Articulating a Vision for Community-Engaged Data Curation in the Digital Humanities

by

Lydia Zvyagintseva

A thesis submitted in partial fulfillment of the requirements for the degrees of

Master of Arts in Humanities Computing and Master of Library and Information Studies

Humanities Computing / Library and Information Studies
University of Alberta

Abstract

The purpose of this study was to identify critical elements in a conceptual model for a community-engaged data curation in the digital humanities, to propose a set of evaluation criteria that would act as guiding principles in pursuing such work in the future, and to explore ways in which community-engaged data curation practice can further the mission of public digital humanities. I selected 28 data curation projects taking place in the digital humanities as my study population and collected data in the form of observations from the project websites, seeking evidence of emergent themes related to the categories of data curation, digital humanities and community engagement. I adopted a grounded theory methodological approach with a pragmatist theoretical perspective, as such approaches aligned well with the evidence-based orientation of this research study.

Through a constant comparative data analysis method typical in a grounded theory study, I was able to identify the key elements of community-engaged data curation in the digital humanities as being systematic, reflective and participative. Having articulated the key elements of the conceptual model, I constructed an evaluation framework for conducting projects of this nature, with such criteria as adherence to established metadata standards, prioritization of data accessibility, and connecting the goals and mission of the project to its operations and data management practices. Finally, based on the examples observed in the study population, I argue that community-engaged data curation can become a form of advocacy for public digital humanities by seeking partnerships with community groups, committing to making research data accessible and providing meaningful ways for the public to contribute to the research projects.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Currently, many research institutions are involved in digital humanities research, such as the production of digital editions of manuscripts, text-mining tools, database-driven interactive websites, gaming and virtual environments, mobile applications, and other textual, graphic and multimedia digital objects. For example, the University of Virginia, University of North Carolina at Chapel Hill, City University of New York, University of Maryland, University of Nebraska-Lincoln, University of California, University of Alberta, University of Victoria, and King's College London, among others, all have established programs, centers, or labs to foster digital research in the arts, humanities and media studies. Data curation is the process that addresses the challenge of managing data produced as a result of research through planning, selection, preservation, description, management, edition, and reuse of data over time. For this thesis, I conducted an exploratory study to understand data curation in the digital humanities specifically. As background for my study, I relied on research on the topic of community engagement, participation and participatory culture. I also brought together concepts, definitions and ideas surrounding data curation and digital humanities. I was ultimately interested in exploring the call made by several scholars in the field for digital humanities to find ways to support broader goals of public humanities and connect the research process with the general public. As a result, my study examines the role of data curation in the landscape of public humanities and digital scholarship. I also argue for a new phase in the consciousness of the digital humanities: that of the critical awareness of issues surrounding impact and community engagement in digital scholarly practice.

Research Problem

The purpose of this study is to identify critical elements in the conceptual model for data curation in the digital humanities with a focus on community engagement. Using an evidence-based, grounded theory approach, this study aims to develop an evaluation framework for data curation projects taking place in the digital humanities. The potential outcome of this type of research is a deeper conceptual and practical understanding of the notion of humanities data curation, the development of a theory of community-engaged

data curation in the digital humanities, as well as evidence-based best practices to guide scholars, librarians, and other practitioners undertaking projects of this nature. As a result, this study aims to address the following research questions:

1. What are the key elements in a conceptual model for data curation in the digital humanities?
2. Based on these key elements, can an evaluation framework for digital humanities projects involving data curation be proposed?
3. How can data curation support the goals of public humanities to foster the pursuit, development and sharing of knowledge beyond the academy?

