“These images are never transparent windows onto the world. They interpret the world; they display it in very particular ways; they represent it.” – Gillian Rose

TITLE: Digitally Born Folklore and Internet Archives: Where the Memes Have No Name

APPENDICES:

• Appendix A: Examples of Meme Genres

• Appendix B: Examples of Web archival issues discovered through the LC’s American Folklife Center.

INTRODUCTION

What is digital folklore? How does folklore present itself in the digital age? What are examples of digitally born folklore? I started this paper with these presumably simple questions and discovered complex, indefinite answers. The focus of this paper centers on the premise of Internet memes as examples of digitally born folklore, due to their commonplace abundance in digital worlds, their proliferate mode of production and transmission that exceeds the expectations of viral entities, and their clear structural characteristics that can be defined as genres and employed as taxonomic means for
further interpretive research. Memes are digitally born artifacts that are produced, transmitted, and housed within Internet Websites and the pages of Social Networking Sites. The nature of Websites and Social Networking Site pages are temporary. Content can remain indefinitely, content can be edited, and entire websites can be removed permanently, destroying all historical record of its contents. Due to the ephemeral nature of websites, it is important to capture and archive content to preserve the record of cultural exchange and values for historical documentation and to perform research. In this paper, I will draw parallels to folkloric methods of analysis to the observable genres of memes to demonstrate the potential of research for memes. I will also provide an overview of what memes are, the current work that defines meme genres, the difference between viral digital objects and memes, and the cultural exchange and participation in the creation and transmission of memes. Finally, I will review current issues with web-archiving and web-scraping methods through a demonstration of the archival issues I discovered through the American Folklife Center’s archive of memes, and the issues this presents for limiting adequate research.

THEORETICAL FOCUS

Memes are digital cultural artifacts and can be archived and studied as examples of digital folklore. In order to begin to structure a review of memes as valuable digital artifacts, I will refer to the study of traditional folklore within the context of fairytales and their digitization. In Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database, Meder et al provide an overview of the Dutch Folktale Database. Meder et al present the algorithmic process of identifying folkloric motifs in
digitized folktales, stories, etc. (Meder et al, 2016) The algorithms were designed through the compilation of historical folklore research methods—there is a structure to folktales that can be identified, categorized, and analyzed. What is the structure of memes that can be identified, categorized, and analyzed? While Meder et al are exploring digitized folklore (read: folklore that was not digitally born), T.R. Tangherlini explores folklore that is digitally born in *Big Folklore: A Special Issue on Computational Folkloristics*. Tangherlini aids in establishing that folklore can be digitally born and expressed through the Internet. (Tangherlini, 2016) Furthermore, Tangherlini expresses the urgency for archiving these cultural artifacts. (Tangherlini, 2016) Trevor Blank echoes Tangherlini in *Public Folklore in Cyberspace*, and presents the Internet as the new media for print technology, and suggests its role in the propagation of digital folklore.

The analysis of visual images can be complex. Gillian Rose offers a series of methods for analyzing images in her book *Visual Methodologies* (Rose, 2016), and Martin Hand contributes to this conversation both through referencing Rose and discussing the approaches in *Visuality in Social Media: Researching Images, Circulations and Practices*. (Hand, 2017) Rose and Hand establish key characteristics of visual images and what ought to be present to conduct thorough analysis. The image attributes that should be analyzed are more than just the image—it is important to consider production, transmission, the intentional presentation for specific audiences, and the contextualization (where was this image viewed?) to complete a thorough analysis. (Rose, 2016) (Hand, 2017)

Limor Shifman has conducted a substantial amount of research on the topic of memes in *Memes in Digital Culture* (Shifman, 2014). Shifman presents a revised
definition of the term meme from Dawkins’ original definition, which I will present later. (Shifman, 2014) She also establishes meme genres and the difference between viral images and/or videos and memes. (Shifman, 2014) Finally, Shifman’s research suggests a model for the structure of memes and what needs to be captured during web archival processes.

