Archiving Media from the "Occupy" Movement: Methods for Archives trying to manage large amounts of user-generated audiovisual media

Howard Besser

The "Occupy" movement that emerged in the US in the Fall of 2011 grew rapidly by posting both text and audiovisual works on online sites that would likely disappear. This paper details the efforts of archivists to preserve Occupy's audiovisual material, and serves as a lesson for any archive trying to preserve and manage large amounts of user-generated media.

The Occupy Movement

The Occupy movement began in September 2011 in New York City, and quickly spread to cities and towns across the US (and eventually to other parts of the world). [1] The movement's key slogan – "We are the 99%" – reflects that the movement was fueled by a moral outrage at the control exerted on society by a small minority of the populace. The movement's name – "Occupy" – points to its tactic of "occupying" public physical spaces for 24 hours per day 7 days per week both to highlight the importance of those spaces to society's discourses, and to maintain a constant presence where people who pass by cannot help but notice the movement. This 24/7 presence in physical space also led to the development of self-organizing and community-building within the movement itself, and is reflected in the communal feeding of large numbers of participants (numbering in the hundreds, or in the case of NYC sometimes numbering in the thousands), and in collective providing of services for all participants (in the form of lending libraries, electrical power, wireless internet services, etc.).

In addition to physically occupying key public spaces, the movement engaged in extensive large-scale demonstrations involving thousands or tens of thousands of participants. Often these demonstrations highlighted what the movement saw as particular examples of systemic problems in society – the government's bail-out of financial firms (while not rescuing the worst-off individuals), the seizure of peoples' houses via foreclosure, etc. A major characteristic of the movement was the broad creativity shown in signs carried in protest marches, and in creative street-theater, where protestors would dress as bankers or governmental officials and act out satiric scenarios.

The Occupy Movement and the Internet

Like the Arab Spring movement that preceded it (and inspired it), the growth of the movement was fueled by communication mediated on the Internet. But, partially because of the high level of broadband Internet access and the ubiquity of smart-phones in the US, the number of digital photographs and video and audio recordings that movement participants posted online was astounding. Statistics from the photographic posting service Flickr show that 6 months after the first Occupy demonstration, more

[1] According to Wikipedia, by October 9, 2011 (3 weeks after its beginning in NYC), Occupy protests had taken place in over 95 cities across 82 countries, and over 600 communites in the US. (http://en.m.wikipedia.org/wiki/Occupy_movement accessed August 19, 2012)

than half a million individual photos had been posted to this service with the tag of "#Occupy".[2] Tens of thousands of individual videos were posted to YouTube during the first few months of the movement. By 6 months after the first Occupy action, 169,000 individual postings to YouTube had been tagged with "#Occupy".[3] The vast amount of content created and the dissemination through commercial websites posed interesting problems for libraries and archives interested in preserving this material.

User-contributed information: How Occupy material resembles what libraries and archives will face in the future

Cultural heritage repositories such as libraries and archives have always collected material from organizations and individuals. But most of this material has entered the repository as part of a large collection that follows some kind of internal organizational and retrieval scheme (such as a newspaper's photo collection with it's own specific numbering and metadata scheme, or an individual's file folders with labeling reflecting that individual's own personal logic). Libraries generally put this type of material in Special Collections departments to help indicate that each collection reflects the organizational schema of the donating individual or organization. And archives create a unique Finding Aid for each collection because each collection's organizational scheme is so different that the archive could not create an organizational scheme that worked across multiple collections.

In the late 20th century as we moved more towards a world of born-digital collections, many challenges for cultural repositories (such as how to collect email correspondence, or how to organize the folders on someone's hard disk) emerged. But for the most part, a repository could expect that all the material coming from an organization or individual would follow a single organizational logic, and would be encoded in a particular set of file formats (such as all documents in a particular version of Microsoft Word for each time period, and all digital movies in a particular version of Quicktime for each time period).

But many of the past artifacts collected by libraries and archives (such as letters and email and other types of correspondence) have been replaced in today's world by more atomized digital artifacts (such as blogs and social network postings) that are more connected to works originated by others. And in a world characterized by networked information and collaborative authorship, the cultural institution collecting material of historical and social importance will need to collect material coming from a variety of different sources, each with its own conception of metadata and file format standards.

The material generated by the Occupy movement looks very much like the type of material that will be entering the archives and library special collections of the future. It is a vast quantity of user-generated everyday material, created by a multitude of different

[2] March 24, 2012 Flickr statistics show 632,089 items tagged with "#Occupy", 164,304 tagged with "Occupy Wall Street", 179,454 tagged with "Occupy Protest", 113,904 tagged with "#OWS", 40,572 tagged with "Occupy Movement", 27,202 tagged with "Occupy Oakland", and 9,164 tagged with "Zucotti Park" (location of the first NYC occupation).