Outline

This study is organized into chapters in the following manner: Introduction, Literature Review, Research Design, Findings, Discussion, and Conclusion. The Literature Review chapter provides a critical and theoretical context for the study and is organized into 3 major categories, reflecting the three main concepts that this study addresses: data curation, digital humanities, and community engagement. Each category is further divided into themes around pertinent trends that appear in the research literature about these concepts. The Research Design chapter outlines the methodology, method, data collection and data analysis procedures used in this study. The Findings chapter, like the Literature Review, is organized around 3 major concepts of data curation, digital humanities and community engagement with detailed examination of each concept according to the evaluation criteria developed as a result of this study. The Discussion chapter answers the research questions posed above and includes a detailed discussion of the key elements that make up the proposed conceptual model for community-engaged data curation digital humanities communities based on the findings of this study. Finally, the Conclusion summarizes the study and addresses further research of critical importance to the subject of data curation in the digital humanities in order to further test the validity and reliability of this study, and to continue to deepen the understanding of this topic in the growing body of scholarship.

Chapter 2: Literature Review

Introduction

While data curation has received some attention in library and information studies (LIS) and science literature, few studies have examined the practice from the perspective of the digital humanities specifically. In addition, a fair amount of scholarship has been produced about information seeking and information behaviour of humanists. For example, we know that humanists work in ways that are markedly different from scientists and even social scientists, and as such, have different information needs and behaviours. Specifically, they prefer to start from a broad collection of texts and narrow down to a specific question. Thus, the ability to browse, search, highlight, annotate, and iterate between texts is critical to their research process (Vandergrift & Varner, 2013). A fair amount of research on the information-seeking behaviour of humanists has revealed practices common to scholars in this field, such as "netchaining" (Sukovic, 2011; Ge, 2010; Talja, Vakkari, Fry & Wouters, 2007). Studies examining the information behaviour of humanists demonstrate that this group of scholars is already using the internet as a large text, relying on web search engines as finding aids for digital resources (Toms & O'Brien, 2006). Furthermore, humanities scholars are embracing digital technologies to enrich their pedagogic practice, aggregate and manage vast amounts of textual data in order to gain new understandings of the texts they study (Juola, 2008). In addition, information-seeking behaviour of humanists has also been analyzed from the perspective of intentionality in seeking digital library services (Chang, Lin, Chen & Chin, 2009), their common use of electronic resources and technologies (Tahir, Mahmood & Shafique, 2008) as well as their perception of such resources (Ge, 2010; Talja, Vakkari, Fry & Wouters, 2007). Humanists, in a sense, "are curators par excellence of scholarly information" because they transform primary "raw" data into secondary "institutional" content (Bernardou, Constantopoulos, Dallas & Gavrilis, 2010). As a result, knowledge about the information-seeking and research process of humanists aids the formulation of policies that guide the development of information services delivered by libraries to this user group.

Is there such a thing as humanities research data? This project operates on the assumption that in order to have data curation in the digital humanities, we must have humanities data. While even digital humanists tend not to think that the material with which they work as "data", many nonetheless acknowledge that their projects either require the production or the processing of certain amounts of digital material – digitized archival content, audio recordings, video clips, large databanks of texts. All of this content can clearly be considered as digital humanities data. This project does not distinguish between data produced as a result of the project or data that existed already and therefore enabled the project to happen. However, not all digital humanities projects can be classified as data curation projects by virtue of having data. The data that make up digital objects found in DH projects are often unstructured, complex, heterogeneous and diverse. For example, projects like the University of Alberta's *Atlas of Alberta Railways* (Lester, 2005), scholarly projects built in Hypercities, University of Victoria's MakerLab blog (2014), University of Washington's "Women Who Rock" (2014) archive, and even Pasanek's "The Mind is a Metaphor" (2010) database of metaphors at the University of Virginia all have various kinds of data, including text, numeric information, images, as well as video and audio recordings. Yet not all organize and make accessible the results of their research through the project site or through their institutional repository, which I argue to be a main component of data curation practice. Managing the data within these complex multi-layered projects involves many activities, skills and competencies, yet few studies looked into the act of data curation in digital humanities projects 'on the ground level'.