Web archives are captured through the creation of web crawling software and Application Programming Interfaces (APIs). Michael Black reviews the differences between web scraping and web archiving in *The World Wide Web as Complex Data Set: Expanding the Digital Humanities into the Twentieth Century and Beyond through Internet Research*. (Black, 2016) Black’s research, in conjunction with Martin Hand’s in *Making Digital Cultures: Access, Interactivity, and Authenticity* indicate the basic functions of Web archiving software and key questions that need to be considered prior to developing the software and programs to ensure all data are completely captured during the process. (Black, 2016) (Hand, 2008) The issues that are tied to Web archival procedures require capturing digitally born content as it is created to ensure that it is archived before it is modified or permanently removed. In other words, digitally born content must be pre-evaluated as culturally significant, and a thorough collection plan must be devised to capture it.
In June 2017, the Library of Congress (LC) announced that it added new born digital collections to the American Folklife Center: “the Web Cultures Web Archive (WCWA), which will feature memes, GIFs, and image macros that surface in online pop culture, and the Webcomics Web Archive (WWA), which will collect comics created for an online audience.” (Peet, p. 14) The addition of memes to the American Folklife Center indicates the potential of cultural and folkloric value within Internet memes. The creation of this archive presents the potential for researchers to explore and analyze the messages of memes, their cultural significance, and the possibility of representing the values of the communities that produced and transmitted them.

According to the LC, they utilize DigiBoard to perform web-archival processes. (“Web Archiving,” 2018) This tool allows the LC to create seed lists (the list of websites to archive), and it manages the permissions required for the archival process. (“Web Archiving,” 2018) The LC uses a tool called Heritrix to perform the web-crawling duties to collect the data for the archive, stores the data in the BagIt Library, and utilizes a location installation of the Wayback Machine to allow the archive to be viewed or replayed. (“Web Archiving,” 2018) The data that the LC has included in the American Folklife Center archive are viewable as a series of websites that are specifically related to the production and/or transmission of memes: KnowYourMeme.com, YTMND (You’re the Man Now Dog!), Meme Generator, and ¡Cuánto cabrón! (“Search Results from Web Cultures,” 2018) Users can view the captures via the American Folklife Center. The captures were recorded from August 2010 to August 2016.
The LC included memes in the American Folklife Center archives, because it has identified them as cultural artifacts. I was not able to locate any indication as to what precisely the LC has distinguished as the cultural value of any meme, but it is not necessarily the role of the archivist to indicate the cultural value so much as it is to preserve a collection of artifacts that can be evaluated and interpreted by researchers to present their findings and meanings. For the scope of this paper, I am presenting the LC’s announcement and preservation of memes as a positive indicator of a motivation, if not an obligation, to explore memes closer as viable representatives of digital folklore, and to examine the results of their archival methods to identify issues that present a potential hindrance to future research.

METHODS OF DATA COLLECTION

I constructed this paper through a review of a collection of secondary sources—scholarly articles, popular publications, and information obtained through the Library of Congress website. The subjects of the scholarly articles that I reviewed ranged from traditional folklore research through the utility of digital tools such as databases and algorithms designed for the purpose of identifying folkloric motifs, to visual methodologies in digital landscapes, to the production and transmission of memes and identifying meme genres, to reviews of web archival practices and issues. I explored the topics of these scholarly articles in order to discover the connections of traditional folkloric methods with post-modern, digital folklore concepts. This required me to have a general understanding of some of the basic and modern methods of analysis of folklore—this process aided in discovering the need for defining structure of digital cultural
McGowan – Digitally Born Folklore and Internet Archives: Where the Memes Have No Name

artifacts that may not exist as physical world objects. Next, I explored methods of visual analysis, particularly within digital worlds. Visual images are not simply entities or artifacts that can be observed and interpreted individually—there is value and meaning that can be found in the production, contextualization, transmission, and the intended audience that ascribe meaning to the image itself that must be considered while interpreting meaning. (Rose 2016, Shifman 2014, Hand 2017) After exploring visual methodologies, I explored what memes were. What are the current discussions of memes as cultural modes of communication and possible contributions to folkloric studies? Finally, I reviewed the basic concepts of web archival practices—this is a process that is constantly in flux as technology and web programming capabilities evolve and digitally born cultural content is continuously created, added to, and removed from the Internet.

