New Review of Academic Librarianship, 19:291–299, 2013
Published with license by Taylor & Francis
ISSN: 1361-4533 print / 1740-7834 online
DOI: 10.1080/13614533.2013.824905

Routledge
Taylor & Francis Group

Beyond the Repository: A Mixed Method Approach to Providing Access to Collections Online

BRIAN WADE GARRISON
University of Kansas, KU Libraries, Lawrence, Kansas, USA

After providing access to over 100 video interviews conducted by a professor with notable entertainers and personalities from film through an institutional repository, an experiment was conducted to discover whether a larger audience could be gained by adding a subset of 32 of these videos to YouTube. The results, over 400,000 views, indicate that libraries should cultivate a mixed method approach to providing access to digital content.

KEYWORDS institutional repository, open access publishing, academic staff, digitization, university libraries

INTRODUCTION

Academic libraries and institutions of memory struggle with the vast range of decisions available when providing access to unique collections online. In contrast with the library catalog and finding aid that have translated trusted discovery methods to online contexts, there is no clear normative way forward for providing access to the complete range of digital content that libraries may now make available. Librarians dealing with unique collections face a range of attractive options that differ from institution to institution depending on collection strengths and resources. However, given the full range of content that a library needs to make available, the question of which method of access is best for any particular collection will not always overlap with the available tools.

This article demonstrates the advantages of a mixed method approach to providing access to unique collections through a digital repository, DSpace,

Address correspondence to Brian Wade Garrison, University of Kansas, KU Libraries, 1425 Jayhawk Blvd, Watson Library, Lawrence, KS, 66045, USA. E-mail: wadeg@ku.edu

and a streaming media service, YouTube.[1] Choosing a mixed method approach, using both a local digital repository and YouTube, mitigates the disadvantages of a single specialized system for discovery and access. Though this article focuses on providing access to archival collections, these same tools can be used for many different purposes throughout university libraries. For a more extensive study of work currently being done with YouTube in university libraries, a recent survey details how various segments of the library have begun experimenting and utilizing the tool (Cho 2013).

The following sections will outline the history of the collection and the process used to add content to DSpace, describe the strategy behind adding content to YouTube, and detail several of the considerations or advantages of this approach.

Collection Background and DSpace Access

In 2010 Associate Professor John Tibbetts of the Department of Film and Media Studies at the University of Kansas approached the Center for Digital Scholarship unit in the Libraries about providing access to his research. In 2005, KU Libraries launched its digital repository, KU ScholarWorks, using the DSpace software developed by MIT Libraries. Aware of KU ScholarWorks as a repository for research, Tibbetts contacted KU Libraries with the intention of having his work placed there. For more than 30 years Tibbetts has been interviewing notable artists, musicians, entertainers, and performers about their work and career. Concerned about access to his materials and the audience that could be gained by making his work available, Tibbetts moved forward in working with the libraries to get them deposited online.

Tibbetts provided a list detailing approximately 1000 audio interviews and 283 video interviews from the early 1980s to the mid-1990s. Most of the interviews Tibbetts initially provided consisted of video interviews done for Kansas City television on press junkets during that period. On average the interviews lasted from 6 to 7 minutes. Since the videos would often have been edited for television, having the full interviews made the collection unique. Many of the 6 to 7 minute segments include interactions with the interviewees that would have never been edited into the final cut for television. Additionally, Tibbetts often collected photos, made drawings or paintings depicting the interviewee, and documented other information surrounding the interview.

The general approach by KU Libraries was to treat Tibbetts' work much like it would a research paper that included additional supplementary materials or data and add the content to the DSpace repository in its raw entirety. While DSpace treats all file types equally, all easily ingestible into the system, it is not ideal for accessing audio, video, and other multimedia. For these types of multimedia files, DSpace requires that the user download the files in

order to open and view them. There are ways of getting around this obstacle, such as creating a streaming media server and providing a link to the content through DSpace. However this is less than ideal, as it neither stores the file in the repository nor provides ideal viewing and access conditions. Plugins to the DSpace installation can allow for the display of thumbnails indicating the relevance of the content for the end-user. However, our particular installation and resources precluded such an installation.