Clearly, the active management of digital humanities research essentially represents the emergence of humanities data curation, a term that might mean more to library professionals than to humanists. Data curation is a fairly new term, which emerged around 2003 to describe the practice that addresses the challenges of maintaining digital information over its entire lifetime, as it is useful to researchers. Data curation, then, involves managing digital research data to facilitate access, dissemination, and preservation of both information content and context –data and metadata. In the LIS research literature, data curation has been discussed both conceptually (Yakel, 2007; Ogburn, 2010; Collie & Witt, 2011; Bengtson, 2012; Tibbo, 2012) as well as in relation

practice, such as from the perspective of various academic institutions (Choudhury, 2008; Witt et al., 2009; Hswe, Furlough, Giarlo & Martin, 2011; Neuroth, Lhmeier & Smith, 2011; Prom, 2011; Rice & Haywood, 2011; Ward, Freiman, Jones, Molloy & Snow, 2011; Wilson, Martinez-Uribe, Fraser & Jeffreys, 2011). There is also a growing awareness that librarians need new types of skills in order to embrace the new roles of digital curators (Heidorn, 2011; Newton, Miller & Stowell Bracke, 2011; Ray, 2012). Benefits of data curation for digital humanists and users of their projects include data integrity and stability, greater interoperability, accessibility and discovery, among others. Data curation ensures the security and availability of research data for future use, and ultimately translates into what Alemu, Stevens, Ross & Chandler (2012) describe as a "dramatically improved infrastructure of knowledge" that makes up the internet today (p. 556). Data curation, in essence, is a value-adding activity carried out in libraries, research teams and non-profit educational organizations, and its importance cannot be overestimated when assessing the value of a digital humanities project. With this in mind, I would like to situate the concepts of digital humanities, community engagement and data curation in a critical and theoretical context that provides a useful perspective on the findings of this study.

1. Data curation

1.1 Definitions

Scholars Lee and Tibbo (2007) point out that the notion of data curation continues to evolve. They provide a definition, which states that data curation

> involves selection and appraisal by creators and archivists; evolving provision of intellectual access; redundant storage; data transformations; and, for some materials, a commitment to long-term preservation. Digital curation is stewardship that provides for the reproducibility and re-use of authentic digital data and other digital assets. Development of trustworthy and durable digital repositories; principles of sound metadata creation and capture; use of open standards for file formats and data encoding; and the promotion of information management literacy are all essential to the longevity of digital resources and the success of curation efforts.

Data curation can also be defined as the active and on-going management of data "through its entire lifecycle of interest and usefulness to scholarship, science, and education" (Noonan & Chute, p. 203). University of California San Diego summarizes the definition as the activity of "managing data to ensure they are fit for contemporary use and available for discovery and reuse." In other words, it is the activity of "managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose" (Yakel, p. 338). The notion of data use and re-use over *its entire lifecycle* is critical to understanding data curation, which is why I believe it is important both to digital humanists engaged in digital scholarly projects and librarians working with them to manage and preserve that data. After all, research project goals and design will determine the kinds of data that are produced as a result of carrying out a digital scholarly activity, which will then influence other research projects that use, cite, transform, edit, describe, and share that data. Curation is implicit in the act of selection and explicit in the act of description and repurposing.

The United Kingdom's Digital Curation Center defines digital curation in exactly the same manner as the sources above (2014), indicating that the terms "digital curation" and "data curation" are interchangeable and may be a matter of cultural preference rather than an inherent distinction. For the purposes of this study, I am not distinguishing between the terms *data curation* and *digital curation*. I am using solely the term 'data curation' in this study because its use by Yakel (2007) and others suggests a North American preference for the term, whereas the use of 'digital curation' by the Digital Curation Center and others suggests a European preference.

Related terms include digital *archiving* and *digital preservation*, which I also define for the sake of clarity. Digital archiving may be described as "a curation activity which ensures that data is properly selected, stored, can be accessed and that its logical and physical integrity is maintained over time, including security and authenticity" (Yakel, p. 338). In other words, archiving is the process of active storage of digital content that exists somewhere between curation and long-term preservation practices. It

is about "ensuring data are properly selected, appraised, stored and made accessible" (Noonan & Chute, p. 203).