I did not conduct any primary research for this paper, so I utilized the meme archives through the LC’s American Folklife Center to review as a case for exploring the methods used for capturing the web archive and my identification of issues the archive presents. The archived websites that feature memes and were captured by the LC have missing images and missing web page levels or depths (these missing web page levels render the search functionality of the archived web page inaccessible or incapable of navigating). Memes are digitally born, visual folkloric artifacts; if they are not present for analysis, they cannot be interpreted for meaning. Furthermore, I will discuss later that a successful meme requires the presence of iterations, copies, and remixes—a meme is a collection of images. If the web archive of a collection of images is missing, any images or missing key navigational properties to explore the collections and relationships of the
collections, then the archive is incomplete. Therefore, it is insufficient to properly execute research initiatives.

The secondary sources and the case of the LC web archive aid in identifying some key characteristics of memes that can aid in performing research to determine the cultural and folkloric value of memes, but this is a large topic, and more in-depth analysis would need to be performed to contribute to this discussion further. In order to expand on this topic further, I would explore two branches in greater detail. First, I would identify and analyze successful examples of web archives beyond the scope of the LC (or even the Wayback Machine) to determine other best methods that exist in the field of web archives and the challenges they are presented with. Second, I would explore the various genres of memes to identify cultural messages and/or values and other themes they present. Are there types of messages that are characteristic of a specific genre? Are there themes that transcend genre? What sub-sections of communities are over-represented or under-represented that parallel these messages? How can web archival practices aid in collecting supplemental metadata that would reveal some of the meaning of messages and the creators that produce memes?

ANALYSIS AND DISCUSSION

I begin this discussion with a review of traditional folklore (read: fairytales) in order to determine archival methods and issues of memes as digital folklore. This review of traditional folklore is centered in how memes can be analyzed in order to determine if they match the criteria to be considered folklore. Folklore, such as fairytales, has motifs that have been observed, identified, and established through historical research. The
stories can adhere to a specific syntax, contain certain types of characters, follow particular story and plotline arcs, and conclude with specific messages. (Meder et al, 2016) These elements describe a structure that can be observed and analyzed, and that information can be translated into computational algorithms that can be employed to conduct rapid analysis of digitized folklore. “These motif sequences serve two purposes: first, they identify a story as a version of a certain tale type, and second, they help reveal how stories of the same type differ from each other.” (Meder et al, p. 88) In other words, identifying structural modalities of folkloric stories allows for analysis of textual, oral, or digitized folklore.

Folklore, at its core, presents tradition, beliefs, and customs communicated through story, song, or dance. The analysis of the structure of these forms of cultural communications aids in grouping the expressions by motifs that can be analyzed further.

“It is known that, in oral tradition, stories and songs vary from performance to performance on many dimensions (Rubin 1997), both in style, structure, and content; in fact, every new representation of a story will generate new variations (Bartlett 1932; Owens and Bower 1979). Although narrative plots and melodies vary, most of the time, they remain recognizable.” (Meder et al, p. 88) The presence of structure does not indicate or presume that any artifact is folkloric, but it does allow for an organized method of analysis to determine if it is a cultural expression and its value. If traditional forms of folklore, cultural expressions documented through text or shared orally, share structural modalities that can be analyzed through human and computational algorithms, how can similar methods be identified and applied to digitally born folklore?
Tangherlini establishes that folklore can be digitally born, and that methods of analysis are needed. The presence of viral images and videos as digital objects requires methods of collection and analysis.

