Analyzing Audio/Visual Data in the Digital Humanities

Taylor B. Arnold  
*Mathematics and Computer Science, tarnold2@richmond.edu*

Lauren Tilton  
*University of Richmond, ltilton@richmond.edu*

Follow this and additional works at: [https://scholarship.richmond.edu/mathcs-faculty-publications](https://scholarship.richmond.edu/mathcs-faculty-publications)

**Recommended Citation**

This Book Chapter is brought to you for free and open access by the Math and Computer Science at UR Scholarship Repository. It has been accepted for inclusion in Math and Computer Science Faculty Publications by an authorized administrator of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.
“With the inauguration of *Computers and the Humanities*, the time has perhaps arrived for a more serious look at the position of the humanistic scholar in the world of data processing,” wrote English scholar Louis Milic in the inaugural issue of the journal (Milic 1966). The opening issue’s invitation to participate immediately offered a capacious definition of humanistic inquiry. Under the editorship of Joseph Raben at CUNY’s Queens College, the opening issue explicitly called for a broad definition that ranged from literature to music to the visual arts as well as “all phases of the social sciences that stress the humane.” The centrality of fields such as music and art history were a given. “The music people have been the most imaginative,” Milic argued. In the second issue, art historian Kenneth Lindsey laid out the state of “Art, Art History, and the Computer.” “Within the past few years, we have witnessed the growth of interest in how sophisticated mechanical instruments can promote both the production of art and a better organisation of the data of art history,” he wrote (Lindsay 1966).

Audio and visual work as a primary source and therefore as data was central to their configuration of DH, which was known at the time as humanities computing. The new journal sought to harness “the phenomenal growth of the computer during the past decade” and demonstrate what could be possible when researchers across disciplines created, analyzed, and communicated their work at the intersection of their object of study and computers. The first issues radiate with excitement.

Fast forward forty years, and the possibilities of the computer could not be more apparent as signaled by the flourishing of what is now labeled as Digital Humanities (DH). However, the institutional configurations that led to “Humanities Computing” and now “Digital Humanities” consolidated around text and specific fields such as literary studies and linguistics. The logocentrism of the field is best demonstrated by the form and content of the field’s publications, which are an important and powerful gauge of credit and prestige because they reflect and shape a field’s priorities. For example, the official journal of the European Association of Digital Humanities, *Digital Scholarship in the Humanities*, was titled *Literary and Linguistic Computing* from its founding in 1986 until 2015. *Computing in the Humanities* relaunched as *Language Resources and Evaluation* in 2005 and situated itself squarely in a particular configuration of linguistics (Ide 2005). They reflect the narrowed configuration of the field in disciplinary terms by the early 2000s.

The centrality of text, specifically literary texts, to the term digital humanities led Lev Manovich to offer the term cultural analytics to signal media and contemporary digital culture as the object of study (Manovich 2020, 7). Of course, researchers have been working in these areas for decades. However, as humanities computing became digital humanities and then Digital Humanities with a capital D and capital H (Johnson 2018), the institutional structures from conferences (Weingart 2017) to journals (Sula and Hill 2019) to hiring have circulated around text and text-heavy fields. Yet, the tide is changing.

We offer an observation, perhaps even provocation: DH is undergoing an A/V turn. The combination of access to data either through digitization or born-digital sources alongside advances in computing from memory to deep learning is resulting in a watershed moment for the analysis of audio and visual data in the field. Decades of work have created streams including scholarship from academic disciplines such as media studies, new and expanded publication avenues such as the International Journal for Digital Art History and the renaming of journals like Literary and Linguistic Computing (LLC) to Digital Scholarship in the Humanities (DSH), and institutional shifts such as the founding of ADHO’s AVinDH SIG in 2014. Colleagues are advocating for DH within the structures of another institutional configuration such as specific academic disciplines and advocating within the institutional structures that form the field of Digital Humanities to see and feature A/V work. Our collective efforts to remove dams and enable streams across institutional structures by individuals, groups, and organizations are making space for audio visual data as an object of study as well as a form of scholarship.

