientists by scholars studying their papers in Special Collections and Archives
• Their correspondence and progressively different drafts of papers reveal their changing thoughts and craft
• But how do we gather these in the Digital Age?

Alasdair Gray's Lanark (Glasgow U Library)

Correspondence
Correspondence

Where can we find these today?
- Do people write letters on paper? Can we see the iterations of changes on manuscripts?
- Where can we find today’s equivalent of these?

This will require
- new interventions (like changing authors’ workflow, or intervening in email handling software)
- New tools (like for analyzing email)
- new approaches like digital archeology, forensics

Stages of the problem
- Stage #1: People write on computers instead of paper
- Stage #2: People no longer store their digital works in places over which they have absolute control
  - Email services (gmail, yahoo)
  - Cloud storage for documents (google docs)
  - Social network services (Flickr, YouTube)

Stage #1 Issues-
- Obsolescence Problem
- Making Sense of the Donor’s desktop
- new approaches like
  - Digital archeology
  - Digital forensics

Old Digital Formats

Obsolescence-the Viewing Problem
- Digital Info requires a whole infrastructure to view it
- Each piece of that infrastructure is changing at an incredibly rapid rate
- How can we ever hope to deal with all the permutations and combinations
Hardware--Issues & steps

- Reading the storage device
  - Finding a device that will play it (8", 5.25", 3.5"; double/single density, zip, jazz, ...)
  - Finding an interface plug to hook it to a current computer (SCSI; Firewire 1,2; USB 1,2,3)
  - Making sure that a recent operating system can recognize the device (driver)

Software—Issues & initial steps

- Create Disk Image of the storage media (so you don’t lose “hidden files”, edit histories, profiles and settings, etc.)
- Begin to explore the Directory Structure of the disk image-
Appraisal & Preservation/Access

- Emulation/Migration just to read the files
- Digital forensics (BitCurator) to decide whether or not to save edit histories, possibly contribute to a Finding Aid, etc.
- Evaluate privacy issues to decide about redaction (more later)

Digital Forensics for Scholarly analysis

Stage #2 Issues

- New tools (like for analyzing email)
- What happens when a service shuts down?
- What happens when the account owner dies?
Analyzing Email

Issue of 3rd Parties & Personal records

- Emails, documents, attachments, photos with someone else’s personal information (social security number, tax returns, libelous accusations)
- Sensitive works involving this individual and others (meetings with Battered Women or victims of oppression in foreign countries, sex photos/videos)

Protecting Privacy of Donors & 3rd Parties

- The only way to try to protect this material is:
  - De-accessioning (but sometimes this involves important records where only small parts are sensitive)
  - Embargoing until the sensitive becomes far less sensitive (ie. Named people die)
  - Redacting the sensitive parts
    - Within the archival records
    - Upon disclosing the records
  - Coding names to protect privacy, human rights, prevent defamation, …

Other Underlying Sensitive Issues—Examples (1/2)

- Human Rights
  - Genocide testimony (can result in revenge on family members)—Rwanda
  - Documentation of oppression of women (Witness in Congo, RAWA in Afghanistan)
- Laws of different countries
  - France, Germany—Nazi advocacy or revisionism
  - Turkey—Armenian records about post-WWI era using word “genocide”
Other Underlying Sensitive Issues--Examples

- Pornography
- Right of Publicity
  - "use of a person’s name, likeness, or other personal attribute"
  - Each state has a different law on this
- Traditional Knowledge (new WIPO protections proposed)
  - tradition-based literary, artistic or scientific works; performances; inventions; scientific discoveries; designs; marks, names and symbols; undisclosed information; and all other tradition-based innovations and creations resulting from intellectual activity in the industrial, scientific, literary or artistic fields
- Knowledge systems, creations, innovations and cultural expressions which have generally been transmitted from generation to generation; are generally regarded as pertaining to a particular people or its territory; and, are constantly evolving in response to a changing environment

Consider this

- The more widely-available you make a record, the more likely it is that someone will find offense in it, and the more likely that someone will challenge you in a foreign jurisdiction (with diff laws)

Privacy is part of our Ethical Standards

- “Archivists protect the privacy rights of donors and individuals or groups who are the subject of records. They respect all users’ right to privacy by maintaining the confidentiality of their research and protecting any personal information collected about them in accordance with the institution’s security procedures.”

