1-1-2021

Exploring film language with a digital analysis tool: The case of kinolab

Allison Cooper
Bowdoin College

Fernando Nascimento
Bowdoin College

David Francis
Bowdoin College

DHQ: Digital Humanities Quarterly
2021
Volume 15 Number 1

Exploring Film Language with a Digital Analysis Tool: the Case of Kinolab

Allison Cooper <acooper_at_bowdoin_dot_edu>, Bowdoin College
Fernando Nascimento <fnascime_at_bowdoin_dot_edu>, Bowdoin College
David Francis <dfrancis_at_bowdoin_dot_edu>, Bowdoin College

Abstract

This article presents a case study of Kinolab, a digital platform for the analysis of narrative film language. It describes the need for a scholarly database of clips focusing on film language for cinema and media studies faculty and students, highlighting recent technological and legal advances that have created a favorable environment for this kind of digital humanities work. Discussion of the project is situated within the broader context of contemporary developments in moving image annotation and a discussion of the unique challenges posed by computationally-driven moving image analysis. The article also argues for a universally accepted data model for film language to facilitate the academic crowdsourcing of film clips and the sharing of research and resources across the Semantic Web.

1. Introduction

Today, decades after the earliest experiments with DH methodologies, scholars hoping to apply DH approaches to the study of audiovisual media continue to find themselves at somewhat of a disadvantage relative to colleagues working with text-based media. Impediments to computationally assisted analysis of moving images have been well documented and are both technological and legal in nature. In recent years, projects like Dartmouth's Media Ecology Project and the University of Richmond's Distant Viewing Lab, among others, have lowered technological barriers by making inroads into moving image annotation and the application of computer vision to moving image analysis. In 2018, the Library of Congress lowered legal barriers in the United States with the most recent round of exemptions to the Digital Millennium Copyright Act (DMCA), granting increased freedom to excerpt short portions of films, television shows, and videos for the purposes of criticism or comment and thereby removing a hurdle to DH-inflected forms of moving image analysis such as videographic criticism. Despite the advances described above, film and media studies scholars are still unable to analyze the moving images digitally that are the subject of their research with anywhere near the ease of DH practitioners working with text or other forms of data. 1

One illustration of this predicament is the ongoing lack of a database dedicated to something as seemingly straightforward as the analysis of film language. As Lucy Fischer and Patrice Petro lamented in their introduction to the 2012 MLA anthology *Teaching Film*, "the scholar of literature can do a keyword search for all the occasions that William Shakespeare or Johann Goethe has used a particular word, [but] no such database exists for the long shot in Orson Welles or the tracking shot in Max Ophüls" [Fischer and Petro 2012]. In response to the improvements to moving image access described above, the authors of this case study set out to develop Kinolab, an academically crowdsourced platform for the digital analysis of film language in narrative film and media (see https://kinolab.org/). This case study describes the opportunities and challenges that participants in the project have encountered in our efforts to create, manage, and share a digital repository of annotated film and series clips broadly and deeply representative of film language as the latter has evolved over time and across countries and genres. In this essay, we contextualize our project within related projects, recent efforts to incorporate machine learning into DH methodologies for text and moving image analysis, and ongoing efforts by AVinDH practitioners to assert the right to make fair use of copyrighted materials in their work. 2

Why should cinema scholars pursue DH approaches when, seemingly, they are so fraught with challenges? One answer to the question can be found in the methodology of a groundbreaking analysis in our field that took place before the first wave of DH scholarship in the 1990s and early 2000s and led to the definition of the group style known as classical Hollywood cinema [Bordwell et al. 1986]. Associated with narrative films made under the Hollywood studio system between roughly 1916 and 1960 and marked by certain recurrent features of narrative, narration, and visual style, classical Hollywood cinema has come to define our understanding of Golden Age cinema and to serve as a benchmark for scholarly inquiries into film form and style. Remarkably, however, the 100 films that made up the sample for the study comprised just a small percentage (roughly .006%) of the approximately 15,000 films produced by American studios between 1915 and 1960 (10). It is eye-opening to consider that such an axiomatic account of American film style and history excludes over 99% of the films produced in the period under investigation, even if, as Bordwell asserts, Hollywood classical cinema is "excessively obvious", having documented its style in its own technical manuals, memoirs, and publicity handouts (3). Today's film scholars may very well wonder how our understanding of this monolithic group style might evolve if we were to radically increase the sample size using DH approaches that didn't yet exist in the mid 1980s. 3

A related answer to the question of why cinema scholars might seek to incorporate DH methodologies into their work can be found on the IMDb Statistics page (see https://www.imdb.com/pressroom/stats/), which at the time of this writing included over half a million film titles in its database. Lev Manovich (2012) has argued that, before the global expansion of digital media represented by these kinds of numbers, "cultural theorists and historians could generate theories and histories based on small data sets (for instance, 'Italian Renaissance,' 'classical Hollywood cinema,' 'post-modernism', etc.)" but now we face a "fundamentally new situation and a challenge to our normal ways of tracking and studying culture" (250). For the Kinolab project, this new situation presents an opportunity to broaden our understanding of how film language works by creating a platform capable of sorting and clustering hundreds of aspects of film language along multiple dimensions such as region, genre, language, or period, among others. 4

We anticipate that our DH approach to the analysis of film language will allow researchers to move between different scales of analysis, enabling us, for example, to understand how a particular aspect of film language functions in the work of a single director, in a single genre, or across films from a particular time period or geographical region. We also anticipate that decontextualizing and remixing examples of film language in these ways will enable us to see what we might not have seen previously, following Manovich's assertion that "Being able to examine a set of images along a singular visual dimension is a powerful form of defamiliarization" (276). We argue that the collaborative development of a data model for film language, essential for the creation of a common understanding among cinema and media studies researchers as well as for their collaboration across the Semantic Web, will clarify and extend our knowledge of film language in the process of making its constitutive components and their relationships comprehensible to computers. And, finally, we expect that these efforts, made possible through the adoption of DH methodologies, will enable us to make more confident statements about the field of cinema studies at large. 5

2. Analyzing Film Language in the Digital Era: Related Projects

Our research has found few scholarly, open access projects dedicated to the digital analysis of film language – a situation likely due at least in part to the technological and legal barriers indicated above. Among the projects that do exist is the Columbia Film Language Glossary (FLG) (see https://filmglossary.ccnmtl.columbia.edu/), a teaching tool designed to offer users illustrated explanations of key film terms and concepts [Columbia Center for Teaching and Learning 2015]. It offers a relatively limited number of clips, with each clip selected to illustrate a single term or concept. This model, while well-suited to the project's pedagogical purposes, precludes users from making significant comparisons between different instantiations of film language. Search options are limited to film language terms and keyword searches, so the FLG does not offer the ability to do advanced searches with modifiers. Finally, it offers no means to research film language diachronically or synchronically. Conversely, Critical Commons (see http://www.criticalcommons.org/) offers an abundant source of user-generated narrative media clips, many of which include tags and commentary to highlight their use of film language. A pioneering project to support the fair use of copyrighted media by educators, Critical Commons accepts moving image media uploads and makes them publicly available on the condition that they are accompanied by critical commentary. This effectively transforms the original 6

clips by adding value to them and protects the users who upload them under the principles of fair use [Critical 2015]. Critical Commons was not designed intentionally for the analysis of film language; accordingly, the site lacks a controlled vocabulary or standardized metadata related to film language to facilitate search and retrieval, although users can execute keyword searches. Lastly, Pandora (see https://pan.do/ra#about) is a non-academic platform for browsing, annotating, searching, and watching videos that allows users to manage decentralized collections of videos and to create metadata and annotations collaboratively.

