Chapter 4
Exploring the Moving Image: The Role of Audiovisual Archives as Partners for Digital Humanities and Cultural Heritage Institutions

Adelheid Heftberger

Audiovisual collections still linger at the fringes of cultural heritage institutions such as galleries, archives, libraries and museums (GLAM). While digitization, cataloging standards and online presentation tools are constantly developing and improving in libraries, the situation in film archives is somewhat less satisfying. However, libraries and film archives are logical partners when it comes to the curation, publication, exploration, and interconnection of audiovisual online resources. In this chapter, I will describe collaborative work that I did with Lev Manovich, Professor in the PhD Computer Science program at The Graduate Center, CUNY, analyzing filmic structures in early Soviet films between 2007 and 2010. I will highlight the primary obstacles which so far have prevented more successful collaborations and projects between film archives and libraries. Additionally, I will present a positive example of interdisciplinary exchange within the field. Being a scholar that once worked in a film archive, I often found myself caught between two worlds. The upside of this is that I have been able to draw on experiences gained from my own research, as well as from interdisciplinary projects, particularly from the interaction with computer scientists.

For collaboration to be successful between different institutions and fields there are many strategic and economic decisions that need to be made. Publication and dissemination activities which are potentially useful for everybody have to be carefully coordinated. Modern universities still tend to separate “pure” scholarship from curation or archival work in general, with the latter reduced to a secondary, supportive role. However, this distinction is no longer appropriate. Many people working in cultural heritage institutions have scholarly backgrounds, yet they are rarely acknowledged as such amongst their peers (Hanley, Heftberger 2012). For my own definition of digital humanities (DH) I would like to draw on one of the earliest publications, the “Digital Humanities Manifesto 2.0” by Jeffrey Schnapp and Todd Presner. The Manifesto was democratic in suggesting an attempt to break up traditional hierarchies by, e.g., incorporating independent scholars who work outside an academic infrastructure. As Schnapp and Presner state (2009, 7), “[t]oday the old theory/praxis debates no longer resonate. Knowledge assumes multiple forms; it inhabits the interstices and criss-crossings between words, sounds, smells, maps, diagrams, installations, environment, data repositories, tables, and objects.” What is particularly noteworthy about the Manifesto, and what has largely been neglected since its publication, is the fact that the authors argue that DH will eventually contribute to tearing down “artificial” hierarchies between academia and cultural heritage institutions. Although Schnapp and Presner have provided many insightful and revolutionary thoughts concerning a new definition of the relation between universities and other institutions, such as archives and libraries, there is certainly still room for improvement, as I have already elaborated on elsewhere (2014). Recent debates in DH, however, seem to be reduced to discussing new ways of publishing scholarly articles, or to demarcations within the field, e.g., traditional versus DH, thus prolonging the familiar debate of quantitative analysis versus hermeneutical tradition. In addition, it appears as if funding agencies are making it increasingly difficult for nonacademics to be included as partners in research projects.¹

¹From a personal exchange with David M. Berry during the Graduate Seminar Digital Archives and Humanities: From Memory Curation to Innovation in Tallinn in 2016.
Partnerships should transcend the familiar setup of mere service provider (archive) on the one side, and the scholar on the other. Media scholar Lev Manovich, who coined the term “cultural analytics,” is among the pioneers working on projects that unite cultures and institutions. My own work with Manovich on the visualization of filmic structures in the work of the Russian/Soviet filmmaker Dziga Vertov (1896-1954) can serve as a model for successful collaboration between film archives and research institutions. I hope to show that successful and meaningful projects require a clear cut division of tasks according to job profiles. To restrict people in cultural heritage institutions to their traditional roles is, in my view, short-sighted and does not leave enough room for (professional) development and true collaboration.

4.1 Digital Humanities: Another Word for Information Sciences?

For me, DH is not so much about which tools the individual disciplines use and why, but rather a new form of scholarship where collaboration and interdisciplinarity are practiced by all involved. Thus, the following quote by Hoyt et al. (2016) appeals to me for its integrative approach to DH:

Rather than take digital humanities as a circumscribed field of research, pedagogy, and outreach, we understand DH as a strategically deployed term of mutual recognition that enables contemporary knowledge workers to signal a shared project interested in the relationship between digital technologies and humanities work. Of course, disagreement exists over what that project is. In a sense, we are all digital humanists. Article databases, online catalogues, search algorithms, word processing software, email, and course management systems already shape contemporary academic work in countless ways.

