GIF Speak: The Development of a Digital Image Collection and Subject Specific Standard Vocabulary

Digital Humanities Project Proposal

Emily Moss

INFO 220

Fall 2017
Abstract

The study of Web 2.0 artifacts and ephemera has become increasingly popular with academicians exploring the use of hashtags, memes, tweets, and the like but the analyses of graphics interchange format files, more commonly known as GIFs (.gif) have thus far been subject only to close readings of their formal, aesthetic, and semantic content. Primarily this is due to the challenge of capturing a GIF corpus and ascribing all the necessary metadata that would allow for valuable distant reading of a GIF corpora. Digital image collections are typically hindered by the laborious process of manual textual metadata entry and while the GIF is not a new technology, its popularity derives from the spreadable quality of Web 2.0 which has only been made possible in the twenty-first century. To that extent, the GIF as a cultural artifact is a relatively new phenomenon and thus, there is not a repository nor standard vocabulary or metadata scheme referenced in discussion and exploration of GIF creation and use which presents further challenges in ascribing textual metadata that captures a GIF’s formal, semantic, and contextual qualities. But because of said qualities, GIF use in online information behavior and virtual exchange is meaningful and to better unpack that meaning, it’s necessary to develop a working database of the digitally native objects to be considered, described, and analyzed. The compilation of metadata generated in application to a GIF corpora can then be subject to the same kind of textual analysis that is currently used in text-driven documents to discern patterns and relationships that exist across a collection of Web-sourced GIFs and to develop a subject specific standard vocabulary that consistently and authentically captures context-driven usage.

Overview
Project Proposal: GIF Speak

Project Outline

The broad aim of this project is two-fold and focuses on

1) developing a digital collection of image files that includes all necessary and meaningful metadata and functions as a database toward analysis of formal, semantic, and contextual attributes such that it’s possible to

2) design a subject specific standard vocabulary that applies to the collection

The specific objectives include:

1) Database assessment and selection

2) GIF ingestion in 4 batches of 25 GIFs each

3) Determining best metadata schema for local application profile and establishing its use

4) Compilation of all textual metadata including standard-driven values and natural language tags for analytic purposes

5) Data visualization through text analysis tools to look for patterns within metadata

6) Weeding/streamlining/identifying best language to describe content based on metadata use

7) Evaluate language against additional corpus

In partnership with the Museum of the Moving Image in Astoria, NY and digital curator, Jason Eppink, an Information Studies associate professor and an American Studies associate professor in the Maryland Institute for Technology in the Humanities (MITH), a graduate program on the University of Maryland campus, are collaborating on a project centered around the development of a digital image collection that functions as a database for web-sourced GIFs with the ultimate objective of designing a subject specific standard vocabulary for
interdisciplinary use when creating, capturing, and studying GIFs. The primary intended audience for the database is an academic one and the project’s output and findings will be presented through scholarly resources and at professional events. However, the database and its content will be public so access is not restricted to members of a designated institution, department, or discipline. Through a local metadata scheme that combines key elements of standard metadata initiatives including Dublin Core and VRA Core while allowing for customized fields as well, comprehensive and detailed metadata ascription will make possible the kind of textual analysis on moving images that has thus far been applied to web-driven language corpora (Zappavigna, 2015). The project seeks to create a database from a pre-selected GIF corpus that includes low-level and high-level metadata attribution. Following file ingestion and record creation, textual metadata analysis will be performed through use of a web-based tool to look for formal and semantic patterns across GIF content such that development of a standard vocabulary to describe the existing GIF collection is possible. Additional ingestion of a new corpus of 25 GIFs will then serve as a test against the preliminary vocabulary to see if its application holds.