The A/V turn brings together two streams of work in DH. The first is what one could call an audio or sound turn. While work on sound is not new in DH, in fact it enjoyed a great deal of prominence in the 1960s and 1970s, the last decade has seen a renewed interest (Sula and Hill 2019). Work includes commitments to sound archives alongside new DH methods such as distant listening. Exciting new publications such as Lingold, Mueller, and Trentien’s 2018 edited volume Digital Sounds Studies signal the emergence of “a new and interdisciplinary field … at the intersection of sound studies and digital humanities,” as Tara McPherson argues (121). While this area of work has not been articulated as a turn, the call to participate and listen is amplifying. As the co-editors write, “we need to bring the insights of sound studies to bear upon the emergent field of digital humanities” (Lingold, Mueller, and Trentien 2018, viii) and vice versa.

The second is the visual turn. In the last few years, there has been a combination of observations and calls claiming that DH is undergoing such a turn. Münster and Terras offer a survey on topics, researchers, and epistemic cultures in what they term the “visual digital humanities” (2020). Wevers and Smits argue that the “visual digital turn” is underway thanks to the affordances of computational advances in computer vision, specifically deep learning and neural networks (2020). For over a decade, media studies scholars such as Lev Manovich and Tara McPherson alongside digital art historians have pioneered the analytical possibilities of visual culture and DH, thereby forging the visual turn. Our own work has been a part of both turns as exemplified by the special issue of Digital Humanities Quarterly on A/V data that we co-edited (Arnold, van Gorp, Scagliola, and Tilton 2021). In this chapter, we make explicit what was implicit in our previous work. Bringing both calls together, we argue that an A/V turn is underway.

In the sections that follow, we begin to lay out a topology of A/V work in DH. The next two sections explore developments that have supported A/V data as a primary source. We begin with digitization and annotation and then turn to machine learning and deep learning. The following
section focuses on A/V as a form of scholarship. Formats such as mixtapes and video essays offer audio and visual ways to produce and communicate scholarship. We end with a return to our provocation and a brief discussion of the challenges and opportunities moving forward.

DIGITIZATION AND ANNOTATION

For decades, digitization and annotation have been instrumental to the field and opening up A/V data as a primary source of research. From archives to institutions of higher education to community-based nonprofits, groups are developing a steady stream of digital materials. How to organize, discover, and analyze through the affordances of digital technologies is a flourishing area of DH. The development of archives and collections as digital public humanities projects is a reason for and result of digitization. Along with facilitating access, a popular area is the development of annotation tools for adding context and information retrieval (Pustu-Iren, Sittel, Mauer, Bulgakowa, and Ewerth 2020). The development of approaches and tools for custom annotation facilitates close analysis for discipline deep-learning specific inquiry and pedagogy.

A prominent area of work is the digitization of materials. Goals include building digital archives and projects, facilitating access and discovery, and developing datasets for computational analysis. Major sites for digitization are galleries, libraries, archives, and museums, often referred to as GLAM institutions. Their commitment to access and discovery has been foundational to work in A/V in DH. They have pioneered digitization processes and models, researched and promoted the importance of collections, developed metadata frameworks such as schemas and ontologies that DH tools like Omeka rely on, and led initiatives to foster open access. For example, the National Film Archive of India has been collecting, researching, and promoting the importance of Indian cinema since 1964. The Akkasah: Center for Photography in Abu Dhabi is working to preserve the visual history of the Arab world while demonstrating how photography has shaped images of the region for over a century. Since 1926, the United States Library of Congress has been collecting sound recordings and remains a pioneer of preservation and digitization methods through the Motion Picture, Broadcasting, and Recorded Sound Division. The final example is the Rijksmuseum, which has set the standard for open access to digital images of museum holdings through their commitment to open access data, with major institutions like the United States Smithsonian Institution following suit. The increased availability of A/V materials and data is spurring research by making A/V data accessible.

As access to digitization equipment and funding has expanded and become less cost prohibitive, digitization and related digital projects are also finding homes in nonprofit organizations and within academia. Given the long history of A/V materials as secondary objects to GLAM institutions, efforts to collect, preserve, and digitize these materials also come from within academia. For example, the Southern Oral History Program at UNC-CH has been digitizing their oral histories of the American South that date back to 1973 to make them accessible. Started in 2015, PodcastRE is creating a “searchable, researchable archive of podcasting culture” and is based at the University of Wisconsin-Madison (Hoyt, Bersch, Noh, Hansen, Mertens, and Morris 2021). Among the most ambitious efforts is the Archives Unleashed Project housed at the University of Waterloo, which has been making petabytes of historical Internet content accessible for study since 2017. Projects such as these are demonstrating how DH is a key interlocutor in the creation, access, and discovery of
audio and visual data and facilitating scholarship while expanding what counts as scholarship and scholarly work in academia.