“In earlier work, I outlined four main areas that will, in the coming years, require significant attention as the materials of study are increasingly represented as digital objects, whether the materials be oral performances or aspects of material culture, or any of the other numerous types of expression that folklorists work with (Tangherlini 2013c). Broadly speaking, these four areas are (1) collecting and archiving, (2) indexing and classifying, (3) visualization and navigation, and (4) analysis (Tangherlini 2013c:8).” (Tangherlini, p. 6)

Folkloric expressions exist within digital worlds, so, as Tangherlini notes above, it is important to establish processes to collect and archive in order to analyze. While historical examples of folklore may have been transmitted and archived through written or spoken word, digitally born folkloric artifacts may include visual components that contribute to the meaning. Furthermore, the nature of content that is produced and transmitted across the Internet includes the presence of visual objects. “The underlying theoretical premise in the application of network theory to folklore is that folklore can be conceptualized as the flow of cultural expressive forms in and across social networks.” (Tangherlini, p. 8)

The analysis of digitally born folklore must include the capture of artifacts that are exchanged through the Internet, where humans are interacting with each other in digital worlds.

Before we can explore what memes are and define structural elements that will aid in archival practices and potential research analysis, it is important to establish essential components of visual methods for research purposes. Gillian Rose establishes the modalities of interpretation for visual images in *Visual Methodologies*. (Rose, 2016)
There are four major modalities that must be taken into consideration when performing analysis: site of image itself, site of production, site of circulation, and site of audiencing. (Rose, 2016) These four elements can be applied to defining the structure of visual objects for analysis. As indicated in the previous section, traditional folklore that is expressed through text or oral tradition has specific syntax, motifs, and other identifiable components that define an observable structure that can be further categorized and analyzed. When Rose’s four modalities for interpreting images are applied to digital images, structure can be observed and defined, to create categories, and to perform analysis.

“If we think qualitatively about specific images, observable streams of images, particular contexts of visual social media use and engagement, or the meaningful activities of producing, consuming and distributing images, then we tend toward the recalibration of established interpretive methods in the social sciences and humanities. This might be content or discourse analyses, social semiotics, surveys, interviews, participant observation, and so on.” (Hand, p. 227)

In the next section, I will explore the definition of memes and common structures that can be observed to create categories for further analysis. It is beyond the scope of this paper to apply Rose’s four modalities to examples of memes to perform interpretation, but I will touch on the image itself, site of production, and site of circulation and how they relate to the essence of a meme in order to determine two key archival processes.

What are memes? The definition of memes starts with a discussion of Richard Dawkins, but extends with a modification by Limor Shifman. The core of the definition is a cultural expression, copied, modified, and shared virally. Dawkins defines memes as singular expressions, but Shifman describes them as collections of expressions.
“My definition departs from Dawkins’ conception in at least one fundamental way: instead of depicting the meme as a single cultural unit that has propagated well, I treat memes as groups of content units. The shift from a singular to a plural account of memes derives from the new ways in which they are experienced in the digital age. If, in the past, individuals were exposed to one meme version at a given time (for instance, heard one version of a joke at a party), nowadays it takes only a couple of mouse clicks to see hundreds of versions of any meme imaginable. Thus, memes are now present in the public sphere not as sporadic entities, but as enormous groups of texts and images.” (Shifman, p. 341)

Dawkins and Shifman purport that memes are examples of cultural expressions. And, if they are examples of cultural expressions then there is value in studying them further to determine the types of messages that are conveyed and if there are themes that are shared across platforms and various structures.

“While memes are seemingly trivial and mundane artifacts, they actually reflect deep social and cultural structures. In many sense, Internet memes can be treated as (post)modern folklore, in which shared norms and values are constructed through cultural artifacts such as Photoshopped images or urban legends.” (Shifman, p. 15)

In order to debate whether or not memes are cultural expressions, and that these expressions are representative of folklore, it is important to further describe the structure of a meme. The structure of memes is essential to define meme genres, and identify what data is required for collection during Web archiving captures to provide sufficient data for further analysis through a folkloric lens. Shifman provides an extensive primer of memes in her book *Memes in Digital Culture*. The key information I will extract from this primer relates to Shifman’s proposal of meme genres and the difference between viral images and memes. These two elements assist in defining Web archival needs: what
needs to be captured? I will start first with the review of viral images as compared to memes. Viral images are singular entities that are shared many times.