The project will take place on the University of Maryland campus over the course of 10 weeks from June 16th to August 25th. The project team will meet and work in the MITH facility which is located on the bottom level of the main library on the University of Maryland’s campus in College Park, MD in Classroom 3 and make use of the resources and equipment available in that space including 8 new desktop computers on which the database software to be assessed in the first two weeks of the project’s activity will installed prior to the first meeting. The project team include the two associate professors who serve as the principal investigators. The project manager is the archival metadata librarian at the university’s main library who has led collection
development initiatives in the past for the library. An information technology consultant has been retained from the Computer Science department who will be in the classroom once a week to oversee software installation, database customization, and general technical troubleshooting. The consultant will receive an hourly wage of $35 dollars/hour and work five hours a week for 12 weeks (as a bookend to the formal duration of the project for preliminary tech setup and output assistance as necessary). The research associates will include four graduate students at the MITH with special interest in Information Studies, Media Studies, and or Digital Humanities. Applications will be accepted for the positions starting February 1st. The graduate students will be unpaid but the project work will be applied to their pursuit of their post-graduate certificate from the Institute. Project meetings will take place on Tuesday, Wednesday, and Thursday from 10 AM to 4 PM with one hour allotted for lunch. The meetings will be overseen by the project manager and at least one principal investigator will always be in attendance.

The project will first focus on tool assessment to determine the best digital asset management system to use moving forward in development of the database. Three DAMS popular in Digital Humanities work and exhibition will be evaluated including Omeka Classic, ContentDM, and Collective Access. This evaluation will include the ingestion of 10 GIFs selected at random from the first batch of 25 to be ingested and recorded in each system. Ease of use as well as interoperability, functionality, metadata capability, and accessibility will be investigated in deciding which DAMS to select for the project. The approved database will then be installed on the three additional computers so that ingestion of the 25 GIFs can proceed. The following process (and is detailed further under Project Tasks) will then be repeated in the subsequent weeks across four stages with growing quantities of metadata until the collection holds 100 separate GIFs.
The GIF: A Brief History

In 1987, CompuServe Information Service released GIF 87 A as a format that was designed to display information that could hold more than one image to eliminate redundant data. Netscape Navigator, in 1995, “took advantage of the Application Extension Block…to enable looping, making the GIF viable for animation online over dial-up speeds” (Eppink, 2014). GIF use in the 1990s primarily occurred on personal webpages and featured rudimentary design (limited color palette, outlined shapes, etc.) but with the explosion of Web 2.0 platforms in the mid aughts including Reddit, Twitter, and Tumblr, the GIF enjoyed a renaissance and became a popular means of virtual communication and artistic or self-expression. Tumblr, in particular, is designed to facilitate the spread of content through its reblogging feature and as such, is a natural extension of the GIF phenomenon which is a file not only almost immediately divorced from its own specific provenance but from the very concept of provenance. While the source text from which the animation was clipped is typically meaningful, the original creator of the GIF file as well as the space in which it was originally published and shared is irrelevant in common GIF use. Similarly, when CompuServe released the format, they did so as an open source specification and encouraged its development and proliferation through networked channels. The GIF is both a digital native and a digital orphan but its status as cipher text allows for myriad
uses in online information behavior. In 2012, at the Library of Congress’s event on Digital Preservation as part of the National Digital Information Infrastructure and Preservation Program, Web expert Anil Dash posited that animated GIFs are the most watched form of video and lamented the lack of any standardized archival practice regarding their publication and use (Weinberger, 2012). Two years later in a blog post on the Library of Congress website, GIFs were proclaimed “a new medium of expression” (Owens, 2014). GIFs, as a means of expression, are best recognized through the reaction GIF whose popularity affirms the significance of GIF use in virtual exchange. Digital curator (and project partner), Jason Eppink, designed an exhibition of 37 GIFs for the Museum of the Moving Image in New York in 2014 entitled The Reaction GIF: Moving Image as Gesture. Eppink characterizes a reaction GIF as the following “[it’s] an animated GIF, typically of a body in motion and primarily excerpted from a film or television show, that is used online as a response or reaction” (Owens, 2014). While there is meaning to be mined from this communication method, currently it’s not possible to access a GIF database that captures not only contextual metadata but descriptive, semantic, and technical metadata as well. Toward better understanding GIF use, the aim of this project is to preliminarily develop such a resource. The following queries inform the subsequent work:

How do you capture context through metadata in a way that is consistent and authentic? How do you reconcile origin metadata for objects that are divorced from their provenance? How do you demonstrate patterns or similarities in GIF content without stripping a GIF of any of its potential meaning when creating a record for it? How do you document relationships across a GIF corpus (database, message board, etc.) using language that is consistent to all GIF content in all GIF contexts?
Primary Source Material: Graphics Interchange Format files or GIFs (.gif); a digitally native moving image of looped, animated content typically sourced from a film or television program and used in a variety of application Web 2.0 platforms

Tools: Digital Asset Management System intended for work in Digital Humanities (TBD), Google docs, university server to host DAMS software, 2017 desktop computers provided through the institution’s work space, web-based text analysis tool Voyant, Wordpress for research associate blogs, personal email correspondence through Gmail accounts

Output: The final digital collection of 100 GIFs, research associate blogs detailing activities, findings, and reflections throughout the project, a narrative report outlining the project and its findings co-authored by the principal investigators, a daily log of all activity maintained by one research associate for clear documentation of the project’s progress

(Subsequent future output may include conference presentations and additional articles.)

Institutional Partners: University of Maryland, Maryland Institute for Technology in the Humanities, Museum of the Moving Image

Environmental Scan

Because of the time and labor challenges inherent to generating a corpus of digital artifacts with ascribed metadata, a GIF database and related taxonomy has not been established toward any great academic, cultural, or commercial use. Most web-based GIF repositories are annotated by users and thus, there is little consistency nor comprehensibility in content description. However, there are bodies of previous analysis that serve as reasonable antecedents to such a project including image archiving, hashtag analysis, and current GIF use in media studies and digital humanities.
Tagging as a means of ascribing metadata to digital content is another phenomenon that rose in prevalence with the advent of Web 2.0. Social tagging or collaborative tagging on websites like Digg and Delicious was first employed as a means of information search and retrieval. However, as tags are derived from natural language, manipulation of their structure and significance evolved over time. The hashtag as a characteristic of microblogging platform Twitter was first used on the platform in 2007 as an annotation to a real world event (fires in San Diego, California). Subsequently, as more users embraced the metadata tool, the hashtag maintained its status as an information retrieval system but additionally provided context, asserted social identification, and served a linguistic function in twenty-first century dialect. As such, scholarly work of the hashtag and its methods of employment grew in prevalence as well with the technological capability to capture and analyze a corpus of tweets inclusive of embedded hashtag use. As antecedent to the study of GIFs, I’m primarily interested in the work on hashtags as a means of self-identification and communication in virtual exchange (Owens, 2014). While I posit that GIF use and hashtag use in this context are related, there are significant differences between the objects that warrant the further study of GIF use.

Brenda Dervin’s conception of sense-making relies on a phenomenological approach to information behavior. Dervin’s theory derives from the phenomenon of discontinuity as always already present and as such, information behavior is ordered through “gap-defining” (as acknowledgement of the discontinuity) and “gap-bridging”. These attempts are recognized as “sense-made” or relational knowledge and inform how the searcher moves forward toward infinite defining and bridging (Dervin, 2003). Given that, bodily experiences are given primacy in sense-making and gap-bridging and the visual of a body or face expressing emotion or engaging in a specific behavior is more phenomenologically meaningful than text describing
those activities. In his work on celebrity fandom, Redmond frames the relationship the fan has to the body of a celebrity (as informed by her own experience as a body) within a phenomenological discourse (Redmond, 2016) and that reading can be extrapolated to include subjects who are made “famous” by their inclusion in popular GIFs. The Reaction GIF then as well as GIFs that are used in the context of expressing a feeling (or many feelings) to a real or imagined audience (Scott, 2015) on Web platforms resonate as a form of gap-bridging and sense-making in virtual exchange in a way that is uniquely distinguished from the way that hashtags are used toward linguistic ends in virtual exchange.