Efforts to build digital archives and collections around a particular topic now garner academic credibility. One form of these initiatives is digital public humanities projects (Brennan 2016). These projects are built around access, discovery, and interpretation, and designed to be accessible to a particular public. Such work is indebted to areas such as archive studies that have long argued that archives and collections are sites of power that shape memory and thereby which stories are told. DH researchers are coming together, often into teams built around multi-institutional collaborations, to create digital public projects around a shared area of study. Examples include the Amateur Cinema Database, Fashion and Race Database, and New Roots/Nueva Raíces: Voices from Carolina del Norte founded in 2017, 2017, and 2006 respectively as well as the quickly growing Covid projects around the world such as covidmemory.lu from C2DH. Importantly, tools developed by DH researchers have responded to and facilitated the development of this area. Among the most prominent is Omeka, which is an open-source publishing platform designed for creating archives, scholarly collections, and exhibitions. A repertoire of plugins includes functions for image annotators, mapping, and multimedia.

The final example we draw on is from our own work creating Photogrammar, which started in 2010. The project uses interactive maps and graphs to facilitate access and discovery of the 170,000 documentary photographs created during the Great Depression and the Second World War by a unit funded by the United States federal government. The Historic Unit of the Farm Security Administration and then Office of War Information, known as the FSA, offers a view into the era through the lens of the state as well as several of the most acclaimed documentary photographers of the twentieth century. We then use the same features to make scholarly arguments about the collection (Arnold, Leonard, Tilton 2017; Cox and Tilton 2019). Our remixing of the archive through data visualizations is designed to communicate new scholarship about the collection while creating space for users to pursue scholarly inquiry based on their questions.

Another aspect of DH that has relied on archives and digitization is annotation. The approach is driven by a plethora of analytical purposes. Among them are disciplinary-specific analysis, pedagogy, and information retrieval. For example, art historians alongside film and media scholars regularly conduct formal analysis. While art historians harness annotation to study composition and iconography, film and media studies scholars study mise-en-scène alongside features such as shot length and cuts. In the classroom, annotation tools facilitate close readings as participants use the built-in features and categories to learn how to engage in close analysis. Another line of work is using these annotation tools to develop metadata that facilitates access and discovery of A/V material through information retrieval systems such as browse and search.

A major area of scholarship in DH is the development of digital tools to support annotation. A prominent area is film and media annotation tools given the field’s focus on close analysis. Building off of linguistic annotation tools like ELAN, teams like the Film Colors Project have developed tools for the annotation of color in film over the last decade. Their work is part of a large ecosystem of projects such as Mediate and the Semantic Annotation Tool (SAT) that are developing tools specifically at the intersection of film and media studies and DH for close analysis and media literacy (Burges et al. 2021; Williams and Bell 2021). Many of these are also designed with pedagogical goals.

Another prominent application is sound annotation. Tools like the Oral History Metadata Synchronizer (OHMS), which was started in 2008, facilitate the creation of transcripts and metadata
from oral interviews. The tool can then be integrated with *Omeka*, which brings us full circle. *Omeka* facilitates the building of a collection, and participants can use the OHMS module to further develop metadata that can then facilitate access and discovery of the collection. The combination brings archive building and annotation together, often through digital public projects such as the Bluegrass Music Hall of Fame and Museum’s oral history project and digital exhibit, which is a partnership with the University of Kentucky’s Louie B. Nunn Center for Oral History. While we have mostly focused on manual annotation, newer tools such as our *Distant Viewing Toolkit* are now looking to scale up the analysis through machine learning, which we will turn to in the next section.

**MACHINE LEARNING/DEEP LEARNING**

Thinking of digitized and born-digital items such as music, paintings, and photographs as data has not been an intuitive move in the humanities (Posner 2015). Yet, many of the fields that DH draws on and is in conversation with, such as statistics, use this terminology (Arnold and Tilton 2019). One outcome is researchers involved in DH are increasingly thinking of the primary sources we study as data. The very addition of a “Data Sets” section in *Journal of Cultural Analytics* is a testament to this shift. The journal is also indicative of the prominence of machine learning in DH, which has not been without challenges.