Regardless of these limitations, making the collection available through KU ScholarWorks still met the objectives of making the collection openly available and storing it for future use. Although Tibbetts has not donated the masters and chosen a permanent home for the collection, he provided digital copies that had already been transferred to DVD. From these files, library staff extracted and inserted a Creative Commons License into them. We agreed on using the Attribution-NonCommercial-NoDerivs 3.0 Unported license, restricting use to citing and sharing free and unedited copies, to ensure that as the videos were potentially propagated they would remain in their current form unless someone received the explicit permission of Tibbetts. Although this is the most closed CC license available, protecting the interests of a faculty member, already willing to take the unusual step of providing free access to his life's work, was in the best interest of KU Libraries as it ventured into this relatively new territory. Using the structure produced by DSpace, videos interviews conducted for the same film were grouped together under a single item record. For the title field, the title of the film was used. For the author field, Tibbetts and the interviewees were included. By mapping the metadata in this manner browsing the collection was possible without necessitating changes to the standard collection interface within DSpace.

As of May 2013, the 134 videos added to the online collection have received 2,984 downloads and 4,831 record views. Before starting the project, neither John nor KU Libraries set benchmarks or goals estimating the size of the possible audience that could be reached. The goals of both parties had been simply to provide access, not reach the widest possible audience. The preliminary model for all further work on the project had been completed, and the project could continue adding the collection of Tibbetts' materials to KU ScholarWorks.

Uploading Videos to YouTube with a Minimalist Approach

As the usage statistics in DSpace indicated relatively less usage when compared with text collections in the repository, finding an alternative or wider audience for the collection that was not being reached seemed like a viable goal. Uploading a small subset of the videos to YouTube would provide a contrast to DSpace in not directly comparable (e.g., views versus downloads) but measurable ways. In the process, KU Libraries would learn how providing access to digital video might be better performed in the future.

When placing the videos on YouTube, the service presents many different options for framing content. Due to the number of variables and difficulty of interpreting how different choices might affect the reception of the materials, the most reasonable solution was to turn off as many features as possible. As a result, the opportunity provided would contrast the relatively organized and curated modes of access on DSpace with the much more typically uncurated quality that most ephemera on YouTube are typically given. Placing the videos online would establish their merit independently of extensive curation. There would be no heavy branding of the institution responsible for making the videos available, other than a quiet reference and handle (i.e. stable unique URL) to the original video in the description on YouTube. Automatically generated tags suggested by YouTube were ignored. User-generated comments were deactivated. The only organizational thread between the videos would be that they were all in the same collection within YouTube, had a Creative Commons license, and contained Tibbetts interviewing someone. In effect the videos would be out there on the internet, on their own, in the wild, without a librarian or archivist to help guide the end-user.

Utility 1: Using Analytics to Track Statistics

Over a period of one month in 2011 a subset of 32 videos were added to YouTube. They were selected in the same logical groupings based on movie title that had been used to organize the videos within DSpace. The videos were also selected based upon the popularity of the films that they represented. From February 2011 through April 2013, the videos have received 440,999 views on YouTube with a monthly average of 8,850 in 2011 and 21,692 in 2012. Even these monthly averages exceed the total videos downloaded or total number of records viewed for the entire Tibbetts Collection in KU ScholarWorks since the project started in August of 2010.

In reporting the large number of views of the content and asserting that other libraries might try a similar experiment, it is important to acknowledge that as a whole the videos possessed an appeal outside the norm of the other open content that KU Libraries could provide access to through YouTube. Many universities that have similar profiles to KU probably have pre-1923 film (i.e., content in the public domain) that has value for smaller audiences, whether archival footage of historical value or of particular interest to their university community, depicting sports or campus life. However, even though the numbers may suggest that this collection is an extreme outlier, they should only encourage others to try experimenting with their collections.

One distinct advantage of offering video through YouTube instead of other means is the robust statistics that YouTube offers the curator. They enable the curator to draw conclusions about the actual use of the collection

and the origin of traffic. Although KU Libraries' installation of DSpace includes a plugin that tracks usage statistics (i.e. numbers of downloads, numbers of views, and which countries are downloading files and records), it does not provide enough feedback to help the curator experiment with methods of curation. YouTube continues to add features to the analytics to help track usage, such as the recently added "estimated minutes watched" enabling curators to further assess the overall impact of the collection, when videos are being played as opposed to when users press play but then click away.