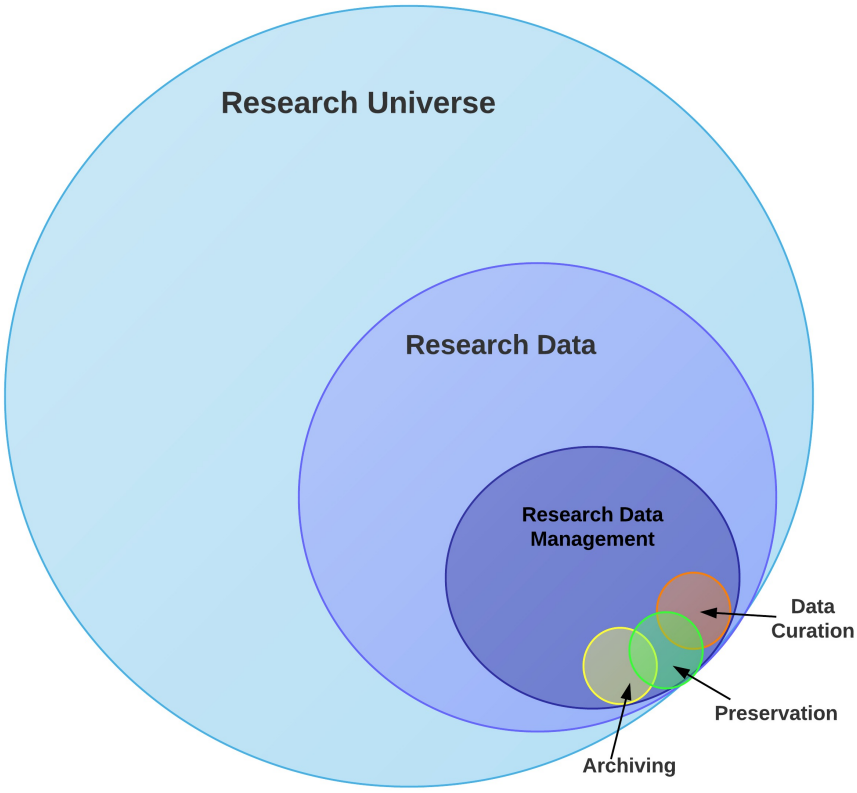
Digital preservation is "an archiving activity in which specific items of data are maintained over time so that they can still be accessed and understood through successive change and obsolescence of technologies" (Yakel, p. 338). In this case, preservation is concerned with the long-term perspective of most institutional libraries to ensure that their collections are available indefinitely. It therefore relates to the technical capacity of making data physically accessible and usable. For example, the University of Alberta Libraries aims to preserve its wealth of knowledge for the next 500 years. As Noonan and Chute put it, preservation ensures "that items or collections remain accessible and viable in subsequent technology environments" (p. 203).

Fundamentally, data curation is about organization, preservation, access and sustainability of knowledge – values also shared by digital humanists (Spiro, 2011). It is a value-adding activity focused on stewardship of digital assets. Flanders and Munoz add that the term curation carries the dual emphasis of protection and "amelioration, contextualization, and effective exposure to an appropriate set of users" (2012, p. 4), thereby enriching and preserving knowledge in the digital form at once. While the use of such terms as "appropriate users" is potentially contestable, they address the growing awareness of public engagement in the digital humanities through such key concepts as interest and usefulness of scholarship that takes place in the form of data curation. They argue that the term scholarship should not be limited to academic activity, since data creation, use and repurposing all take place both in research teams and beyond them. In light of recent and growing emphasis on crowd-sourcing initiatives and community-driven data creation, the boundary between the academy and the public sphere is becoming more permeable. As a result, "it is reasonable to anticipate that definitions of 'scholarship' may continue to broaden rather than narrow over time" (2012, p. 4). In this way, data curation may constitute an apt embodiment of community engagement in a research environment.

It is clear that despite my effort to differentiate the terms archiving, curation and preservation, they overlap conceptually as they all relate to the notion of information access and usability. To represent the relationship between these terms, I have

constructed a diagram represent their conceptual positions. Figure 1 below illustrates these positions, with data curation, archiving and preservation being very specific sub-sets of a broader range of activities that make up research data management, but sharing certain conceptual elements, and thereby intersecting inside the research data universe.