7 The efforts described above to make narrative moving image media available digitally for educational and scholarly purposes are complemented by projects developing promising tools for the digital analysis of moving images. Estrada et al. [Estrada et al. 2017] identify nearly 30 suitable tools for digital video access and annotation, evaluating in particular the professional video annotation tool ELAN and the qualitative data analysis software NVivo. While Kinolab relies upon a custom-built platform, ELAN and VIAN are two preexisting solutions that can be adapted to a variety of digital film analysis projects. ELAN (see https://archive.mpi.nl/tla/elan) is an annotation tool for audio and video recordings initially developed for linguists and communications scholars that has been adopted successfully by film studies researchers, whereas VIAN is a visual film annotation system targeting color analysis with features to support spatiotemporal selection and classification of film material by large vocabularies [Halter et al. 2019]. The brief overview that follows here concentrates more narrowly on current software and projects we have identified as best suited to work in a complementary way with Kinolab to support its focus on the digital analysis of film language. The Media Ecology Project (MEP), for example, develops tools to facilitate machine-assisted approaches to moving image analysis. These include, among others, a Semantic Annotation Tool enabling moving image researchers to make time-based annotations and a Machine Vision Search system capable of isolating formal and aesthetic features of moving images [Media Ecology Project 2019]. Similarly, the Distant Viewing Lab develops tools, methods, and datasets to aid in the large-scale analysis of visual culture [Distant 2019]. The Video Analysis Tableau (VAT) facilitates the automated comparison, annotation, and visualization of digital video through the creation of a 'workbench' – a space for the analysis of digital film – that makes available essential tools for the job but leaves the definition of the job itself up to individual media researchers and their collaborators [Kuhn et al. 2015].

8 Even as machine learning projects like the MEP and Distant Viewing Lab bring scholars of moving images closer to the kind of distant reading now being performed on digitized literary texts, their creators acknowledge an ongoing need for human interpreters to bridge the semantic gap created when machines attempt to interpret images meaningfully. Researchers can extract and analyze semantic information such as lighting or shot breaks from visual materials only after they have established and encoded an interpretive framework [Arnold and Tilton 2019, 2]: this work enables computers to close the gap between the pixels on screen and what they have been told they represent. The digital analysis of film language generates an especially wide semantic gap insofar as it often requires the identification of semiotic images of a higher order than a shot break, for example the non-diegetic insert (an insert that depicts an action, object, or a title originating outside of the space and time of the narrative world). For this reason, analysis in Kinolab for now takes place primarily through film language annotations assigned to clips by project curators rather than through processes driven by machine learning, such as object recognition.

3. From Textual Analysis to Moving Images Analysis in DH

9 A frequent topic in digital humanities concerns the balance between data annotation and machine learning. Manovich [Manovich 2012] rejects annotation for the purposes of Cultural Analytics (the use of visualization to explore large sets of images and video), arguing that the process of assigning keywords to every image thwarts the spontaneous discovery of interesting patterns in an image set, that it is not scalable for massive data sets, and that it cannot help with such data sets because natural languages lack sufficient words to adequately describe the visual characteristics of all human-created images [Manovich 2012, 257–262]. Notwithstanding researchers' increasing success in using computers for visual concept detection, the higher-order semiotic relationships that frequently constitute film language remain resistant to machine learning. When, then, should one annotate, and for what types of information? Projects and initiatives dedicated to text analysis, which is a more historically developed DH methodology, form an instructive continuum of the many ways in which manual annotation and machine learning techniques can be combined to retrieve information and perform digital corpora analysis. In many cases, digital projects rely solely on manually encoded digital

texts to provide their representational and analytical tools. Other models seek to add annotations on higher-level semantic entities such as spatial information [Pustejovsky et al. 2011], clinical notes [Tissot et al. 2015], and emotions [Alm et al. 2005]. A brief survey of the relationship between annotation and machine learning in text analysis provides insight into how this relationship may apply to time-based media and specifically to moving image analysis.

10 In the field of Natural Language Processing (NLP), annotations of parts of speech have greatly assisted in the advancement of text mining, analysis, and translation techniques. Pustejovsky and Stubbs have suggested the importance of annotation to enhance the quality of machine learning results: "machine learning (ML) techniques often work better when algorithms are provided with pointers to what is relevant about a dataset, rather than just massive amounts of data" [Pustejovsky 2012]. In another development of the annotation and machine learning relationship, some unsupervised machine learning models seek through statistical regularities to highlight latent features of text without the extensive use of annotations, such as the Dirichlet distribution-based models, including the model proposed by Blei et al [Blei et al. 2002] for Latent Dirichlet Allocation. Topic modeling has gained considerable attention over the last decade from the digital textual corpora analysis scholarship community. These models take advantage of the underlying structures of natural language coding forms. Despite its intrinsic semantic ambiguity, the code of natural languages textual structure follows syntactic patterns that can be recognized through algorithms that, for example, try to reproduce how texts are generated, following a generative hypothesis.

11 Even more recent advances in machine learning, especially in the area of neural networks and deep learning [Young et al. 2017], have opened new perspectives for data analysis with simpler annotation mechanisms. Deep neural networks have shown great success in various applications such as object recognition (see, for example, [Krizhevsky et al. 2012]) and speech recognition (see, for example [Sainath et al. 2015]). Moreover, recent works have shown that neural networks could be successfully used for several tasks in NLP [Cho et al. 2014]. One of the most used models in recent years has been word2vec, which represents semantic relations in a multidimensional vector space generated through deep learning [Mikolov et al. 2013]. This method allows the exploration of more sophisticated semantic levels without or with little use of annotations external to the text structure itself. More recently, models that use the attention mechanism associated with neural networks known as transformers [Vaswani et al. 2017]have empowered a new wave of advances in results on several areas of natural language processing such as text prediction and translation [Devlin et al. 2018].

12 These advances of digital text analysis seem to point to a trend toward a diminishing need for annotation to achieve results similar to or superior to those that were possible in the past with annotated data set training alone. However, despite the many advances we have described so far, there are still higher levels of semantic information (such as complex narrative structures or highly specialized interpretative fields) that require manual annotation to be appropriately analyzed.

13 From this brief exploration of the relationship between annotation and machine learning algorithms in the context of text analysis, we highlight three related observations. First, there has been a continuing and evolving interplay of annotation and machine learning. Second, recent machine learning algorithms have been reducing the need of extensive annotation of textual corpora for some interpretative and linguistic analyses. And thirdly, manual annotation still has a role for higher-level semantic analyses, and still plays an essential role in the training of machine learning models. With these three observations related to developments in text analysis, we are better positioned to understand a similar relationship in the context of time-based media. For this purpose, we take as reference the Distant Viewing framework proposed by Arnold and Tilton, which they define as "the automated process of making explicit the culturally coded elements of images" (5). The point, well noted by the authors, is that the code elements of images are not as clearly identifiable as the code elements of texts, which are organized into lexical units and relatively well-delimited syntactic structures in each natural language. Indeed, as Metz [Metz 1974] argues, film is perhaps more usefully understood as a system of codes that replace the grammar of language.

14 Thus, digital image analysis imposes the need for an additional level of coding – in Kinolab's case, curatorial annotations – so that the semiotic elements comprising film language are properly identified. As discussed earlier, Arnold and Tilton highlighted the semantic gap that exists between "elements contained in the raw image and the extracted structured information used to digitally represent the image within a database" [Arnold and Tilton 2019, 3].