  -http://www2.archivists.org/standards/code-of-ethics-for-archivists

What I know from my prior work

- InterPARES—If we hope to preserve electronic records, archivists need to be involved early in the life-cycle of that record, long before the record enters the archive
- Preserving Digital Public Television—Pushing metadata gathering upstream into the production cycle

WorldFocus

- Nightly news program begun Oct 2008
- We began working with Workflows six months before program began
- Had ability to engineer metadata gathering into the creation/production process

What type of Content are we discussing today?

- For an academic library:
  - Departmental records (recruiting works, student work, …)
  - Campus: television/radio records, newspaper photos
  - Recordings of campus events/activities
  - *Records from the Campus PR department
- For a public or academic library:
  - Personal "papers", photographs from groups of individuals
  - Records of a local volunteer organization
  - *Material from local television or radio channels
  - *Local newspaper morgue
  - Audio or video recordings of local events (sports, cultural, …)
Famous visitor to UCLA

Why is this different than other problems we face?
• Within a collection, very little consistency in terms of quality, file format/compression, file naming conventions, folder arrangement, attached metadata
• A bigger challenge than what we face with other types of collections:
  – Digital records from a business usually follow some kind of consistent enforced guidelines
  – Collection organized by a single individual usually has internal consistency (at least for given periods of time)
  – Much material of this kind must be harvested from social networks

What challenges do these pose?
• We need to find smart ways to harvest metadata and analyze files, as well as to influence behavior of potential contributors

3 Examples-
• Occupy Movement
• Northwestern University
• UCLA

Occupy Movement
• Ideas from Activist Archivists-

Activist Archivist Website
Some Activist Archivists’ Projects & Ideas-

• Try to alter practices of user/contributors
  — “Why Archive” postcard & video
  — 7 Tips to Ensure Your Video Is Usable in the Long Term
  — Study of metadata loss through uploading to services
  — Best Practices for Creators/Collectors
• Redundancy in recording metadata
• Crowd-sourcing of selection/appraisal

Why Archive Postcard

7 Tips to Ensure Your Video Is Usable in the Long Term

• Collect details while filming
• Keep your original raw footage, unaltered
• Make your video discoverable
• Contextualize it
• Make it verifiable
• Allow others to collect and archive
• Or archive it yourself

“How Archive” video

“How Archive” postcard

ACCOUNTABILITY. Archives collect evidence that can hold those in power accountable.

SELF-DETERMINATION. We define our own movement. We need to create and maintain our own historical record.

SHARE. Archives are a point of entry to our movement’s rich record. We can use them to ensure transparency, generate discussion, and enable direct action.

EDUCATE. Today’s videos, flyers, web-pages, and signs are material for tomorrow’s skill-shares, classes, and mobilizations.

CONTINUITY. Just as past movements inspire us, new activists will learn from the experiences we document.

RECORD & COLLECT what’s happening around you.

PRESERVE the record.
Study of metadata loss through uploading to services

Best Practices for Content Collectors

- Security
  - Sensitive material
  - Scraping for content
- Content Search
  - Internet Archive, Archive-It, YouTube, Vimeo, Bit-Torrent
- Receiving Content
- Metadata Extraction
- Copyright

Best Practices for Content Creators

- Security
  - Hidden camera laws, parties’ consent laws
- Capturing Content
  - Highest quality, set date and time-stamps, note location
- Offloading Content
  - Raw files directly onto computer, keep material organized
- Uploading Content
  - Importance of tagging, review of diff services
- Depositing with an Archive
- Copyright

Creative Commons Guidance

- Creative Commons lets you mix-and-match four different conditions:
  - Attributing: You let others copy, re-use and distribute your video, but they must credit you.
  - Share-Alike: You let others copy, re-use and distribute your video, only if they do the same with the work they create.
  - Non-Commercial: You let others copy, re-use and distribute your video for non-commercial purposes only.
  - No Derivative Works: You let others copy and distribute your video, but not to create new works using it.
- You can use these conditions in different combinations to share your work in a controlled way. Creative Commons licenses are legal tools that depend on pre-existing copyright laws. Having a Creative Commons license on your work may give you legal recourse, but it may not actually prevent people from downloading and re-using your video illegally.