Utility 2: Judging the Effectiveness of Social Media

YouTube and the social web have grown together and correspond with birth of videos that go viral and receive a million or more views. Therefore, it would be reasonable to assume before analyzing the data that a significant portion of the collection would receive notice through these channels. However after analyzing the data and the different patterns of those getting to the videos, it became clear that the newer social media functions YouTube enables were mostly irrelevant to successfully gaining viewers for the Tibbetts videos. When considering the total number of videos actually uploaded to YouTube each day, 72 hours of video every minute,[2] the number of the harmlessly banal to genuinely educational videos with very few views exceeds the number of cute and bizarre outliers that receive so many views through social propagation.

Receiving a substantial volume of traffic generated through users embedding videos in Facebook, blogs, or commercial websites may be the distribution path for some videos that "go viral," but the Tibbetts videos received a large majority of their traffic, over 75%, from within the YouTube ecosystem. Embedded content, videos placed on a blog or website, or direct traffic, videos referenced by sharing a URL, account for only 23% of the traffic over the life of the collection.

Utility 3: Benefiting from YouTube Usage Behaviors

For the majority of the views that the Tibbetts content received, one can only speculate that the very nature of YouTube browsing, doing a search, selecting a video to watch, noticing relevant recommended videos by YouTube to the one selected, and then clicking on the next video, and then the next, and so forth, resulted in the majority of the traffic that the Tibbetts videos received. In order to prove this hypothesis, one would either need to conduct behavioral studies of YouTube use or gain access to and analyze much more granular data about all instances of traffic related to the videos. On a more limited

scale, YouTube tracks data indicating which video from within the YouTube environment a user viewed prior to accessing a video in the collection. This kind of browsing, via "YouTube suggested video," accounts for over 50% of the views for the collection. Ironically, this kind of browsing behavior precedes the domination of search on the web and parallels similar models of the ways that librarians usually think of users browsing their physical collections.

In this respect, YouTube helps solve a notable limitation of video or other multimedia in digital repositories since they cannot be indexed in the same way that full text files can be. One common denominator of library collections on the web is that they typically live many clicks away from the libraries' homepage. As a result Libraries make great efforts to make sure that their content is indexed, making the content easily discoverable through a search engine. If prospective patrons construct the proper query, it is likely that at least some portion of library collections will end up in relevant search results via library catalogs or a search engine, especially where texts that are machine readable are concerned. In the case of audio that essentially consists of a conversation, it may be possible to generate a transcript that reproduces most of the content in text form. However, most videos cannot possibly be fully captured and indexed in this way. In the case of the Tibbetts collection, transcripts could effectively summarize a large piece of the content, given that they are conversations. However, neither Tibbetts nor KU Libraries have the time or resources to commit to transcribing over 1,200 interviews. Placing the videos on YouTube enhanced their discoverability without making large additional investments.

Utility 4: Tracing Interest in Specific Content

Over the life of the collection there have been several events that can be traced back through the metadata to their original post on the web. For example, the noted media scholar Henry Jenkins posted several of the videos on his blog along with a narrative provided by Tibbetts.[3] In a separate instance, the motion picture visual effects artist, Gavin Rothery, embedded the Blade Runner interviews in a post on his blog.[4] These are two examples of human intervention that demonstrate the potential for social tools to propagate content and may allow scholars or curators to engage with interest groups, scholars, or experts both inside and outside academia. Prior to this project one might guess that the assistance of scholars or media professional promoting content would increase traffic significantly, these kinds of posts usually only result in a sharp spike in traffic.

Gaining access to YouTube's larger audience can help track mass phenomena resulting in increased views too. Such an increase in views was created by the release of the Muppet Movie in late 2011. The release of

the movie appeared to coincide with a more gradual and even increase in views of the interview with Jim Henson and the Tibbetts video periodically appearing in the top 10 hits in YouTube searches for "Jim Henson." Another spike occurred on April 8, 2013 for the video containing an interview with the recently deceased Annette Funicello. On that day the video received 9,595 views. Many librarians and archivists experience and respond to world events, increased use, and interest in parts of their collections. The Jim Henson example demonstrates how having already made content available electronically, the library was able to provide relevant and timely resources in sync with the current zeitgeist. While the Funicello example demonstrates how making more content available more rapidly could provide access to a wider audience in moments when a library could not possibly respond rapidly enough to demand.