Figure 1. "Hannibal and Clarice Meet" in *The Silence of the Lambs*. Directed by Jonathan Demme. Strong Heart/Demme Production, 1991. Kinolab, https://kinolab.org/FilmClip.php?id=726

Mechanisms to bridge this semantic gap may either be built automatically through computational tools or by people who create a system of annotations to identify these semiotic units. Moreover, these semiotic units can be grouped hierarchically into higher levels of meanings, creating a structure that ranges from basic levels of object recognition, such as a cake, to more abstract levels of meaning, such as a birthday party. Such analysis becomes more complex when we consider time-based media since its temporal aspect adds a new dimension to potential combinations, which adds new possible interpretations of meanings to images considered separately. An example taken from Jonathan Demme's *Silence of the Lambs* (1991) illustrates this challenge. In Figure 1, Anthony Hopkins as the murderous psychopath Hannibal Lecter appears to gaze directly at the viewer, ostensibly 'breaking the fourth wall' that traditionally separates actors from the audience. Both curator and a properly trained computer would likely identify this single shot – a basic semiotic unit – as an example of direct address or metalepsis, "communication that is explicitly indicated as being targeted at a viewer as an individual" [Chandler 2011], often marked by a character looking directly into the camera. But, as Figure 2 demonstrates, this single shot or basic semiotic unit is actually part of a more complex semiotic relationship that reveals itself to be *also* or *instead* an embedded first-person point-of-view shot when considered in the context of immediately preceding and subsequent shots. The shot itself is identical in both of these cases, but the film language concept that it illustrates can only be determined in light of its syntagmatic (sequential) relation to the shots that precede and follow it [Metz 1974] or other properties, such as an audio track in which direct address is or isn't communicated explicitly. This semantic ambiguity is a key component of the scene's success insofar as it aligns the viewer with the perspective of Lecter's interlocutor, the young FBI trainee Clarice Starling – an alignment that is felt all the more profoundly through the chilling suggestion that the spectator has lost the protection of the fourth wall, represented here through the metaphorical prop of the plexiglass partition separating the two characters.

15

The Distant Viewing framework proposes an automatic process to analyze and extract primary semantic elements "followed by the aggregation and visualization of these elements via techniques from exploratory data analysis" [Arnold and Tilton 2019, 4]. Based upon the evolution of digital text analysis following the new advances brought about by machine learning techniques described above, we predict that such evolving techniques will also allow the recognition and automatic annotation of more complex semiotic units, further narrowing the semantic gap for meaningful image interpretations.

16

Figure 2. Timeline showing embedded first-person point-of-view shot in "Hannibal and Clarice Meet" clip.

17 Kinolab creates a framework to explore the intermediate levels in this semiotic hierarchy by defining annotations that form a set of higher-level semiotic units of film language relative to basic units such as the cut or other types of edits and allows the description of common categories for understanding time-based media characteristics. Such semiotic units form the basis of a film language that describes the formal aspects of this type of digital object.

18 Kinolab is structured to help researchers reduce the semantic gap in digital film language analysis in three distinct ways. The most basic form is through a collaborative platform for consistent identification of semiotic units of film language in film clips, allowing sophisticated searches to be done immediately utilizing them. The Kinolab software architecture is also designed for integrating distant viewing plugins so that some film language forms can be automatically recognized by machine learning algorithms from the scientific community. This plugin would also allow subsequent exploratory data analysis based on Kinolab's archive. Finally, Kinolab can serve as a resource for applying, validating, and enhancing new distant viewing techniques that can use the database with information about film language to develop training datasets to validate and improve their results. Given Kinolab's architecture, it can produce a standard machine-readable output that supplies a given clip URL with a set of associated tags that a machine learning algorithm could integrate as training data to learn examples of higher-level semantic annotations, such as a close-up shot. What is lacking in Kinolab towards this goal is specific timestamp data about when a certain film language form is actually occurring (start/stop) which, combined with automatically extracted basic sign recognition (e.g. objects, faces, lighting), would be extremely valuable for any machine learning processes. The existing architecture could be expanded to allow this with the addition of a clip-tag relationship to include this duration information, however the larger work would be identifying and inputting this information into the system. One possible way to address this limitation is to integrate a tool like the aforementioned Media Ecology Project's Semantic Annotation Tool (SAT) into Kinolab. The SAT can facilitate the effort to create more finely grained annotations to bridge the gap between full clips and respective tags, providing a more refined training dataset.

19 With these extensions and within this collaborative ecosystem of complementary tools we believe that Kinolab could serve as an ideal platform for exploring the full spectrum of combinations between manual annotations and machine learning techniques that will foster new interpretative possibilities of time-based media in a manner analogous to advances in the area of digital text analysis.

4. Kinolab: A Dedicated Film Language Platform

20 Kinolab is a digital platform for the analysis of narrative film language yet, as previous discussion has suggested, 'film language' is a fluid concept that requires defining in relation to the project's objectives. The conceptualization of film as a language with its own set of governing rules or codes has a rich history that dates back to the origins of the medium itself. This includes contributions from key figures like D.W. Griffith, Sergei Eisenstein [Eisenstein 1949], André Bazin [Bazin 2004], and Christian Metz [Metz 1974], among many others. Broadly speaking, film language serves as the foundation of film form, style, and genre. Kinolab focuses on narrative film, commonly understood as "any film that tells a story, especially those which emphasize the story line and are dramatic" [Chandler 2011]. To tell a story cinematically, film language necessarily differs in key ways from languages employed for storytelling in other mediums. As the example drawn from *The Silence of the Lambs* demonstrates, this is particularly evident in its treatment of modalities of time (for example, plot duration, story duration, and viewing time), and space (for example setting up filmic spaces through framing, editing, and point of view) [Kuhn 2012]. Film language can also be understood as the basis for, or product of, techniques of the film medium such as mise-en-scene, cinematography, editing, and sound that, when used meaningfully, create distinctive examples of film style such as classical Hollywood cinema or Italian neorealism. Finally, film language is a constitutive aspect of genre when the latter is being defined according to textual features arising out of film form or style: that is, an element of film language such as the jump cut, an abrupt or discontinuous edit between

two shots that disrupts the verisimilitude produced by traditional continuity editing, can be understood as a characteristic expression in horror films, which make effective use of its jarring effects. Kinolab adopts a broad view of film language that includes technical practices as well as aspects of film history and theory as long as these are represented in, and can therefore be linked to, narrative media clips in the collection.

Our primary objective in developing Kinolab was to create a rich, DMCA-compliant platform for the analysis of narrative media clips annotated to highlight distinctive use of film language. 21

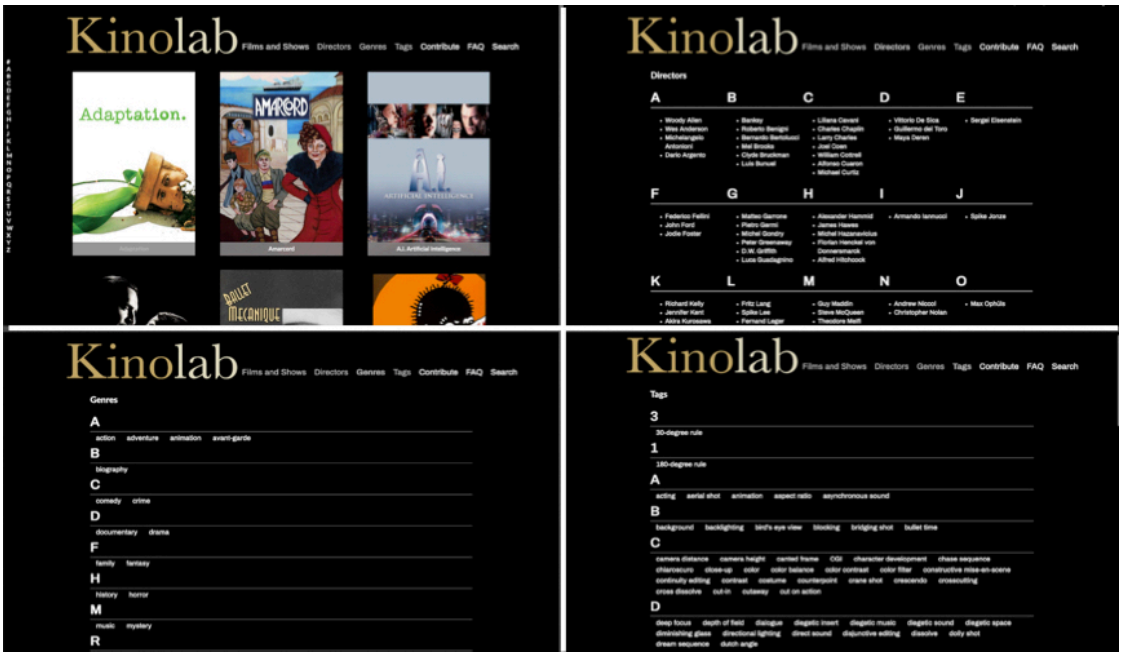

Figure 3. Principal entry points to the Kinolab clip collection.