The challenge has been twofold. One is the development of the data and metadata to study. For decades, researchers have been working to digitize and make features they want to study machine-readable, primarily through the construction of metadata or translation of a feature into text. Linguistic scholars have been at the fore, by manually tagging formal elements of spoken and written language through the construction of encoded corpora (Burnard and Aston 1998; Davies 2010). Manual annotation and related tools have been pivotal. The second development is the form of data that computers can process. Advances in storage and processing combined with new computational methods are opening up the scale of analysis as well as what can be analyzed. The advances in audio and image processing combined with the last several decades of digitization have opened new possibilities for large-scale data analysis. One of the most recent additions to the computational repertoire—machine learning and specifically deep learning—are facilitating the A/V turn in DH.

Machine learning involves the use of data to create and improve algorithms. Approaches such as topic modeling, named-entity recognition, and principal component analysis have become commonplace among DH scholars in areas such as computational literary analysis, visual digital humanities, and digital sound studies. A subset of machine learning in artificial intelligence known as deep learning has had a particularly large impact in opening areas of research in A/V DH (Arnold and Tilton 2021). They are expanding the scope and scale of analysis, particularly the kinds of evidence (Piper 2020). For example, art historians studying iconography in art through object detection, media studies scholars studying visual style and narrative structure in TV using face recognition, and historians using deep-learning algorithms to extract machine-readable text from scans of old newspaper pages.

An exciting development is that DH researchers are not just adopting these methods and their embedded theories as is but developing theories and frameworks that draw on and build off machine learning. We are now seeing the development of DH frameworks, methods, and
theories that bring together scholarship across fields such as computer science, film studies, media studies, and sound studies under new transdisciplinary spaces, including cultural analytics, distant listening, and distant viewing (Clement 2013a; Bermeitinger et al. 2019). They have led to calls for tools that reflect the interpretative possibilities and commitments of these theories and methods such as ImagePlot, HIPSTAS (Clement 2013b), and the Distant Viewing Toolkit (Arnold and Tilton 2020).

As an example of how researchers are combining method and theory to formulate scholarship that emerges specifically from DH, we turn to our own work. We situate distant viewing as a theory and method for the critical use of computer vision to study a corpus of (moving) images. The Distant Viewing Toolkit is built to facilitate two of the four parts of our DV method: annotate and organize. Rather than manual annotation, we draw on machine learning to automatically annotate at scale to pursue our scholarly questions about images, specifically twentieth- and twenty-first-century visual culture from the United States in our case. Along with building custom annotators, we then aggregate the annotations into features such as the categorization of shot length and detection of film cuts. In other words, the toolkit enacts the theory and facilitates the method. Such developments demonstrate how DH scholars are not only borrowing, but adjusting, remixing, and building approaches to machine learning embedded with and facilitating the kind of scholarly inquiry that animates the humanities. This scholarship is also a part of expanding the scope of scholarship in the field, from traditional articles to software, and joins an even larger ecosystem of DH researchers forging new paths of scholarly argumentation.

**FORMS OF SCHOLARSHIP**

An exciting area of A/V work is how researchers are pushing the forms of scholarship in the field. As scholars ranging from rhetoric to media studies argue, *how* we make our claims shapes the argument. Or, as Marshall McLuhan famously stated, the medium is the message. Or, at least, a part of the message. Exploration of how one can use A/V to argue and communicate insights about A/V is an exciting area in DH, which is forging an expanded notion of what counts as DH.

Work in this area comes in audio, visual, and audio visual formats. As researchers have long lamented, forms such as text articles and books often reduce the aural and visual complexity of A/V to words. There are times when those words approximate the explosion of meaning and feeling that a piece of art or stanza of music can elicit, but they cannot replicate the embodied process of knowing that seeing and hearing garner. A growing area of work now draws on the forms that we employ to communicate our scholarship.