Defining the typical DH scholar is difficult, because they transcend disciplines and institutions. Nonetheless, DH has become something of a buzzword for a relatively recent development in academia, particularly among literary studies. On the other hand, humanities scholars are typically among the fiercest critics of DH initiatives, and tend to foreground the problems (of which there are indeed many) rather than the possibilities. DH centers and study programs, meanwhile, are flourishing all over the world, conferences are being held regularly and a lively discussion continues to manifest itself in a variety of online and print publications (including blogs). Since Franco Moretti’s daring proposal to include what he called “distant reading” into literary studies (Moretti 2005), an ongoing debate has arisen on whether or not the humanities should involve quantitative analysis, or if that would compromise their raison d’être as a discipline. However, leading figures in this debate underline the importance of combining “both worlds” by training people in more than one discipline (Unsworth 2004):

We will need English majors who have a background in logic, who can handle statistics, who do maths, if we are going to turn out a generation of disciplinary specialists who can bring the accumulated wisdom of the humanities to bear the computational contexts – perhaps in helping build ontologies for scholarly projects in disciplinary contexts, or building tools for data-mining in the context of humanities research.

This description could actually apply to the not insignificant number of information scientists already working in libraries and, unfortunately to a much lesser extent, in film archives. Building ontologies is, however, only one possible area of (inter)action. Instead, I would like

---

3 For example, in Germany a new funding line, “E-Humanities,” was announced in 2013, which led to various activities in terms of new study programs and centers for digital humanities.
to shift our perspective, away from perpetuating customer-service relationship models, e.g.,
between the data specialist in an archive and the historian.

Whatever our opinion of DH may be, the humanities nowadays face a very different situation when it comes to studying arts, culture and society. Tony Hey, Research Officer at Microsoft, has coined the term “fourth paradigm” to denote our current data driven society, where scholars increasingly perform data intensive science, which according to Hey succeeded experimental science, theoretical science, and computational science. Hey rightfully claims that “science must move from data to information to knowledge.” The use of so-called “big data” presents a challenge to both humanities scholars and GLAM institutions. Previously, the objects of study for humanities scholars were often to be found in libraries, archives, and museums. Scholars could either examine these objects on site or, in some cases, were provided with digitized copies on demand in low quality strictly, for viewing purposes. Nowadays, more than 300,000,000 photos are shared on Facebook and 80 million photos on Instagram every day (Manovich 2016). Schnapp (2013, 10) puts these figures into perspective with his statement that, “[e]very two minutes we now take as many photographs as were taken during the entire 19th century.” The situation is similar for audiovisual material, as Luke McKernan, lead curator of news and moving images at the British Library, writes on his blog: “I estimate that there have been 2.7 billion videos uploaded to YouTube since 2005. 400 hours of video are added to the site every minute.” In reality, most film archives are not in a position to archive or catalogue these videos systematically, as they already have an immense backlog of films still to be catalogued and digitized and won’t be finished any time soon.

Humanities scholars can now employ digital tools and techniques in their research, like annotation, data mining, text linguistics and audiovisual analysis. While the digital revolution has proved overwhelming for some, it has been, and continues to be, enthusiastically welcomed by others. Some disciplines such as literary studies or linguistics have integrated computer aided analysis to a large extent. Other disciplines, however, are more reluctant. This is especially true of film and media studies, where conceptual discussions are still in the formative stages, e.g., the theory and practice of video annotation is only now being developed by infrastructure networks that include information specialists as valuable partners, such as DARIAH-DE in Germany or CLARIAH in the Netherlands. I would argue that GLAM institutions are logical collaborators in this process, owing to their preoccupation with the organization of metadata and the building of information structures. On the other hand, not every GLAM institution has information specialists among its staff, and those that do are more often than not unable to spare them for humanities scholars’ projects. Ideally, these projects also involve computer scientists, as even the most technically savvy humanities scholar is unlikely to have the same grasp of tasks such as programming and coding. And while I can appreciate the benefits of an idealized DH scholar with a broader professional range and expertise that transcends disciplinary boundaries, I would nonetheless argue for a division of tasks within interdisciplinary teams, especially when it comes to projects on a larger scale.