Additionally, the origin of the web-sourced GIF is providing additional meaning and situates GIF analysis within the discipline of digital humanities as the cultural artifact retains many of the formal artistic qualities of its source text. The act of “poaching” (Jenkins, 1992) that occurs in GIF creation and dissemination and in a poached culture such as Web 2.0, Jenkins via Bakhtin maintains that a text is absent inherent meaning; meaning is ascribed through consideration of the text’s relationship to other texts. The GIF as a looped, soundless excerpt from a formal filmic work is both data and metadata; it paradoxically is defined as a text and a context and thus, its use as a communication tool in web-based dialogues is laden with meaning that justifies further exploration of semantic and formal attributes therein contained.

The dominant mode of GIF use in digital humanities at this time includes the scholarly creation of GIFs to better understand formal and artistic elements of production. Digital storytelling and computer science instructor at the University of Mary Washington, Jim Groom has developed what he refers to as a GIF curriculum or GIFculum in collaboration with Chinese History professor, Sue Fernsebner, for her Chinese History through Film course. “The idea was to experiment with GIFs to capture a moment (or several moments) in cinema in order to replay
it again and again to get a sense of how it is working both technically and thematically” (Groom, 2013). As a result of this initial exploration, Fernseber revised her course to incorporate student GIF work and made it central to their analysis of film in the final paper. In this context however, emphasis is on GIF creation for the purposes of analysis rather than analysis of web-sourced GIF content. That exploration has been conducted as well (Gursimsek, 2016; McDonald; 2013) but thus far relied on close reading exegesis of an extremely limited number of GIFs (totaling no more than five) because of the challenge in surveying a GIF corpus in any meaningful way. While the web platform, Tumblr, allows one to search GIFs by subject, a visual scan of the results resonates only the most superficial commonalities, differences, and patterns across the collective GIF content. To derive real, quantitative knowledge on GIF use necessitates analysis of not only the image but its textual metadata which requires the development of a digital collection or GIF database.

The primary challenge however in developing a collection that allows for the kind of analysis necessary toward understanding GIF behavior is akin to that which is posed in the development of any digital image collection. The process employed to organize, search, and retrieve images through digital archives maintains primary influence on the process depicted within this proposal as one that continues to rely on textual metadata for access (Fear, 2010). Metadata is, in many ways, what determines the value of a digital collection particularly when one is being amassed to further explore the context of object use. Effective metadata ascription is consistently inhibited by two factors of equal significance – 1) manual textual metadata entry is incredibly laborious and 2) often suffers from inconsistent attribution language. Toward reconciling these two key issues, some institutions have come to rely on social or voluntary metadata ascription (Fear, 2010; Mayer, 2013). Further some archives have adopted the concept
of “More Product, Less Process” which includes capturing as many objects as possible and providing metadata for the collection as a whole or segments of the collection while eschewing detailed object metadata entry (Greene & Meissner, 2005). However, this proposed solution generates additional problems in that user-generated metadata is problematic in consistency, authenticity, and context. Therefore, this project proposal acknowledges the time, effort, and expertise necessary to generate an object record that reflects detailed and comprehensive administrative, descriptive, and structural metadata and foregrounds that process in the project’s work plan by pre-selecting the collection objects and making use of the services of advanced graduate students in the fields of digital humanities, library and information science, and media studies as well as the expert knowledge of the institution’s metadata librarian.

**Project Tasks**

**Stage 1** includes testing three different content management systems through ingestion of the same 10 reaction GIFs sourced from the museum’s exhibition of 37 total and selected at random. The three systems under assessment include Omeka, ContentDM, and Collective Access and their advantages and disadvantages are enumerated in Appendix A.

Two of the three systems operate through open source web-based software installation and all requisite system requirements have already been installed on the work computers (seven in all). That installation was performed by the IT consultant prior to the project’s commencement. Each system works with Dublin Core and VRA Core metadata schemes so for this initial assessment, each will be employed (with redundant elements omitted) to determine which scheme or combination of schemes is best to capture all descriptive, administrative, and structural metadata of each object. The research associates will present their findings on the
systems and then one system will be selected to move forward with the project. At this point, the metadata scheme will also be determined for initial ingestion and customized as necessary through the implementation of a local installation profile.