The platform we envisioned would facilitate comparisons across clips and, to this end, feature advanced search options that could handle everything from simple keyword searches to searches using filters and Boolean terms. A secondary objective was to develop an easy-to-use contribute function so that users wishing to add their own legally obtained narrative media clips to the collection could do so with relative ease, thereby building into Kinolab the capacity for academic crowdsourcing. Ultimately, the simple design that we settled on invites verified academic users into the collection through four principal entry points accessed via the site's primary navigation (see Figure 3): Films and Series, Directors, Genres, and Tags. The terminus of each of these pathways is the individual clip page, where users can view a clip and its associated film language tags, which link to other clips in the collection sharing the same tag, and, if desired, download the clip for teaching or research purposes. Additional entry points accessed via the primary navigation bar include the Contribute (see Figure 4) and Search (see Figure 5) functions. Users can contribute their own narrative media clips via a simple interface designed to facilitate the curatorial process for project members working in Kinolab's back end. Academic crowdsourcing is standardized via a controlled vocabulary of film language terms (discussed further in Section Five: Working Toward a Data Model for Film Language). The Search function queries all of the fields associated with a clip in Kinolab's database, including informational metadata akin to what one would find in an IMDb film or series episode entry and content metadata supplied by Kinolab curators and contributors. Kinolab curators – project faculty, staff, and students – have access to the back end of the Contribute function, where they can evaluate and edit submitted clips and their metadata (informational and content metadata including film language tags) and approve or reject submissions to the collection. 22

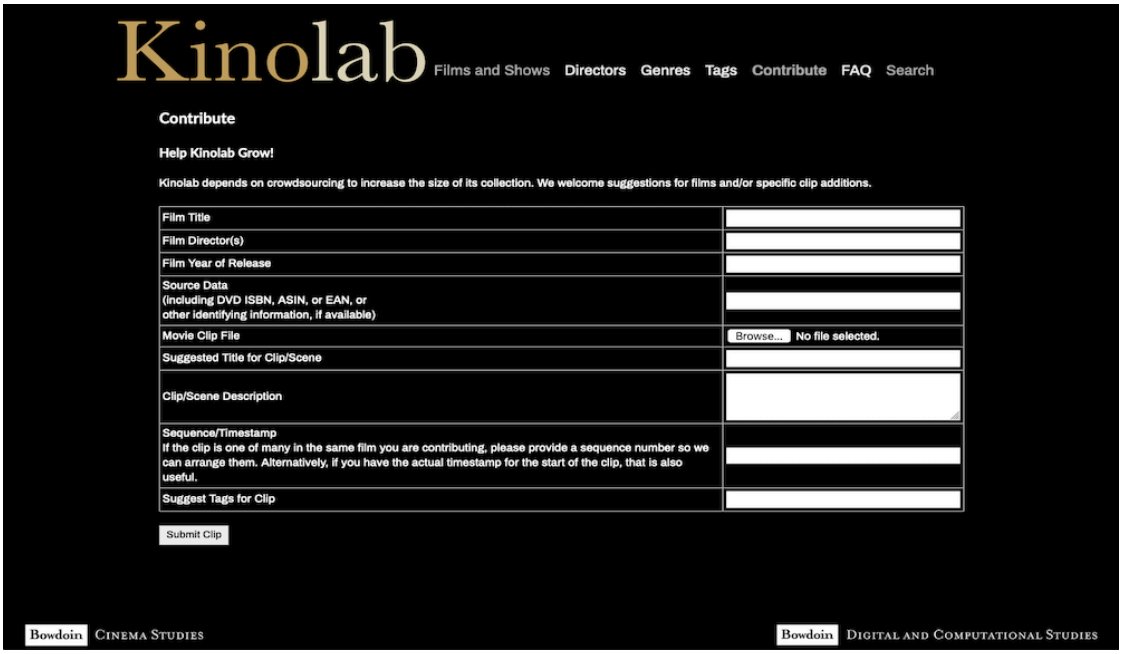

Figure 4. Kinolab Contribute page.

The vast majority of Kinolab's file system overhead goes to storing audiovisual clips. Accordingly, we built the first implementation of Kinolab on a system that could handle most of the media file management for us. Our priority was finding an established content management system that could handle the intricacies of uploading, organizing, annotating, and maintaining digital clips. To meet this goal, we initially adopted Omeka, a widely used and well-respected platform with a proven record for making digital assets available online via an easy-to-use interface (see https://omeka.org/). Built to meet the needs of museums, libraries, and archives seeking to publish digital collections and exhibitions online, Omeka's features made it the most appealing out-of-the-box solution for our first release of Kinolab. These features included: an architecture stipulating that Items belong to Collections, a relationship analogous to clips belonging to films; almost limitless metadata functionality, facilitating deep descriptive applications for film clips; a tagging system that made applying film language identifiers simple and straightforward; a sophisticated search interface capable of performing complex searches; and, finally, a built-in administrative backend capable of handling a significant part of the project's file and database management tasks behind the scenes.

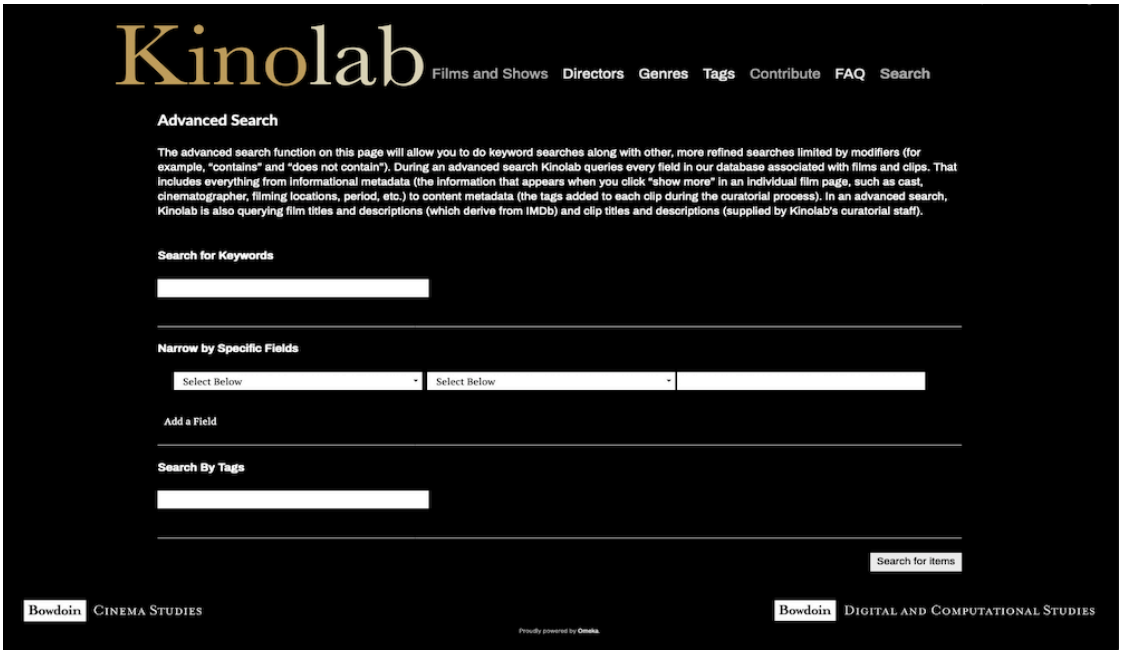

Figure 5. Kinolab Search page.