Utility 5: Reaching the Right Audience

YouTube provides a demographic breakdown of viewers into several categories (e.g., sex, age, country). However, YouTube will not provide scholars wanting to provide access to their content with any sense of whether their peers in academe are viewing the content similar to the way one can track citations through tools like Google Scholar. The only two gauges through YouTube of the prestige of receiving attention is the volume of views and, the number of minutes each video has been played per day. In fact those weary of the kind of attention some items may receive may take note that one video in the collection, an interview with Sean Young on Blade Runner, has received 25% of the views in the collection as a whole. One could speculate about why this is the case, but few would conclude the traffic was from over 100,000 scholars. Naturally an abundant audience is part of the result in opening up the collection to access points familiar to the wider public. For now, alternative metrics only supplement the benefits of more extensive studies of citation tracking (Kousha et al. 2012).

Whether the audience visits YouTube or DSpace to view or download content, the thousands of people that access a collection will vastly exceed points of comparison for other delivery mechanisms, such as journal subscriptions. The numbers point to the need for a reevaluation throughout the academic community about what counts for scholarship and how the impact of scholars on their field and society can be measured and evaluated. Alternative metrics for scholarly publishing have already begun to challenge the standard assessments of value throughout academia. Libraries however win with every single download and every single viewing of their material. Increasing those numbers in every ethically sound manner possible propels libraries and institutions of memory toward a new standard for dealing with multimedia in academic librarianship.

A Pragmatic Approach to Providing Access

At the University of Kansas, while several staff members have an interest in the access and preservation of moving images and other types of multimedia, no group or individual dedicates their time to these types of materials to the exclusion of other kinds of digital or material objects. In addition to dedicated personnel, a university library such as KU may have significant audio visual collections, such as those donated by Oldfather Studios where Centron Films (a prominent educational film company in the mid-20th century) was once located, but no significant block of resources to deal with the particular needs of such a collection systematically on an annual basis. Moreover, the infrastructure required to transfer such film or videos using the same kinds of standards by the motion picture industry remain prohibitive. Many organizations find outsourcing the transfer and creation of digital files the best and only solution to get quality deliverables with a minimum amount of wear on materials.

In the case of the Tibbetts materials, Tibbetts had already had his materials transferred before approaching KU Libraries. The materials themselves, various kinds of magnetic tape formats such as Beta SP, are still under his care, and he has retained rights to the materials. In effect KU Libraries has only been a steward of providing access to the materials, not the long term preservation and care of the original materials or even digital masters of the materials.

These details represent the kinds of factors that concern all collections that do not have dedicated staff. Often librarians consider best practices not in terms of the optimal outcome for the materials in the abstract, but for the materials and resources at hand. Tibbetts may or may not intend to eventually donate his materials to KU Libraries. In the present, despite not having some degree of control of the originals, KU Libraries decided that the solution that best meets his needs and the mission of the organization is to make his materials accessible. Many librarians and archivists that do not have dedicated staff to these types of materials may place these materials at the end of their cue of possible collections to approach next. Many institutions create lists of priorities that not only depend on the needs of a collection but the feasibility of making progress with a collection. Physical collections present similar challenges and also depend on the staffing and resources available.

The numbers of views gained through YouTube serves as a justification for a philosophy that manages to do the best with the available resources. Though this exercise revealed the obvious, the largest viewing audience in the world would likely create the largest amount of traffic, the success of the project rests with the totality of means by which different users with different kinds of behaviors may access the content and the relatively low cost of adding multiple distribution mechanisms beyond a repository once the initial work has been done.

NOTES

1. For access to the collection on DSpace and YouTube, see: http://kuscholarworks.ku.edu/dspace/handle/1808/6581 and http://www.youtube.com/user/jt283and1000
2. http://www.youtube.com/t/press_statistics (05/20/2013)
3. http://henryjenkins.org/2011/03/over_the_rainbow_john_c_tibbet.html(5/20/2013)
4. http://www.gavinrothery.com/my-blog/2011/12/21/blade-runner-cast-interviews-circa-1982.html(5/20/2013)

REFERENCES

Cho, Allan. "YouTube and Academic Libraries: Building a Digital Collection." *Journal of Electronic Resources Librarianship* 25.1 (2013): 39–50. Web. 9 May. 2013.

Kousha, Kayvan, Mike Thelwall, and Mashid Abdoli. "The Role of Online Videos in Research Communication: A Content Analysis of YouTube Videos Cited in Academic Publications." *Journal of the American Society for Information Science and Technology* 63.9 (2012): 1710–1727. Web. 9 May 2013.