**Stage 2** includes initial GIF ingestion of 25 GIFs sourced from the museum’s exhibition of 37 total GIFs. GIFs are typically sourced from the Web but to work with an existing corpus that has already undergone one level of vetting for inclusion in an exhibit about the GIF, this ingestion will include 25 reaction GIFs captured as GIF files from the Museum of the Moving Image website. Each of these GIFs is a digital object of 3-5 second looped animated content. The author and original publisher of the GIF file is typically unknown but the object records will include information on the provenance of the source text from which the GIF derived. While GIFs can include any looped animated content (including original and remixed artwork or media), these files are only sourced from film or television programming (inclusive of commercials, music videos, television series and specials, and feature films). Significantly, in this initial GIF collection, works displaying social actors inclusive of interviews, live events, and reality programming are excluded from consideration and analysis. Object records reflect specified metadata standards and natural language tagging annotation.

**Stage 3** includes capturing all the textual metadata that has been entered for each of the 25 GIFs and compiling it in two separate Google docs including one document containing all the natural language tags and a separate document containing all the metadata field attributes. A text analysis tool, voyant-tools.org, will be employed to provide visualization and statistical information on the two bodies of metadata through a simple copy and paste function via a textbox on the website’s homepage such that observations can be made regarding most frequent or popular terms, term relationships (i.e. man to woman ratio), and patterns established through
the terms. (Appendix B) This analysis will lay the groundwork for revisiting the corpus with the goal of accounting for all the previous metadata through a working standardized application – that is, can information gleaned from natural language tags be effectively and consistently incorporated into a metadata scheme and additionally, does the data bear out the possibility of a controlled vocabulary without loss of semantic, formal, or structural attribution? This question doesn’t have to be answered at the conclusion of Stage 3 but it informs the work moving forward.

**Stage 4** includes the repetition of Stage 2 with a new batch of 25 GIFs sourced from Eppink. The initial corpus was pulled from Eppink’s curated exhibit and subsequent GIF corpora in batches of 25 separate objects will additionally be sourced from Eppink to reduce influence or manipulation of the results. Eppink is not aware of what the initial metadata revealed and therefore his selection is free of confirmation bias while the project team members are unable to use their knowledge of the initial metadata to influence the selection of new GIFs for related analysis. The criteria for the second GIF batch repeat the stipulation that the content is sourced from film or television and spans three to five seconds. As with the first batch, records will be created for each of these new objects that reflect all available administrative, descriptive, and structural metadata in adherence to the collection’s established metadata scheme. Following the second ingestion of GIFs, the new metadata and tags will be similarly compiled in the pre-existing Google documents and analyzed through Voyant via the following comparisons:

1) Batch 1 Metadata WITH Batch 2 Metadata
2) Batch 1 Tags WITH Batch 2 Tags
3) Batch 2 Metadata WITH Batch 2 Tags
4) Batch 1 Metadata WITH Batch 2 Tags
5) Batch 2 Metadata WITH Batch 1 Tags

Documentation of findings will occur through two methods – a shared document that includes Voyant’s data visualization models and narrative annotation to be maintained by a designated secretary among the team project members. Additionally, the four graduate students will be required to keep individual blogs through Wordpress that capture their reflections, predictions, and other observations following each stage of metadata comparison and analysis. Certain questions posed at this stage may include- is it possible to account for the title of a source text in a standardized vocabulary? Is it worthwhile to design or abbreviate the language of common or repeated values e.g. ‘white older male’ as ‘WOM’ or ‘subject: WOM’ or ‘2WOM’? Do such designations hold up against a new corpus of GIFs? That is, is there a continuity across GIF content that allows for a common language to be referenced when ascribing metadata?

Stage 5 includes the ingestion of a third batch of 25 separate GIFs sourced from Eppink and the requisite metadata entry, capture, comparison, and analysis that’s proceeded the previous two batch ingestions.