On the other hand, we will inevitably need scholars who are able, to a certain extent, to use statistics, data analysis, and visualization software in their daily work. This will enable them to combine quantitative and qualitative approaches in their work if necessary. Archivists will also need to understand data, and what can be done with it. Both archivists and scholars ultimately need to bring a wide array of qualities to the table, above all curiosity and the willingness to learn from other scientific fields, in order to find a common language and goals.

---

4.2 Galleries, Libraries, Archives, and Museum Institutions as the Logical Partners

The reasons why libraries, at least those in the United States, have embraced DH more willingly than archives and museums is summed up in the “Special Report: Digital Humanities and Libraries” (Varner, Hswe 2016): “the values of librarianship inform a deep interest in information access, a concern for information preservation, and a desire to make room for our diverse user communities.” Regrettably, cultural heritage institutions tend to see themselves as service providers, rather than as equal partners with their own research agenda. Schnapp remains one of the few scholars who continue to develop projects in which cultural heritage institutions figure as equal partners. By now he prefers to engage in what he calls “knowledge design” as an overarching concept that encompasses disciplines as well as institutions. He sees the current situation in the humanities as one of experimentation, rather than of using clear-cut methods. According to Schnapp (2014), there are new challenges arising, such as “how to construct arguments that zoom back and forth between the micro, the meso, and the macro, perhaps even overleaping those middle layers of analysis and narrative that once constituted the home turf of the arts and humanities disciplines?” He suggests the following fields of activities as possible entry points (ibid.): storied collections (innovative ways of working with and across collections), social lives of things (multimedia approaches to the description and representation of three-dimensional objects as networks of relations), new learning containers (rethinking learning spaces and models), and ubiquitous curation (the world as laboratory).

As I have stated previously, cultural heritage institutions can contribute to this ongoing process of designing knowledge by providing not only documents, but also expertise, in metadata and information science. They are the experts when it comes to preparing objects for digitization, and can provide reliable data (and metadata). On the other hand, in doing so, they can also profit from software as an aid to cataloging, organizing, exchanging, and publishing their archival holdings. Metadata for archival catalogues needs to be added, cleaned, and enriched on a comprehensive basis, and I can think of several ways in which digital tools (and DH scholars) could be useful here, e.g., cleaning metadata, facilitating metadata exchange across film archives and libraries, performing automated indexing and abstracting, importing data from relevant web sources (such as DBPedia, Wikidata, IMDB etc.), providing and utilizing Linked Open Data. However, audiovisual archivists are faced with multiple challenges which slow down and sometimes even prevent these developments outright. To give just a few examples: firstly, film is multimodal. Analogue film usually consists of two separate elements just for image and sound. Secondly, rights issues impede both the preservation of and access to archival documents by third-party users. Thirdly, the exchange of metadata and information amongst film archives has traditionally, and for various reasons which shall not be elaborated here, been a delicate matter, and the prevailing situation is only slowly changing.

It is important not only to strengthen the ties among the cultural heritage institutions in order to jointly solve these challenges, but to reach out to academia as well. In many cases, scholars can be valuable allies when it comes to making heritage public and meaningful by, e.g., taking an active interest in little known or unidentified parts of archival collections.
4.3 Collaboration with Lev Manovich in and Beyond the Project “Digital Formalism”

The interdisciplinary project “Digital Formalism” was carried out between 2007 and 2010 as a joint collaboration between the Department of Theater, Film and Media Studies at the University of Vienna, the Technical University of Vienna, and the Austrian Film Museum as the archive partner. Matthias Zeppelzauer, computer scientist and one of the project partners, sums up the different working approaches from his point of view (Olesen, forthcoming 2017) thusly:

At some point, interestingly, the project's direction became influenced by the demands and requirements of the computer scientists much more than anticipated. The reason was that the computer scientists required precise and complete annotations of the films for quantitative evaluation of the algorithms, and these annotations did not exist. Subsequently, they were generated by the film scientists and the archivists in the project who provided the necessary background information and knowledge about the film material. This stimulated research in the computer science domain and led to a shift from qualitative analyses to quantitative evaluations. Especially the archivists recognized the great potential of the annotations for visualizations and for developing novel perspectives on the material.