Omeka's ease of use came with some significant restrictions, however. Its functionality for describing Collections through metadata was far more limited than that for Items. This limitation makes sense for the cultural heritage institutions that are Omeka's primary users, which need extensive descriptive metadata for individual items comprising a collection rather than for the collection itself. In Kinolab's case, however, an Omeka 'Collection' was analogous to an individual film, and we struggled with our inability to attach key metadata relevant to a film as a whole at the Collection level (for example, cinematographer, editor, etc.). The constraints of Omeka's model became more pronounced as the project expanded beyond films to include series. This expansion entailed moving from a relatively straightforward Film-Clips relationship to the more complicated relationship between collections and items Series-Seasons-Episodes-Clips, which Omeka's generic model couldn't represent. The inclusion of series also confounded Omeka's search operation, which did not operate in a way that could factor in our increasingly complex taxonomies. As Kinolab grew, so did our need for functionalities that Omeka could not provide, ranging from the ability to select thumbnail images from specific video frames to the ability to specify extra relational concepts. Omeka's rich development community and plugins could have moved us toward some of these goals, but as we continued to add plugins and to customize the core feature set of Omeka, we were forced to recognize that the time and cost of the alterations were outweighing the benefits we gained from a pre-packaged system. Indeed, we had altered the base code so much that we could no longer claim to be using Omeka as most people understood it. That meant that upgrades to Omeka and its plugins could prove problematic as they could potentially affect areas of code we had modified to meet our goals.

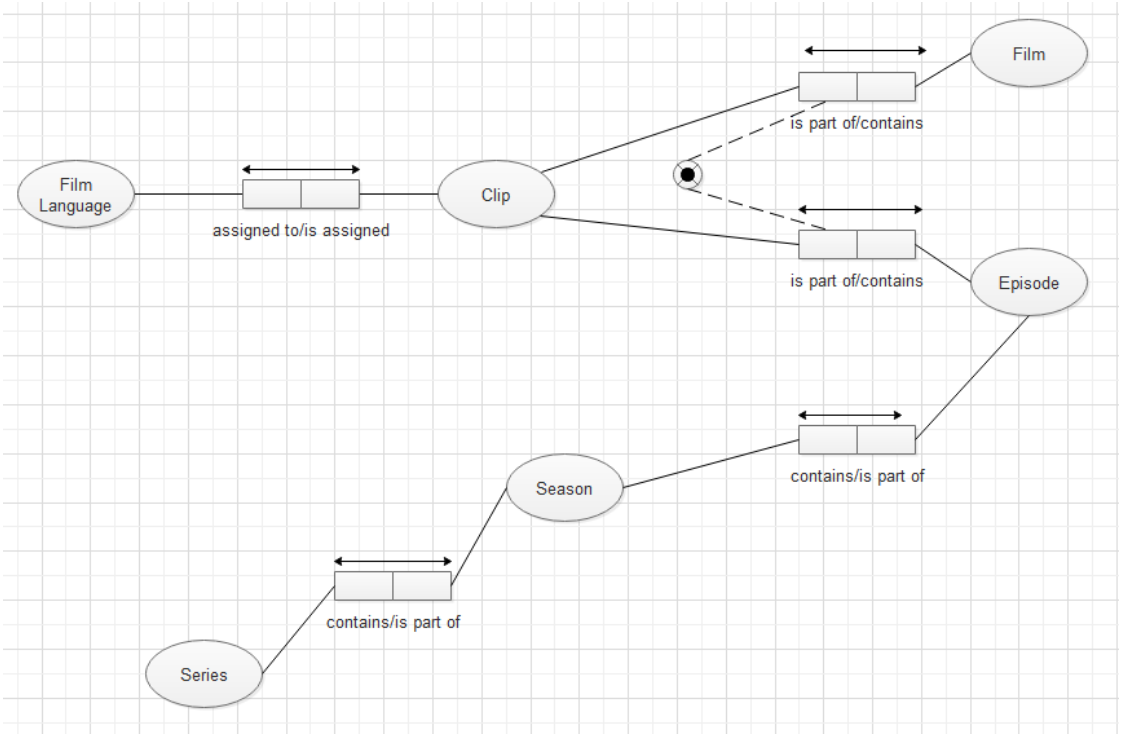

Figure 6. Basic object-relational schematic of Kinolab films and series.

Moving away from Omeka gave us the freedom to take the Kinolab concept back to the data modeling phase and define a database backend specifically for our project. We were able to implement the user interface collaboratively, module by module, with all team members, which helped flush out additional requirements and desirable features in easy-to-regulate advances. The system we ended up building used many of the same tools as Omeka.

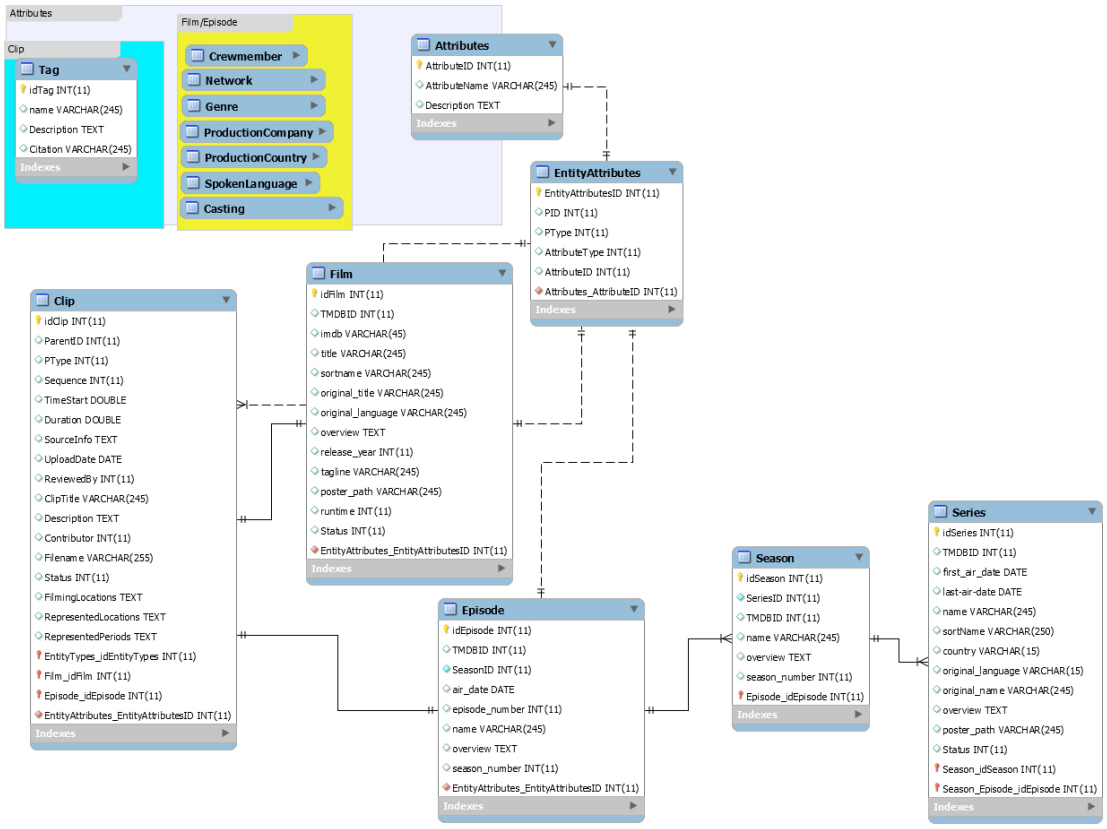

Figure 7. UML diagram of Kinolab's database.