Stage 6 repeats Stage 5 with the final batch of 25 separate GIFs sourced from Eppink and the requisite metadata entry, capture, comparison, and analysis. Following this stage, the database contains 100 records for 100 separate GIF files.

Upon compilation and analysis of all the metadata contained therein, observations on the language of GIF metadata are made, discussed, and documented toward a subject specific standard vocabulary that provides coverage for the collection with the potential of providing information on GIF creation, publication, dissemination, and utilization in a Web 2.0 environment. This concludes this preliminary project of creating a GIF database through the development of a digital image collection and its attendant standardized vocabulary. The
findings will determine if the work should continue at which point, the team, funding, resources, and scope will be assessed and likely, expanded.

Participants

- Female Information Studies Associate Professor at UMD – Co-principal Investigator; as an Information Studies professor, she is primarily concerned with the GIFs as they relate to information behavior and will drive discussion of how GIF context is captured through metadata as well as advise on how to improve upon the process.

- American Studies Associate Professor at UMD – Co-principal Investigator; as an American Studies professor and the Director of the Design Culture + Creativity Program, he is primarily interested in the formal qualities of GIFs and if there are formal commonalities in popular GIF use. He will supervise and inform how descriptive and technical metadata is generated and ascribed.

- Archival Metadata Librarian at UMD – Project Manager; as a metadata librarian with previous project management experience developing digital collections, she is well-suited to advise on metadata schemes and record creation and will lead each stage of the project.

- Digital Curator at Museum of the Moving Image – Consultant and Contributor; as a “GIF expert” having curated several exhibitions for the museum, he will serve as the primary source for providing GIF batches for ingestion as well as be available to consult on content and context queries related to GIF record creation and provenance.

- Advanced Graduate Student at MITH – Research Associate

- Advanced Graduate Student at MITH – Research Associate

- Advanced Graduate Student at MITH – Research Associate
• Advanced Graduate Student at MITH – Research Associate/Project Secretary
  The research associates will handle the tool assessment, GIF ingestion, record creation, metadata capture and contribute to metadata analysis through in person meetings and on required individual blogs where they will document their predictions, findings, and reflections once a week for the duration of the project. Three RAs will evaluate potential databases in the first stage of the project while the fourth will serve as the project secretary and be responsible for maintaining the daily log of activity for formal documentation.

• IT Consultant at UMD – Technical Support; sourced from the university’s Computer Science department, they are a web developer who will be available for software installation, installation profile design, and general technical support at an hourly wage.

Budget

Given the exploratory nature of this project, the principal investigators sought to severely limit the financial costs associated with its execution. The primary expenses are those of time and
labor which are covered by the research associates working for course credit rather than monetary compensation. The space and equipment are provided via the institution’s facility and the one explicit confirmed cost is the salary of the IT Consultant at $35/hour for 5 hrs./week for 12 weeks.

<table>
<thead>
<tr>
<th><strong>Equipment fees: CDM test ingestion</strong></th>
<th>One time cost of $75 dollars</th>
</tr>
</thead>
<tbody>
<tr>
<td>To upload &amp; create records for 10 GIFs</td>
<td>If CDM is selected, budget revision will be required</td>
</tr>
<tr>
<td><strong>Consultant fees: IT Salary</strong></td>
<td>$35/hour for 5 hrs./week for 12 weeks = $2,100.00</td>
</tr>
<tr>
<td><strong>Incidental fees: unexpected expenses</strong></td>
<td>&gt; $750.00</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>$2,925.00</strong></td>
</tr>
</tbody>
</table>

References


Appendix A

**Omeka (Classic)** is a free open source content management system developed at the Roy Rosenzweig Center for History and New Media at George Mason university that is specifically designed for digital humanities use. It has interoperability because of its profile within the
discipline and utilization of standard organization and metadata schemes. It’s also possible to specify how to use elements within the database for local collection. The platform is easily accessible and doesn’t require any prior coding knowledge to populate object records following installation of the software. The system allows for customized functionality through plug-ins and as an open-source platform, encourages collaboration and development through site forums and Github. Semantic HTML5 markup is available for front end and back end design. Plug-ins currently available include those for VRACore metadata scheme, controlled or standard vocabulary design, and tagging (though the tagging schema is problematic).