In the project, I annotated eight of Dziga Vertov’s films using the free software ANVIL, and the data gained served as the Ground Truth for subsequent computer aided analysis. The basic aim was to gain insight into the highly formalized artistic work of the director by applying quantitative and formal analysis, as well as close readings, and to correlate the data with surviving original documents from the collection of Dziga Vertov materials held at the Austrian Film Museum. It was essentially a rare case of a successful and internationally hailed collaboration between an archive and a research institution, in which a substantial amount of archival holdings could be shared in digital form with a scholar, in this case Lev Manovich. This facilitated not only the scholarly exchange between members of different “fields”, but also lead to the creation of a series of innovative visualizations that dealt specifically with the representation of filmic structures. Many of these visualizations were included in Manovich’s article “Visualizing Vertov” (2013), as well as in my own PhD thesis and subsequent book publication “Kollision der Kader” (2016). In this book, I describe the entire process, and use the visualizations for my own interpretations and analyses of Dziga Vertov’s films. Many of the visualizations are also available online, and can be viewed and explored there.

---

6 For more information, see: https://www.flickr.com/photos/culturevis/albums/72157622608431194 (last accessed: 23.3.2017).
This collaboration was possible due to a number of fortunate circumstances. Firstly, the films are all rights free. Secondly, the expensive digitization of the analogue film elements could be financed through the “Digital Formalism” project. There are two ways one can digitize a film, either frame-by-frame, or via a continuous process which is commonly referred to as “telecine.” The latter is the cheaper and faster of the two options, but also the more problematic as the equipment, in Europe at least, runs at a default speed of 25 frames per second. From the raw telecine video transfers, an image sequence was therefore exported containing several thousand individual image files per film (with each image corresponding to one frame on the analogue film strip). This procedure was quite unusual at the time, but it enabled us not only to carry out the manual annotation work as defined within the project, but also to create visualizations using the free software ImageJ, which Manovich regularly uses in his work. Thirdly, the archive partner trusted Manovich to use the material only for scholarly purposes and, while this would be a normal case of fair use in the United States, it is important to note that there is no equivalent in European copyright law. Such a decision can thus only be taken if those in charge (e.g. archive directors, project leaders) are open to collaboration and can see the potential benefits.  

---

7 For this, I am grateful to the project team, in particular to Michael Loebenstein and Alexander Horwath, both of the Austrian Film Museum.
Fig. 2: Montage of the first frame of each shot and respective shot lengths underneath (as number of total frames) divided into three 300 meter reels (5 reels in total) of “The Eleventh Year” (Dziga Vertov, 1928). Detail of the whole visualization for increased visibility. The bars underneath the frames represent the shot length created by manual annotation. Source material provided by the Austrian Film Museum, Vienna. Visualization by Lev Manovich/Software Studies Initiative.

Fig. 3: Montage of the first frame of each shot and respective visual activity of “The Eleventh Year” (Dziga Vertov, 1928). The longer the bar, the more vivid the action within the shot. The movement was calculated automatically by Lev Manovich. Source material provided by the Austrian Film Museum, Vienna. Visualization by Lev Manovich/Software Studies Initiative.

The visualization of archival holdings has subsequently become more widespread since my initial collaboration with Manovich, and there is an ongoing interest in our work and the visualizations. A number of interesting projects have been developed by cultural heritage
institutions such as the German Digital Library, the New York Public Library or the Harvard Art Museums. However, it was the specific focus on visualizing film works on a micro and meso level, taking both scholarly expertise and archival knowledge into account, which made our work so pioneering.

4.4 Conclusions

With DH the question of high quality source material has once again become relevant. There is nothing wrong per se with using files extracted from a commercial DVD, or downloading files from internet sources. However, for scholars like myself it has proven necessary to work with digital copies made from reliable sources under carefully controlled conditions. Working within interdisciplinary groups makes the research process more transparent, as communication has to be more explicit. Having archivists on the team guarantees reliable data. Drawing on archival experience is not only good scientific practice but, in the case of the “Digital Formalism” project, it also informed our research and my own subsequent work with Manovich. Technical knowledge of analogue film prints, as well as the production, screening, and archiving of analogue film, proved not only beneficial when it came to the annotation and analysis of the films, but even sparked new research questions (Heftberger et al. 2009).