Figure 1: Research data universe



1.2 Is there a theory of humanities data curation?

In the *Companion to the Digital Humanities*, Schreibman, Siemens & Unsworth (2004) define the digital humanities as the emerging "discipline in its own right" (History section, para. 1) in which critical inquiry involves "the application of algorithmically facilitated search, retrieval, and critical process" to humanities-based work. According to the authors, "exemplary tasks traditionally associated with humanities computing hold

the digital representation of archival materials on par with analysis or critical inquiry, as well as theories of analysis or critical inquiry originating in the study of those materials" (Principles, Applications and Dissemination section, para 1). In other words, making digital scholarship possible by building large corpora, software and social systems that support it is as much a scholarly pursuit as analyzing the digital data themselves.

In her 2009 article "The Digital Future is Now: A Call to Action for the Humanities," Christine Borgman asserts that "digital content, tools and services all exist, but they are not necessarily useful or usable." Data curation addresses this gap by emphasizing usability of data through systematic planning, organization and preservation of information. In fact, it is a critical piece to making the results of research usable. Despite relative progress in the digital humanities, the development of multitude of applications, projects, and academic programs, Alex Poole remains critical of the lack of impact that data curation has had in this world of digital data. He argues, then, that "ultimately, it remains unclear when a critical mass of case study evidence will be assembled to address the stubborn concerns" regarding infrastructures, professional development and education in data curation, archiving and preservation services, and new professional practices that include advocacy of the expanding discipline (p. 70). In other words, there is plenty of talk and not always a lot of action with respect to data curation in the humanities.

Do the digital humanities, through such practices as data curation, enable what Delanty (2001) argues as the "articulation of technological and cultural forms of citizenship to complement the older civic, political and social rights of citizenship" (p. 5)? In my view, failure of the digital humanities to demonstrate a sense of impact – in terms of relevance and support from the general public – has resulted in the resource gap identified by Borgman and Poole. Nevertheless, due to the content with which they work – historical texts, maps, visual art, narratives, and film, for example – the digital humanities potentially have a high degree of appeal to the general public and are well positioned to support the mission of public humanities. Data curation in the digital humanities, by that logic, is one way to sustain that appeal and capitalize on the power of the humanities to communicate that sense of impact.

2. Digital humanities

I approach this study from the belief that the humanities are important because they allow us to make sense of our world. In other words, the humanities exist not just in the sphere of the academy, but also in the public at large, in popular culture, in language and in the arts. I am arguing, then, for a renewal of the idea of truly public humanities – those that exist for the people. As a result, I would like to explore the concept and reality of data curation in the digital humanities from the perspective that data curation can live out the values of public humanities by incorporating elements of community engagement. With this in mind, I would like to define the terms of my argument and provide a bit of background to their formation.

2.1 What are the digital humanities?

As mentioned in the introduction, this project aims to contribute to the growing body of knowledge regarding stewardship of digital scholarship in the form of data curation. Specifically, I would like to focus on data curation in the digital humanities (DH). For the purposes of this project, I am treating the *humanities* and the *digital humanities* as separate concepts, and would like to begin this chapter by outlining the difference in their definitions as well as a brief history of this distinction. While many scholars point to the fact that hardly any humanist practicing today does not engage in some way with the digital, not all humanities scholars employ *digital methods* in their investigations. By digital methods, I mean the use of tools, software and approaches that enable information processing beyond the capacity of human effort in terms of size, speed and function. For example, the use of geographic information systems software like ArcGIS allows researchers to relate multiple data points on a digital plane. While such a tool may not replace mapping as a conceptual process of locating objects in space as represented on a plane of the map, it nevertheless allows for a new kind of process of research not possible before. The use of software to approach information in ways that

are significantly differently from the human brain in the humanities arguably both supports new research avenues and alters the very approach to research as a process. Certainly, not all humanists using computers are necessarily "doing digital humanities". Even those using digital methods such as analyzing the exchange of letters across time and space using a network-mapping approach, for example, do not necessarily think critically about the role of digital approaches to knowledge in their research.