[3] March 24, 2012 YouTube statistics show 169,000 items tagged with "#Occupy", 98,400 tagged with "Occupy Wall Street", 70,500 tagged with "Occupy Protest", 50,300 tagged with "#OWS", 54,800 tagged with "Occupy Movement", 13,400 tagged with "Occupy Oakland", and 6,690 tagged with "Zucotti Park" (location of the first NYC occupation).

users (Besser 2011). There is no easy way to control for quality, file format, or metadata. Unlike most organizational collections that try to enforce standards for metadata and file formats, there are not even guidelines suggesting what schemes should be followed. And because the content comes from so many individuals, it lacks even the semi-consistency that a single individual would apply to the items that he or she creates. And what might logically constitute a future "Occupy" media collection is actually found today spread over a multitude of commercial social networks (such as Flickr, YouTube, and Facebook) that each add their own organizational idiosyncrasies, and offer no guarantee that the material will remain posted for any length of time.

So, in order to preserve this type of material, we need to find smart ways to harvest metadata and analyze files, as well as to influence the behavior of potential contributors. A number of the methods that might be useful for future user-contributed collections were explored in the projects of the Activist Archivists, which are outlined later in this paper. Many of these methods are based upon the findings in prior projects on preserving born-digital material that Activist Archivists had participated in. From the InterPARES II project (http://www.interpares.org/ip2/ip2_index.cfm) (2002-2006) we learned that 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. From the Preserving Digital Public Television project (http://www.thirteen.org/ptvdigitalarchive/) (2004-2010**)** we learned the effectiveness of automating metadata collection from the moment of first recording (Besser & van Malssen 2007).

Activist Archivists

In response to the Occupy movement, in October 2011 students and recent graduates of NYU's Moving Image Archive and Preservation Program – MIAP (see other paper for this conference explaining MIAP in more detail) began efforts to explore the archiving and preservation of the media being generated by the Occupy movement. They felt that much of the spirit, decentralization, self-organization, playfulness, and whimsy of this protest movement would be lost to history if the media that documented this did not survive. Joined by MIAP Director Howard Besser, the group took on the name Activist Archivists, and began work on about a dozen different projects to archive the born-digital media content related to this movement (http://www.activist-archivists.org/), with most of the projects having potential impact on the archiving and preservation of all types of material that might be collected by cultural repositories in the future.

Many of the sub-projects involved collaboration with various partners. These included both collecting institutions (such as the NYU Library's Tamiment Collection) and "working groups" from the Occupy Wall Street movement (including both the "Archives" working group, which mainly dealt with collecting non-digital artifacts such as posters and signs, and the "Media" working group).

It is important to note that certain predispositions of the Occupy movement may not be relevant to libraries and archives building collections from other sources. Those in the Occupy movement were very suspicious of conventional organizations, including universities and libraries, and often needed convincing that a conventional cultural institution might be a good repository for the artifacts that they created. Occupiers could also be characterized as having a "do-it-yourself" (DIY) mentality, not wanting to rely on professionals outside their community to organize and provide access to the material. This was part of a critique of conventional media dissemination outlets which appeared

to not do a good job of explaining the movement, and appeared to manipulate news coverage. The Occupiers wanted to control their own story. Their ideology also made them suspicious of any type of exclusive arrangement, including giving their material only to a particular repository. And their consensus decision-making process made it difficult for a repository to try to come to an agreement with the group, as a group discussion on a topic such as this might range over several meetings, and each meeting might be composed of a slightly different group of participants (and discussion from previous meetings had to be repeated to and accepted by the newcomers).

Activist Archivists Projects

Because many in the Occupy movement did not immediately recognize the value of saving artifacts representing their movement, one of the first projects undertaken by the Activist Archivists was to create a short simple postcard that succinctly explained why saving these artifacts should be important to the movement. Much effort was made to explain this using the value system of the movement, and significant effort went in to avoiding what many in the movement referred to as "outsider language". The "**Why Archive" postcard** briefly explained that saving this material would serve values dear to the movement, including: Accountability, Self-Determination, Sharing, Education, and Providing Continuity. It explained that it was important to "Record and Collection" what was happening in the movement in order to "Preserve the record".

Another important Activist Archivist project sought to provide advice to those recording Occupy events. The "**7 Tips to Ensure our Video is Usable in the Long Term**" offers important advice to those recording video, still images or sound. Advice includes: collecting details about the recording, keeping the original raw unaltered footage, making the recording discoverable/accessible, contextualizing it, making it verifiable, allowing others to collect and archive the recording, or being very careful about archiving it yourself. (For this last point, the Tips suggests looking at the Library of Congress' website on Personal Archving.) Ideas from these "7 Tips" were expanded into a much lengthier set of advice in "**Best Practices for Video Activists**". This document also discussed legal restrictions involving getting permission from those you record (which in the US differs depending on which state the recording takes place in), as well as issues involving copyright (and stresses the idea of executing a Creative Commons license which will allow a repository to archive the material and make it available in the future.