Nonetheless, I would like to argue for a more formalized working relationship between GLAMs and research institutions, so that reliable archival documents can be made available for further research in sufficient quality. In order to have more material at our disposal for the kinds of studies mentioned here, and to seriously engage with big corpora of (high quality) moving images, there are a number of obstacles which must first be overcome. Firstly, cultural heritage institutions require continuous and sufficient financial support if they are to be able to preserve and digitize their collections. They also need the support of the academic community in their fundraising efforts, and to make their work meaningful and visible. In the end, only through collaboration can the film heritage be made available for scholars and the general public alike. Secondly, there is a need to establish research infrastructures where archivists and curators can participate at equal level to the scholars, rather than being viewed merely as the content providers.

In order to understand the scholars’ needs, archival staff members must ideally be scholars themselves. My collaboration with Manovich initially came about because we both consulted Cinemetrics, a well-established collaborative online project for quantitative film analysis. Not only were we both willing to spend considerable time outside our usual work schedule to exchange ideas and create something together, but by sharing the same network, interests and goals, mutual trust and respect for different professional backgrounds and knowledge cultures could be built. The same applies to working in interdisciplinary teams. Furthermore, if the archivist is part of the scholarly circle, he/she can come up with new ideas for research projects involving cultural heritage institutions that build bridges to the archival community, who may not immediately see the benefit of collaboration. This allows archivists to express their own needs, to raise awareness for possible applied research fields, and to draw on academic expertise in a “friendly” environment and a structured set-up. Sometimes

---

8 For more information, see: [http://infovis.fh-potsdam.de/ddb/index_en.htm](http://infovis.fh-potsdam.de/ddb/index_en.htm) [last accessed: 23.3.2017].
9 For more information, see: [http://publicdomain.nypl.org/pd-visualization/](http://publicdomain.nypl.org/pd-visualization/) (available for download and re-use) [last accessed: 23.3.2017].
10 For more information, see: [http://metalab.harvard.edu/](http://metalab.harvard.edu/) [last accessed: 23.3.2017].
11 Cinemetrics was founded by the film scholar Yuri Tsivian ([www.cinemetrics.lv](http://www.cinemetrics.lv)) and consists to a large degree of manually produced data on the average shot lengths of films. The website functions as a hub for all those interested in exploring formal film analysis and participating in the network.
very practical problems may have to be solved, while other topics require a more substantial
exchange on a theoretical level. For example, visualization techniques could prove extremely
useful to film restorers but have yet to be developed for this purpose.

Manovich was very open to take input from “the archivist/scholar”, and modified his
visual representations of the filmic structure accordingly. As a result, valuable information
for film historians, e.g., how an entire film was split over single reels, was made visible for
further exploration. Similarly, the computer-aided comparison of film prints presents
challenges to scholars and archivists alike, as has been shown in the “Digital Formalism”
project (Zaharieva, Breiteneder 2010). Additionally, large-scale applications could prove
useful to cultural heritage institutions.

Finally, I would argue that in a data-driven world, the roles within the GLAM
institutions have become increasingly volatile and diverse. While this might seem scary and
demanding at first glance, there is huge potential in rethinking standard workflows and the
long-standing tasks of cultural heritage institutions to “collect, preserve and show.”
Personally, I welcome this development as I think the complex tasks involved, whether in the
daily business of GLAM institutions or research projects, demand different levels of expertise
and job profiles. Creating favorable conditions for working together on every level would and
should be advantageous for everyone.

Literature

Heftberger, A., Tsivian, Y., Lepore, M., 2009. Man with a Movie Camera under the Lens of
Cinemetrics. Gruber, K., Wurm, B., Kropf, V. (Eds.), Digital Formalism. Die kalkulierten

descriptions and the challenges of the digital era. MedieKultur – Journal of media and
communication research 30, 57, 135–153. Available from:
http://ojs.statsbiblioteket.dk/index.php/mediekultur/issue/view/2200. [last accessed:
23.3.2017].

Hanley, O., Heftberger, A., 2012. Scholarly Archivists/Archival Scholars: Rethinking the
Traditional Models. The Velvet Light Trap, no. 70, Fall, 64–65.

Acland, C. (Eds.), The Arclight Guidebook to Media History and the Digital Humanities,
Sussex, 1–29.

Manovich, L., 2013. Visualizing Vertov. Available from:
http://softwarestudies.com/cultural_analytics/Manovich_Visualizing_Vertov.2013.pdf [last
accessed: 23.3.2017].

Manovich, L., 2016. Challenges of Digital Art for our Societies, online lecture, published
March 8, 2016. Available from: https://www.youtube.com/watch?v=sYSd0hoBFZA [last
accessed: 23.3.2017].

London.