The system requirements for Kinolab read much like those for Omeka and include a Linux operating system, Apache HTTP server, MySQL, and PHP scripting language. 26

Perhaps the most significant change that we made in the move from Omeka to a platform of our own design concerns metadata collection. In the first, Omeka-based implementation of Kinolab, project curators manually gathered informational metadata for films and series from IMDb.com and physical DVDs, subsequently uploading that metadata into Omeka's back end as part of a labor-intensive curatorial workflow. We eventually understood the project to be less about collecting media data than about aggregating annotations in service of film language analysis. We recognized that, if we were to continue attempting to collect and store all of the significant metadata describing films and series ourselves, we would be spending considerable energy duplicating efforts that existed elsewhere. This realization led us to partner with a third party, TMDb (The Movie Database) to handle the project's general metadata needs. For our new Kinolab implementation, we do store some descriptive data particular to the project in order to seed our search interface, but for the most part we rely on TMDb to be the actual source data and direct our users to that source whenever possible, enabling us to focus more narrowly on clip annotation. 27

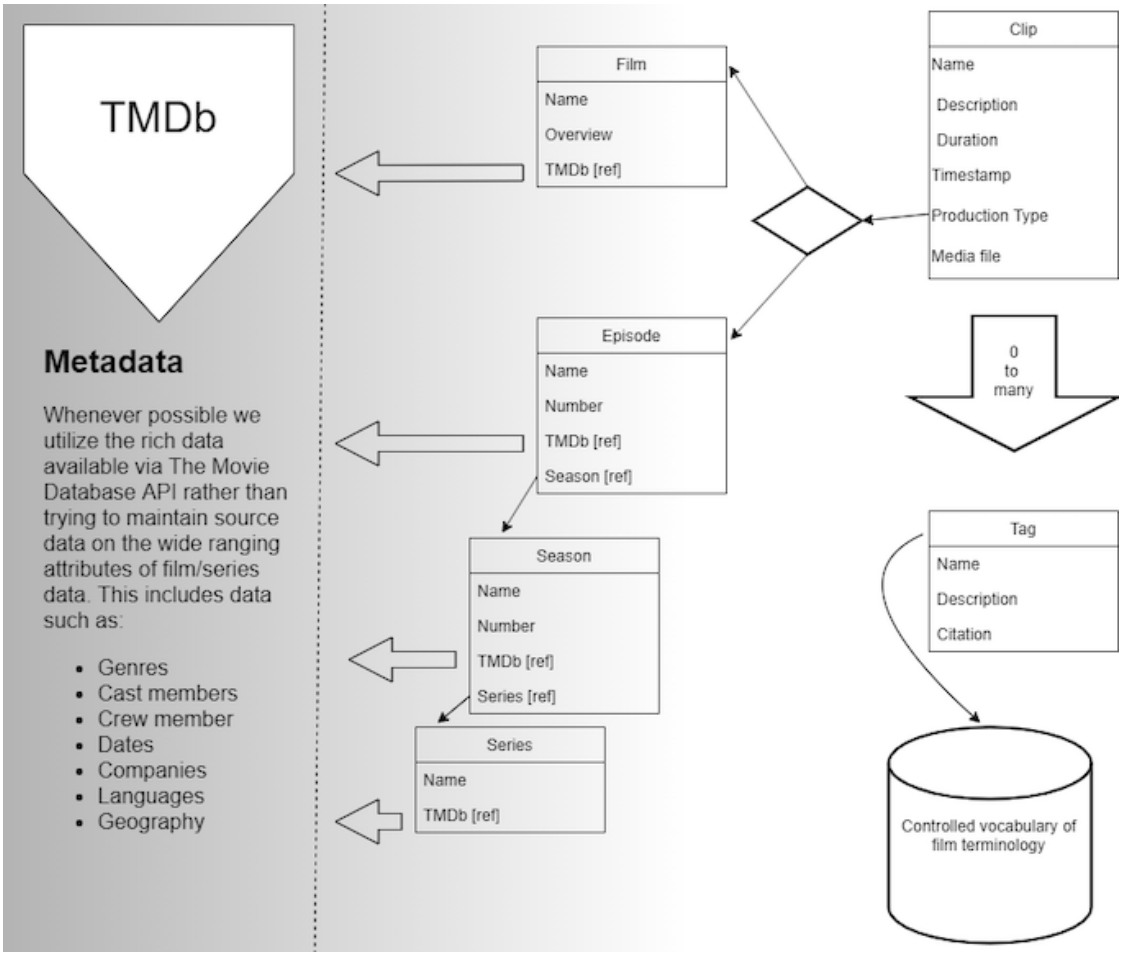

Figure 8. Kinolab metadata collection

Unlike IMDb, TMDb has a clear message of open access and excellent documentation. In testing, it offered as much and sometimes more information than one could access on IMDb. We have concerns about the long-term reliability of a less established source like TMDb over a recognized entity such as IMDb, but since we only make use of this data tangentially we decided that it is provisionally the best option. The metadata that TMDb provides is important for helping to locate and contextualize Kinolab clips, but the project is not attempting to become a definitive source for providing information about the films and series from which they are excerpted. Consequently, we simply reference this kind of metadata via TMDb's APIs or direct Kinolab users to the TMDb site itself. The lack of an accessible, authoritative scholarly database dedicated to narrative films and series is an ongoing problem shared by the entire field of media studies [Fischer and Petro 2012]. In the case of the Kinolab project, it has represented a challenge almost as significant as the legal and technological ones outlined elsewhere in this case study. 28

5. Working Toward a Data Model for Film Language

Early in Kinolab's development, we confronted a tension between the expansive concept of film language and the need to define it methodically for computational purposes. Problematically, clips initially contributed to the project, for example, could illustrate the same cinematographic concept using synonymous but different terms, complicating the indexing and retrieval of clips. For example, a shot in which the camera frame is not level with the horizon was defined differently (and correctly) by contributors as either dutch angle, dutch tilt, or canted angle. Alternatively, a clip might be identified with a single form of film language but not with its parent form. For example, the sequence shot, in which an entire sequence is rendered in a single shot, is a child of the long take, a shot of relatively lengthy duration, thus identifying the one ought to also identify the other. 29

Though different in kind, these and other related issues we encountered demonstrated the need to situate individual film 30

language concepts within a broader, machine-readable model of film language such as a thesaurus or ontology. The first case cited above, involving the interchangeability of dutch angle, dutch tilt, or canted angle, is a straightforward problem of synonymy, resolvable through the adoption of a controlled vocabulary for film language spelling out preferred and variant terms and including synonym ring lists to ensure Kinolab's ability to return appropriate clips when queried. The second case cited above, however, demonstrates the need to conceive of film language hierarchically. Both problems reveal how Kinolab could benefit from a data modeling approach capable of explicitly defining the "concepts, properties, relationships, functions, constraints, and axioms" of film language, akin to those proposed by the Getty Research Institute for art, architecture and other cultural works [Harpring 2013].