Two areas include audio, specifically podcasting and mixtapes, and moving images, specifically videographic criticism. Along with being a form to reach wider audiences, podcasting is providing a space to weave in the sounds that are often the object of study. The rise of podcasting has created an exciting space to literally voice and listen. Podcasts like *Books Aren’t Dead*, which is a part of the Fembot Collective, are using audio technology to produce intersectional feminist DH (Edwards and Hershkowitz 2021). Scholars such as Tyechia Thompson are using DH in the form of pedagogical mixtapes to study and communicate knowledge about Afrofuturism (Thompson and Dash 2021). The second area comprises video essays and videographic criticism. As scholars such as Jason Mittell have argued, videographic criticism is a DH method to study moving images and sound which then uses the same forms to convey scholarly arguments (Mittell 2019, 2021). DH
on A/V is pushing boundaries by expanding our forms of scholarship and how we communicate scholarship in new and more expansive directions.

The final section that we will highlight is visualization. We return to Eric Champion’s point from 2016 that DH is “visualisation light and simulation poor.” An exciting development is that this is changing as forms of scholarship such as interactive visualizations become more prominent. The change is the result of a combination of factors that include new frameworks like React.js that have lowered the barrier to building interactive graphics, accessing A/V data, and increased credit in the academy that shapes which forms and kinds of scholarship are prioritized. Publications such as Reviews in DH, more capacious ideas of peer review such as grants, and awards like the American Historical Associations’ Rosenzweig Prize are expanding the kinds of scholarship that DH counts. Interactive data visualization projects such as American Panorama’s Mapping Inequality and Kinometrics, interactive 3D visualizations such as Virtual Angkor Wat and Victoria’s Lost Pavilion, and tools such as Image Plot are just several examples of projects that are pushing the boundaries of visualization as a way of knowing and form of scholarship in DH.

CONCLUSION

The sections above are just a few of the parts of the topology of A/V in DH. It is not exhaustive and reflects our angle of view, which comes with its own limits due to our positionality and subjectivity. However, we see this work as a part of recognizing and naming an exciting development in the field. The work outlined above alongside those that have been featured and organized in other spaces such as Digital Humanities Quarterly support our observation that DH is undergoing an A/V turn. It is an exciting time.

We end by returning to our framing of the A/V turn as a provocation. In spaces such as conferences, we have argued that DH has been focused on text and left little space for other ways of knowing, such as images, sound, and time-based media. The call has been met with discomfort and even hostility. Often, it is interpreted as a threat to text-based scholarship and specifically text analysis, which has enjoyed copious space in formal DH structures such as conferences and journals. We love this area of DH, and actively engage with it. However, we are calling for a bit more space for other kinds of knowledge and ways of knowing in these same DH structures. Often, there is pressure not to provoke but to be “nice,” which has created an affective and discursive culture where telling others to be nicer is a way to diminish or silence other calls within the field as we have seen with efforts such as #transformDH and #pocoDH. We think the tension that our provocation may produce can be productive and generative for it calls into question the status quo and offers other possible directions for the field.

So, we end with calling not just for more space, but for an A/V turn in DH. The consolidation of the field around text by the early 2000s led scholars like McPherson and Manovich to call for fields such as film studies and new frames such as cultural analytics to disrupt the logocentrism of DH. This groundwork combined with technological, institutional, and financial developments has facilitated a quickly growing area of scholarship around A/V in DH. The development and analysis of A/V data, and therefore its use as a primary source, as well as the creation of A/V work as a form and medium of scholarly interpretation and argumentation are two ways that DH is seeing an A/V turn. There is more work to do to imagine the areas of inquiry we could address and questions we could ask. What version of DH could we build if we moved beyond the field’s logocentrism?
What kinds of scholarship could we pursue? What kinds of collaborations and partnerships could we imagine? What if we then thought across text, image, and sound? What version of the field could we create? We won’t know unless we make more spaces to hear from other parts of DH as well as expand what counts. Let’s explore this turn together.

NOTES

1. For a longer history of the relationship between Art History and DH in the twentieth century, see Greenhalgh (2004).
2. In 2016, the Riksmuseum announced that 250,000 images of artworks would be released for free.
3. For example, see projects such as the Early African Film Database, the Amateur Movie Database, and Cinema, and the Santa Barbara Corpus of Spoken American English.

REFERENCES

Clement, Tanya. 2013b. “Distant Listening or Playing Visualisations Pleasantly with the Eyes and Ears.” Digital Studies/Le champ numérique 3 (2).