“The main difference between Internet memes and virals thus relates to variability: whereas the viral comprises a single cultural unit (such as a video, photo, or joke) that propagates in many copies, an Internet meme is always a collection of texts.” (Shifman, p. 56)

Memes are, as Shifman describes, a collection of texts. This means that a meme is not a singular image, but a series of images. In order for a meme to be successful or complete its existence or express its essence, it must exist among many iterations, variations, and copies of itself.

“Combining these two principles, I define an Internet meme as:

(a) a group of digital items sharing common characteristics of content, form, and/or stance, which (b) were created with awareness of each other, and (c) were circulated, imitated, and/or transformed via the Internet by many users.” (Shifman, p. 41)

Here we have our first principle for the proper development of Web archiving memes:

The collection or capture of content during Web archiving must include the many iterations of a meme—this can include hundreds of images, and they will all contribute to the overall meaning or theme of the meme as a whole, during analysis. Furthermore, Rose’s modalities can be applied to assessing the capture and collection of a meme: each iteration of the meme will share essential components of the image itself; the site of production will require similarities throughout all iterations in order to ensure clear parallels and shared characteristics; finally, the site of circulation will reveal modes of
sharing or communication of the iterations of the memes, define the collection, and identify the variations that occurred through the exchange of each iteration.

The second element I extracted from Shifman’s meme primer is her proposal of meme genres. Shifman outlines a total of nine meme genres through a survey of memes: Photo fads, flash mobs, reaction Photoshops, lipdubs, misheard lyrics, recut trailers, LOLCats, rage comics, and stock character macros. (Examples of these genres are included in Appendix A.) Shifman further categorizes these nine genres into three groups.

“An initial observation stemming from this survey is that meme genres can be divided into three groups: (1) Genres that are based on the documentation of “real-life” moments (photo fads, flash mobs). These genres are always anchored in a concrete and nondigital space. (2) Genres that are based on explicit manipulation of visual or audiovisual mass-mediated content (reaction Photoshops, lipdubs, misheard lyrics, recut trailers). These genres—which may be grouped as “remix” memes—often reappropriate news and popular culture items. Such transformative works reveal multifaceted attitudes of enchantment and criticism toward contemporary pop-culture. (3) Genres that evolved around a new universe of digital and meme-oriented content (LOLCats, rage comics, and stock character macros). (Shifman, p. 118)

Shifman’s meme genres establishes the second principle for the proper development of Web archiving memes: Memes must be collected with all images present and sufficient metadata to identify and then categorize by the structure through which they were designed—the structure of the memes can indicate textual syntax and the types of images used, which may impact the interpretation of meaning. Here again, Rose’s modalities can be applied to assessing the capture and collection of a meme: the image itself to indicate meaning; the site of production to define the genre that the meme belongs to; and what
changes occurred through the circulation of the original image when compared to the collection of each iteration.

Web archival requires software to scan and capture data associated with the websites that have been assigned for collection. The Web archival software or Application Programming Interfaces (APIs) are traditionally designed and modified related to the specific archival needs of the project. The software is pre-defined with a series of algorithmic functions that will scan and capture pre-determined websites for a Web archive.

“While there are a number of common tools found across a variety of web scraping projects, there are no one-size-fits-all programs available due to the wide variety of content hosting platforms and the pace at which both they and the web language standards change.” (Black, p. 97)

The software has to be modified to fit the scope of the archival project, which requires a clear understanding of the needs and desires of the archival collection. This presents an issue for Web archiving because it requires the archivist to pre-determine the digitally born content to have enduring value and to know where to collect this content.