The system requirements for installation are as follows: Linux operating system, Apache HTTP server, MySQL version 5.0 or greater, PHP scripting language version 5.2.11 or greater (PHP version 5.3+ is highly recommended), ImageMagick image manipulation software

Database back up is possible through PHPMyAdmin.

**Content DM** is a proprietary database software available through OCLC. The database is scalable for collection size and metadata is stored in XML. The system is compliant with Dublin Core and VRA Core metadata standards and allows for development of custom controlled or standard vocabularies. Workflow is flexible because it’s web-based. The maximum number of metadata fields a user can create for each collection is 125. It’s a DAMS that’s been adopted by several libraries, academic institutions, and museums because of its functionality.

The primary detraction to using ContentDM is that there is a cost associated and because it’s proprietary, customization is limited and relies on the system IT staff. A price quote is only available through direct inquiry and dependent on the size and scope of collection. The limited information on pricing available via the Web suggests that a base fee may be $750/year (or was
when the system launched in 2007) but can go as high as $40,000 for initial ingestion and $10,000/year for maintenance. CDM is ‘software as a system’ or SaaS.

**Collective Access** is a free open source content management system accessible through software installation. The system allows for alternate names fields, it’s developed in a collaborative environment, boasts easy installation, and is used by museums and archives as well as by universities and for DH projects. As a platform, CA is customizable and the design emphasis is on display and interoperability. The system works with Dublin Core and VRA Core metadata standards and allows the use of standard vocabularies and thesauri. There is comprehensive wiki documentation and user forums that are regularly updated. The system requirements for installation are as follows: Linux operating system, Apache HTTP server, MySQL version 5.0 or greater, PHP scripting language version 5.2.11 or greater (PHP version 5.3+ is highly recommended), ImageMagick image manipulation software.

The two main components of CollectiveAccess are Providence, a "back-end" cataloging interface, and Pawtucket, a "front-end" publication and discovery platform. Installation instructions for Providence are [available here](#). Instructions for Pawtucket are [here](#). Pawtucket draws from the Providence database to display data in a read-only website.

Appendix B

Sample GIF ingestion, metadata ascription, tag compilation and analysis using Omeka.net

10 GIFs pulled from The Reaction GIF exhibit curated by Eppink and ingested include:

1. 50 Cent Drive By
2. George Michael AD Fall
3. Crying Dawson
4. David Tennant Rain
5. Liz Lemon High Five
6. Lucille AD Excited
7. Carrey Liar Liar
8. Jeremiah Johnson Nod
9. Troy Abed Good Point
10. Anchorman Don’t Believe You

Browse Items Page - Available to view at http://giflook.omeka.net/

Tags: 2012, black man (2), comedy (6), Community, complex (2), Dan Harmon, Danny Pudi, Donald Glover, drink (2), duo, gesture, Indian man, men of color, NBC (2), parody (5), reaction (10), sitcom (4), sitting (3), 2.0 (3), 2004 (3), Arrested Development (2), excited, Fox (2), funny (6), Jessica Walter, Lucille Bluth, out of character, repeat, senior, spread (7), white woman (2), 2008 (2), 30 Rock, high five, Liz Lemon, Tina Fey, 1997, angry, disgusted, film (3), impressed,
Jim Carrey, Liar Liar, raised eyebrows, smiling (3), white man (6), 1970s (2), 1972, beard, Jeremiah Johnson, nod, Robert Redford, western, wilderness, zoom, Anchorman, Apatow, caption, cigarette, credits, Will Ferrell, 2000, crying, Dawson’s Creek, face (2), James Van Der Beek, sad (3), season finale (2), WB, falling, Fox, Michael Cera, tired, BBC, cinemagraph, Doctor Who, rain, 2009, 50 Cent, car, driving, Entourage, HBO, rapper, smirking

Data Visualization through Voyant