Our research revealed the lack of preexisting, authoritative models for film language. The International Federation of Film Archives (FIAF), for example, offers a "Glossary of Filmographic Terms" designed to assist film catalogers in the consistent identification and translation of credit terms, as well as a "Glossary of Technical Terms", for terms used in film production and the film laboratory, but neither resource could provide the kind of guidance we sought in organizing and deploying film language consistently. The Large-Scale Concept Ontology of Multimedia (LSCOM, see http://www.ee.columbia.edu/ln/dvmm/lscom/) is, for now, limited to concepts related to events, objects, locations, people, and programs and therefore lacking labels related to film form. The AdA Ontology for Fine-Grained Semantic Video Annotation (see https://projectada.github.io/) is promising for its focus on film-analytical concepts, but remains only partially complete. This led us to take an exploratory first step in that direction in the form of a controlled list of film language terms, drawn primarily from the glossaries of two widely adopted cinema studies textbooks, Timothy Corrigan and Patricia White's *The Film Experience* [Corrigan and White 2018] and David A. Cook's *A History of Narrative Film* [Cook 2016] (see https://kinolab.org/Tags.php for a complete list of terms). The controlled list currently includes approximately 200 aspects of film language and their accompanying definitions and serves to regulate Kinolab's academic crowdsourcing by ensuring that concepts are applied consistently across the platform. All metadata and particularly the application of film language tags are reviewed by Kinolab's curators before being added to the Kinolab collection. Annotation for Kinolab works by allowing a curator to define a one-to-many relationship of a clip to a limitless number of tags, bounded only by the number of available tags in our controlled list. Tags are linked to the clip by reference only, so if there is a need to change the name or description of a tag, it can be done without having to resync all tagged clips. So, for example, if it were decided that a dutch angle should be called a canted angle that could be updated at the tag level and would automatically update wherever tagged. 31

This is a modest solution that notably excludes specialized terms and concepts from more technical areas of film language such as sound, color, or computer-generated imagery. Moreover, relying upon authoritative introductory texts like *The Film Experience* and *A History of Narrative Film* threatens to reproduce their troubling omissions of aspects of film language like 'blackface', which doesn't appear in the glossary of either book despite being a key element of historical film language and narrative in the United States and beyond. Our flat list is admittedly a makeshift substitute for a more robust form of data modeling that could, for example, deepen our understanding of film language and provide further insight into which aspects of it might be analyzable via artificial intelligence, or enable us to share Kinolab data usefully on the Semantic Web. We have, however, anticipated the need for this and built into Kinolab the possibility of adding hierarchy to our evolving controlled vocabulary. For example, tags like 32

- color
- color balance
- color contrast
- color filter

will eventually allow a user to drill down to 33

- color
    - color balance
    - color contrast
    - color filter

Our experience thus far in developing Kinolab has demonstrated that there is a genuine need for development of a film language ontology with critical input from scholars and professionals in film and media studies, information science, computer science, and digital humanities. Beyond the uses described above, this kind of formalized, machine-readable conceptualization of how film language works in narrative media is also a logical information-age extension of the critical work that has already been done on film language and narrative by the figures cited earlier [Eisenstein 1949] [Metz 1974] as well as contemporary scholars such as David Bordwell [Bordwell et al. 1986], among others. 34

6. Fair Use and the Digital Millennium Copyright Act

A robust, well-researched body of literature exists in support of U.S.-based media scholars wishing to exercise their right to assert fair use [Anderson 2012] [Keathley et al. 2019] [Mittell 2010] [Center for Social Media 2008] [Society for Cinema and Media Studies 2008] [College Art Association 2015]. Simultaneously, legal exemptions permitting this kind of work have broadened in the United States over the past two decades. Notwithstanding these developments, aspiring DH practitioners interested in working with moving images may be put off by a complex set of practices and code that necessitates a clear understanding of both the principles of fair use and the DMCA. They may also encounter institutional resistance from university or college copyright officers who reflexively adopt a conservative approach to fair use claims made by faculty and students, especially when those claims relate to the online publication of copyrighted moving images. Kinolab's policy regarding fair use and the DMCA builds upon the assertive stances toward fair use and the DMCA adopted by fellow AVinDH practitioners, especially those of Anderson [Anderson 2012] in the context of Critical Commons and Mittell [Keathley et al. 2019] in the context of videographic criticism. Kinolab's policy also reflects (and benefits from) loosening restrictions authorized by the Librarian of Congress in triennial rounds of exemptions to the DMCA. These have shifted gradually from the outright ban described above to broader exemptions in 2015 for "college and university faculty and students engaged in film studies classes or other courses requiring close analysis of film and media excerpts" [Federal Register 2015, 65949] and, in 2018, for "college and university faculty and students [...] for the purpose of criticism, comment, teaching, or scholarship" [Federal Register 2018, 54018]. The 2018 exemption should be of particular interest to the AVinDH community in that it does away with the earlier rule that capturing moving images (or motion pictures, in the language of the Register of Copyrights) be undertaken only in the context of "film studies classes or other courses requiring close analysis of film and media excerpts," replacing that language with the more expansive "for the purposes of criticism, comment, teaching, or scholarship." 35

The Kinolab team authored a comprehensive statement detailing the project's adherence to the principles of fair use as well as its compliance with the DMCA in order to secure critical institutional support for the project, which was granted after vetting by Bowdoin College's copyright officer and legal counsel (see http://kinolab.org/ for Kinolab's Statement on Fair Use and the Digital Millennium Copyright Act). Essential as this kind of work is, it is time-consuming and somewhat peripheral to the project's main goal. Moreover, our confidence about finding ourselves on solid legal footing is tempered by the knowledge that that footing does not extend outside of the United States, where Kinolab would fall under the jurisdiction of diverse and, in some cases, more restrictive copyright codes. For now, we echo colleagues whose work has paved the way for Kinolab when we observe that the right to make fair use of copyrighted materials is a key tool that will only become more vital as audiovisual work in DH increases, and that members of the AVinDH community should continue to exercise this right assertively. For our part, we make Kinolab's work available under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC), which gives users permission to remix, adapt, and build upon our work as long as their new works acknowledge Kinolab and are non-commercial in nature. 36

7. Conclusion

This case study highlights several of the challenges and opportunities facing DH practitioners who work with audiovisual materials: in particular, the recent shift in digital text analysis (and, to some extent, in moving image analysis) away from annotation as a basis for data set training in favor of newer forms of machine learning; the ongoing need for an authoritative data model for film language; and the changing legal terrain for U.S.-based projects aiming to incorporate AV materials under copyright. The fact that each of these challenges is simultaneously an opportunity underscores just how dynamic AVinDH is in 2021. It also explains why this case study describes a project that is still very much *in medias* 37

res.

As of this writing, the Kinolab team is testing its new platform and seeking user feedback on ways to improve it. We are also taking steps to ensure the thoughtful, intentional growth of Kinolab's clip collection and the project's long-term sustainability. These include, among others, 1) expanding the project's advisory board to include members broadly representative of an array of scholarly interests in film language and narrative, including sound, color, and computer-generated imagery (the use of 3D computer graphics for special effects), but also animated media, national and regional cinemas, horror, ecocinema, science fiction, silent cinema, television, queer cinema, classical Hollywood cinema, transnational cinema, and/or issues related to diversity and inclusion, among others; 2) independently developing and/or contributing to existing efforts to create a robust data model for film language; 3) encouraging colleagues to contribute to Kinolab by supporting the ongoing work of clip curation at their home institutions, either by internally funding undergraduate or graduate student clip curation or through student crowdsourcing in their classrooms; 4) testing and implementing where appropriate machine vision technologies such as those in development at the Media Ecology Project and the Distant Viewing Lab; 5) developing relationships with likeminded groups such as Critical Commons, Domitor, the Media History Digital Library and the Alliance for Networking Visual Culture, among others; and 6) developing national organizational partnerships with the Society for Cinema and Media Studies and/or the University Film and Video Association. Through these and other strategies, we hope to become a genuinely inclusive platform for the analysis of narrative media clips, built from the ground up by the scholars and students using it.

Works Cited

Alm et al. 2005 Alm, Cecilia Ovesdotter, Dan Roth, and Richard Sproat. “Emotions from Text: Machine Learning for Text-Based Emotion Prediction.”In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 579–86. HLT '05. Stroudsburg, PA, USA: Association for Computational Linguistics. (2005).

Anderson 2012 Anderson, Steve. “Fair Use and Media Studies in the Digital Age.”*Frames*, 1 [Online]. Available at: https://framescinemajournal.com/article/fair-use-and-media-studies-in-the-digital-age/ (Accessed 30 November 2019).