“Not only are websites often constellations of thousands of files, each of which can thought of as an archival record or a publication, decisions have to be made about the appropriate ‘depth’ of web ‘harvesting.’” (Hand, p. 138)

Again, the archivist must pre-determine what is to be collected in order to design the software to capture all necessary data for the collection. If the software is not designed to travel through the many layers of a website (follow the hyperlinks to sub-sections and other pages) not all data will be captured. This can negatively impact the completeness of the Web archive.
When a Web archive is created it is capturing the website at a particular moment in time. It is common for the software to be programmed and scheduled to scan and capture the website during many instances over a period of time. The data is typically stored and time stamped to indicate when the scan was captured. The purpose of the multiple scans is to capture content as it is added to a particular website. Websites can be updated daily, weekly, monthly, or less frequently so the software has to be programmed to perform scans that suit the behavior of the site. However, this requires multiple collections of data for the same website to capture the changes over time. The software will capture the changes, but it will not indicate the changes for you.

“Although a broad Web archive is a collection of Web material, it cannot be considered a corpus, that is, a clearly delimited and structured set of elements (words, texts, images, etc.). When studying the online Web, one can select a corpus to study, for instance, a set of URLs or file types. But since a number of versions of each element exist in a Web archive (cf. above), one has to construct not one corpus, but two. First the URLs that should be included in the study, and second the specific versions of each of these URLs.” (Brügger et al, p. 77) In order to analyze the contents of a Web archive the researcher must explore multiple collections of a collection. Consider this within the context of memes, which are large collections of images by definition. “If everything from the past is saved, it becomes close to impossible to actually find significant materials.” (Post, p. 69) (There is further discussion that is required here in order to determine how metadata is captured during the Web archival process that can aid in filtering out multiple instances of the same content over time to eliminate unnecessary reviews of duplicate information.)

When I explored the captures of KnowYourMeme.com through the American Folklife Center archives, I discovered some issues. (Appendix B includes screenshots of
some of the issues I encountered while exploring this archive.) First, there were images missing from pages. The missing images suggest that there may have been web page loading delays or lags, and/or the web-crawling software could have been processing too long and completed the scan before all images and content were captured. Second, there are layers or depths of the website that are missing that prevent full navigation of the historical capture. When I attempted to perform a search for a specific meme that was prevalent in 2016 there were no search results returned. Instead, the LC directed me to the current KnowYourMeme.com website to perform my search where the content was originally captured from. It may be the case that this content still exists on the original website, but if it is missing from the archive it leaves me to question what other content is missing, and is the archive complete enough for myself or other researches to extract adequate visual data and metadata to perform analysis and interpretation?

**SUMMARY**

Folklore is categorized and evaluated through established methods of structure and algorithmic analysis. In this paper I presented memes as a candidate for research as an example of digitally born folklore. In order to analyze and interpret memes as digitally born folklore I presented the essential definition of memes and the genres they can be categorized within in order to apply visual methodologies. Finally, I presented a basic overview of how Web archival programs work and the issues that can impede the completeness of the collection.

Memes are visual images that exist as large collections of iterations and are shared across the Internet. In order for a Web archive to be complete and support an
adequate collection of data for analysis it is important that images and metadata are not lost, missed, or overlooked during the scanning and capture process. These issues were prevalent in the review of the American Folklife Center, so they are valid concerns for future research. Subsequently, future research should focus on the methods of archival collection of visual collections such as memes in order to capture their essential components for analysis.
REFERENCES


BIBLIOGRAPHY


Dorst, J. D. (2016). Folklore's cybernetic imaginary, or, unpacking the obvious. *Journal Of American Folklore*, (512), 127.


APPENDIX A

Figure 1: Photo Fad, planking.

Figure 2: Flash mob.

Figure 3: Reaction Photoshops.
Figure 4: Lipdubs.

Figure 5: Misheard Lyrics.

Figure 6: Recut Trailers.
Figure 7: LOLCats.

Figure 8: Rage Comics.

Figure 9: Stock Character Macros.
APPENDIX B

Figure 1: Images are missing from the archive.
Figure 2: More images are missing from the archive.

Figure 3: Failed archive image search results.