An Activist Archivist research project revealed significant issues for any organization seeking to preserve digital videos posted to commercial websites. MIAP student Rufus de Rham's "**Metadata Loss During Upload or Download**" empirical study uploaded digital video files from 3 different consumer recording devices (iPhone, Android, and Canon t2i) to 4 different internet dissemination services (YouTube, the Internet Archive, and both Vimeo's paid and free services). He then examined over 200 fields of metadata present in the original files, and found that YouTube and Vimeo's free service stripped out almost all the metadata, and only the Internet Archive's service maintained all the metadata intact. Important metadata such as date, time, and GPS location were stripped out of the file header by YouTube and Vimeo's free service. This means that cultural organizations seeking to collect videos posted to online services would have to reconstruct date, time, and location information for anything they collected from YouTube or Vimeo's free service.

Activist Archivists also worked on a "**Best Practices for Content Collectors**". This document aimed to pull together information gleaned from other Activist Archivist projects, and to offer advice to cultural repositories on how best to collect this type of content. It suggested that with content pulled from the Internet Archive, metadata for date and time recorded (and possibly geographic location as well) could be automatically extracted from metadata in the file headers. It discussed some Occupiers' sensitivity around traditional organizations and exclusivity, and suggested ways in which and archive could collect materials without being tied to an exclusive agreement. It discussed the sensitivity of some protesters around security, and offered suggestions for using software tools that could automatically detect faces and place small black bands over peoples' eyes in order to make individual identification difficult.

Two Activist Archivist members worked closely with NYU's Tamiment archival collection, which had arranged for the **digital audio recording of daily 2-hour "Think Tank" discussions of strategies and tactics** by Occupy Wall Street participants. The Activist Archivists suggested ways in which important metadata could be automatically extracted from the recordings, as well as working on the spreadsheets of metadata related to each recording. Guidelines for the recordings insisted that key metadata be captured in redundant ways. So, for example, the date that should have been captured by the Zoom H2n digital audio recording device was also embedded within the file-name, and was supposed to be orally recited as part of an initial script that was read into the recorder at the start of each session. This redundancy was important because the recording device's date and time were not necessarily reset when the device's battery ran out of energy, and the oral script was not always read at the beginning of the recording. This form of redundancy would also protect against a future time when the metadata might become separated from the content. And the reading of the same exact script at the beginning of each session would allow for future, more sophisticated speech recognition software to automatically extract important metadata. The script also is designed to systematically record information including who is responsible for the recording, what is expected to be discussed in this session, and that participants desire to execute a Creative Commons license that will allow archives to preserve and disseminate the recording.

Activist Archivists also advised NYU's Tamiment archival collection on **preserving digital videos posted on YouTube**. Selection of which videos to capture and preserve had been a tedious process, with the Tamiment Director viewing each YouTube video, and deciding whether or not it was worthy of preservation. This process would not scale to the 100,00 videos that had been posted to YouTube tagged with "Occupy Wall Street" just 6 months into the movement. Activist Archivists suggested that categories for the videos be developed (celebrity visits; internal workings such as the library, kitchen, and media; confrontations with police; labor union involvement; housing/foreclosure protests; etc.), and that selection of the most important videos be crowd-sourced to members of the Occupy Wall Street Working Groups (both Archives and Media). Each participant could fill out a web form indicating what they thought were the 5 most important YouTube videos to preserve in each category. This method would scale up as the number of videos posted to YouTube increased, and it was more in line with the participatory nature of the Occupy movement. This method has still not been implemented due to the illness and eventual death of the Tamiment Director.

Conclusion

Activist Archivists tackled a wide variety of problems inherent in selecting, capturing, and preserving media related to the Occupy movement. They explored various methods for convincing the individuals who record these events to make sure that their recording devices were set to automatically record date, time and location; to use consistent metadata and file-formats; to post to sites that will not throw away their metadata; and to execute creative commons licenses that will give cultural repositories the legal right to preserve and make available these recordings. They also advised collecting institutions on how to articulate their need to preserve to the individuals that make recordings, and provided advice on both redundant methods of capturing metadata for recordings, and on scaling the selection process in creative ways. These efforts are likely to prove useful for solving a problem that most collections will face in the near future – how to organize, preserve, and provide access to that large amount of user-generated content that most collections will receive in the future (and not have the time to catalog or convert).

Citations

Besser, Howard (2011). *Amateur Collections and Scholarship: Lessons from the impact of amateur collections on Cinema Studies research and on Film Archives' practice.* Photo Archives and the Photographic Memory of Art History, Part III, March 25-26, Institute of Fine Arts, New York University, NY, NY, USA.

Besser, Howard and Kara van Malssen (2007*). Pushing Metadata Capture Upstream into the Content Production Process: Preliminary Studies of Public Television,* DigCCurr 2007: an international symposium in digital curation, April 18-20, Chapel Hill, NC, USA (http://www.ils.unc.edu/digccurr2007/program.html).