Arnold and Tilton 2019 . Arnold, Taylor and Lauren Tilton. “Distant Viewing: Analyzing Large Visual Corpora.”Digital Scholarship in the Humanities. (2019) https://academic.oup.com/dsh/advance-article-abstract/doi/10.1093/digitalsh/fqz013/5382183.

Bazin 2004 Bazin, André. *What is Cinema?* Volume 1. University of California Press, Berkeley (2004). Available at: ProQuest Ebook Central (Accessed 30 November 2019)

Blei et al. 2002 Blei, David M., Andrew Y. Ng, and Michael I. Jordan. “Latent Dirichlet Allocation.”In Advances in Neural Information Processing Systems 14, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani, (2002) 601–8. MIT Press.

Bordwell et al. 1986 Bordwell, David et al. *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. Columbia University Press, New York (1985).

Center for Social Media 2008 Center for Social Media. *Code of Best Practices in Fair Use for Online Video*. Available at: https://cmsimpact.org/code/code-best-practices-fair-use-online-video/ (Accessed: 30 November 2019).

Chandler 2011 . Chandler 2011. *A Dictionary of Media and Communication*. Available at: https://www.oxfordreference.com/view/10.1093/acref/9780199568758.001.0001/acref-9780199568758-e-1041 (Accessed 30 November 2019).

Cho et al. 2014 Cho, Kyunghyun, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches.”(2014) arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1409.1259.

College Art Association 2015 College Art Association. *Code of Best Practices in Fair Use for the Visual Arts*. Available at https://www.collegeart.org/programs/caa-fair-use (Accessed: 30 November 2019).

Columbia Center for Teaching and Learning 2015 Columbia Center for Teaching and Learning. *About the Columbia Film Language Glossary* [Online]. The Columbia Film Language Glossary. Available at: https://filmglossary.ccnmtl.columbia.edu/about/Columbia Center for Teaching and Learning (Accessed: 30 November 2019).

Cook 2016 Cook, David A. *A History of Narrative Film.* W.W. Norton & Company, Inc., New York (2016).

Corrigan and White 2018 Corrigan, Timothy and Patricia White. *The Film Experience.* Fifth edition. Bedford/St. Martin's, Boston (2018).

Critical 2015 *How To* [Online]. Critical Commons. Available at http://www.criticalcommons.org/how-to (Accessed: 30 November 2019).

Devlin et al. 2018 Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding."(2018) arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1810.04805.

Distant 2019 *Analyzing Visual Culture* [Online]. Distant Viewing Lab. Available at https://distantviewing.org/ (Accessed: 30 November 2019).

Eisenstein 1949 Eisenstein, Sergei. *Film Form: Essays in Film Theory.* Harcourt, Inc., San Diego, New York, London (1949).

Estrada et al. 2017 Estrada, Liliana Melgar, Eva Hielscher, Marijn Koolen, Christian Gosvig Olesen, Julia Noordegraaf, and Jaap Blom. "Film Analysis as Annotation: Exploring Current Tools and Their Affordances"In *The Moving Image* 17.2: 40-70 (2017).

Federal Register 2015 . Available at: https://www.federalregister.gov/documents/2015/10/28/2015-27212/exemption-to-prohibition-on-circumvention-of-copyright-protection-systems-for-access-control (Accessed 30 November 2018).

Federal Register 2018 . Available at: https://www.federalregister.gov/documents/2018/10/26/2018-23241/exemption-to-prohibition-on-circumvention-of-copyright-protection-systems-for-access-control (Accessed 30 November 2018)

Fischer and Petro 2012 Fischer, Lucy and Patrice Petro. *Teaching Film*. The Modern Language Association of America, New York (2012): 6.

Halter et al. 2019 Halter, Gaudenz, Rafael Ballester-Ripoli, Barbara Flueckiger, and Renato Pajarola. "VIAN: A Visual Annotation Tool for Film Analysis."In *Computer Graphics Forum*, 38: 119-129. (2019)

Harpring 2013 Harpring, Patricia. *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and other Cultural Works.* Getty Research Institute, Los Angles (2010).

Keathley et al. 2019 Keathley, Christian, Jason Mittell, and Catherine Grant. *The Videographic Essay: Criticism in Sound & Image.* Caboose (2019): 119-127.

Krizhevsky et al. 2012 Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks."In *Advances in Neural Information Processing Systems* 25, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger. (2012) 1097–1105. Curran Associates, Inc.

Kuhn 2012 . *A Dictionary of Film Studies*. Available at: https://www.oxfordreference.com/view/10.1093/acref/9780199587261.001.0001/acref-9780199587261-e-0460 (Accessed 30 November 2019).

Kuhn et al. 2015 Kuhn, Virginia Alan Craig et al. "The VAT: Enhanced Video Analysis."in *Proceedings of the XSEDE 2015 Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure.*, a11, ACM International Conference Proceeding Series, vol. 2015-July, Association for Computing Machinery, 4th Annual Conference on Extreme Science and Engineering Discovery Environment, XSEDE 2015, St. Louis, United States, 7/26/15. https://doi.org/10.1145/2792745.2792756.

Manovich 2012 Manovich, Lev. "How to Compare One Million Images?"In *Understanding Digital Humanities*. Ed. David Berry. Palgrave Macmillan, New York (2012). 249-278.

Media Ecology Project 2019 Media Ecology Project. Available at http://mediaecology.dartmouth.edu/wp/ (Accessed: 30 November 2019).

Metz 1974 Metz, Christian. *Film Language: A Semiotics of the Cinema.* Oxford University Press, New York (1974).

Mikolov et al. 2013 Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed Representations of Words and Phrases and Their Compositionality."In Advances in Neural Information Processing Systems 26, edited by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (2013) 3111–19. Curran Associates, Inc.

Mittell 2010 Mittell, Jason. "Letting Us Rip: Our New Right to Fair Use of DVDs."*The Chronicle of Higher Education*, 27 July [Online]. Available at: https://www.chronicle.com/blogs/profhacker/letting-us-rip-our-new-right-to-fair-use-of-dvds/25797 (Accessed 30 November 2019).

Pustejovsky 2012 Pustejovsky, James, and Amber Stubbs. *Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications*. O'Reilly Media, Inc. (2012)

Pustejovsky et al. 2011 Pustejovsky, James, Jessica L. Moszkowicz, and Marc Verhagen. “ISO-Space: The Annotation of Spatial Information in Language.”In Proceedings of the Sixth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation, 6:1–9. (2011) pdfs.semanticscholar.org.

Sainath et al. 2015 Sainath, Tara N., Brian Kingsbury, George Saon, Hagen Soltau, Abdel-Rahman Mohamed, George Dahl, and Bhuvana Ramabhadran. “Deep Convolutional Neural Networks for Large-Scale Speech Tasks.”Neural Networks: The Official Journal of the International Neural Network Society 64 (April) (2015) 39–48.

Society for Cinema and Media Studies 2008 Society for Cinema and Media Studies. *The Society for Cinema and Media Studies' Statement of Best Practices for Fair Use in Teaching for Film and Media Educators*. Available at: https://cmsimpact.org/code/society-cinema-media-studies-statement-best-practices-fair-use-teaching-film-media-educators/ (Accessed 30 November 2019).

Tissot et al. 2015 Tissot, Hegler, Angus Roberts, Leon Derczynski, Genevieve Gorrell, and Marcus Didonet Del Fabro. “Analysis of Temporal Expressions Annotated in Clinical Notes.”In Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11) (2015) aclweb.org. https://www.aclweb.org/anthology/W15-0211.

Vaswani et al. 2017 Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. “Attention Is All You Need.”(2017) arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1706.03762.

Yang et al. 2018 Yang, Xiao, Craig Macdonald, and Iadh Ounis. “Using Word Embeddings in Twitter Election Classification.”Information Retrieval Journal 21 (2) (2018): 183–207.

Young et al. 2017 Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. “Recent Trends in Deep Learning Based Natural Language Processing.”(2017) arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1708.02709.