Marking Creative Commons licenses

- There are a few ways to mark your video with a Creative Commons license. One way is to include a Creative Commons “bumper” or text card in your video. Creative Commons has created some with graphics that you can download from their website. This method is useful if your video is going to be shared offline (e.g. on DVD, live screenings), as the license information is attached to the video itself.
- Another way to mark your video with a Creative Commons license is to publish your video on platforms that are Creative Commons-enabled, such as YouTube, Vimeo, or Internet Archive. These platforms allow you to easily select a license during the upload process. This method is useful because the license is machine-readable. A search engine, for example, can detect the license.

Redundancy in recording metadata

- Think Tank material collected by NYU’s Tamiment Library-
Collecting – Think Tank

- Daily, 2 hours
- Audio capture hardware provided by NYU library (Zoom-H2n)
- Bi-weekly digital file transfers

Think Tank metadata redundancies

- Guidelines stipulate that person holding recording device will check to see that time and date stamp are correct before beginning recording (mostly didn’t happen)
- Guidelines stipulate that a script be read verbatim at the beginning of the recording, with date, time, proposed subject, etc. (and would eventually allow voice-recognition software to create appropriate metadata). Script also stated that all participants agreed to Creative Commons licensing of the recording
- Guidelines requested that date/time be embedded in the applied file-name

Think Tank Guidelines

- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video

Collecting – Think Tank

- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video

Collecting – Think Tank

- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video
- Guidelines for Populating Audio and Video
Crowd-sourcing of selection/appraisal

- Suggestions to Tamiment Library re collecting YouTube videos of Occupy

Tamiment YouTube collecting

- plug-in for FireFox (downloadhelper.net)
- As of April 2012, ca. 250 items, policy: large events
- Fair Use: 2012 ARL Code of Best Practices
  - “transformative” collecting with context
- Tamiment was selectively browsing through YouTube Occupy videos, trying to choose which ones to keep, then cataloging them with
  - Title, Creator, Creation Date, Upload Date, Description, URL, Youtube Username, License, Format, Codec, Source Media, On Internet Archive, CC License type

Tamiment YouTube Cataloging

But this won’t scale!

March 24, 2012 YouTube stats
(just 6 months after start of movement)

- “#Occupy” 169,000
- “Occupy Wall Street” 98,400
- “Occupy Protest” 70,500
- “Occupy Movement” 54,800
- “#OWS” 50,300
- “Occupy Oakland” 13,400
- “Zucotti Park” 6,690

Alternative approach to YouTube Selection process

- Develop categories of important YouTube videos
  - Celebrity visits, Internal workings (library, kitchen, media), Confrontations with police, Labor, Housing, etc.
- Have Occupiers fill in an online form listing the 5 most important videos in each category
Advantages of YouTube Collaborative Filtering Selection Process

- Scalable and manageable
- Consistent with Occupy ideas of inclusiveness and of managing own story
- Tamiment can still choose to be selective in collecting only a portion of what is voted in, but the total set for review is a manageable scale
Managing locally-produced born-digital A/V Content

- [http://activist-archivists.org/](http://activist-archivists.org/)
- [http://visions.indstate.edu/pda2014/](http://visions.indstate.edu/pda2014/) (past PDAs)

Howard Besser, NYU
[http://besser.tsoa.nyu.edu/howard/Talks](http://besser.tsoa.nyu.edu/howard/Talks)

Managing locally-produced born-digital A/V Content

- **Howard Besser**, Director, Moving Image Archiving and Preservation Program
- **Sady Sullivan**, Director of Oral History, Brooklyn Historical Society
- **Brendan Allen**, Archivist, Democracy Now!
- **Natalie Milbrodt**, Director, Queens Memory Project
- **Yvonne Ng**, Archivist, WITNESS