Interestingly, there is a certain earnest assumption that exists among some digital humanists that eventually, and likely soon, 'all H will become DH'. Or rather, much like all science is now digital insofar as no distinction is made between *e-science* and *science* in academic discourse, no distinction will need to be made between digital humanities and broader humanities. I am concerned, then, with exploring how digital humanities might become a flagship for public humanities, and I address both its definition and its articulation in practice further in this chapter.

What might we mean when we say "digital humanities?" Some view DH as "an understanding of new modes of scholarship, as a change not only in tools and objects but in scholarship itself" (Parry, p. 436). For instance, when all published literature of the Nineteenth Century is digitized and publically available online, how does that change the kinds of questions scholars can ask of this body of knowledge? What happens to the nature of research when humanists have at their disposal both unprecedented sources and tools with which to engage those sources? Hayles (2012), for example, regards the digital humanities as a diverse field of practices "associated with computational techniques and reaching beyond print in its modes of inquiry, research, publication, and dissemination" (p. 27). Lisa Spiro defines it most broadly, arguing that "the digital humanities reconfigures the humanities for the Internet age, leveraging networked technologies to exchange ideas, create communities of practice, and build knowledge" (p. 21). In other words, the digital humanities are not only about *what* is being studied, but also *how* and *why* it is being studied.

2.2 Brief history of the digital humanities

While the purpose of this project is not to rewrite the history of DH, it is nonetheless useful to set a critical assessment of the digital humanities with respect to

data curation in a larger context of intellectual developments that took place in the discipline over past 20 years. Several critics argue that there are two phases of digital humanities scholarship that have emerged in a fairly chronological order. Evans and Rees, for example, see the emergence of a division between digital humanities as a set of quantitative tools and research methods, and the digital humanities as "a newly emerging field influenced by computation as a way of accessing, interpreting, and reporting the world itself" (29).

From this perspective, Phase I of the digital humanities includes building of large databanks, massive digitization initiatives, and establishment of digital infrastructures. This period of digital humanities work was quantitative, "mobilizing the search and retrieval powers of the database, automating corpus linguistics, stacking hypercards into critical arrays" (Schnapp & Presner, 2009). Phase II, on the other hand, includes reflexive and experiential types of scholarly inquiry, such as the focus on digital pedagogy, gaming and multimedia studies as well as critical reflection on the discipline itself. The second wave of digital humanities is "qualitative, interpretive, experiential, emotive, generative in character" (Berry, 2012, p. 3). Some scholars claim, for example, that there are, in fact, at least two types of "digital human*isms*: one that sees the digital as a set of tools to be applied to humanistic inquiry (design, project, tools, data) and another that sees the digital as an object of study (social media, digital games, mobile computing)" in and of itself (Parry, p. 433).

We might also think about these phases in terms of methodological strategies employed in the course of scholarship, which Hayles proposes as the notions of *assimilation* and *distinction* (2012). By her definition, assimilation "extends existing scholarship into the digital realm: it offers more affordances than print for access, queries, and dissemination; it often adopts an attitude of reassurance rather than confrontation." Activities such as corpus construction, annotations, making scholarship more accessible online might follow this DH strategy. Distinction, in contrast, "emphasizes new methodologies, new kinds of research questions, and the emergence of entirely new fields" (Hayles, p. 46). These fields might include such subject areas as augmented-reality studies, gaming theory, geospatial humanities, and critical and reflexive digital modes of pedagogy, for example.

Finally, scholars like Manovich (2013) call for a Phase III, which might include purposeful interrogation of the very *computationality* of digital scholarship with a focus on the materiality of code and computational techniques. For example, Berry calls for an effort to develop "methods, metaphors, concepts and theories in relation to this software and code to enable us to think through and about these systems, as they will increasingly have important critical and political consequences" (2014, p. 40). If we follow this logic, then, I would argue for a fourth phase in the consciousness of the digital humanities. This stage of DH would include a critical awareness of issues around impact and community engagement in digital scholarly practice. The emergence of community engagement in theory and practice in the academy, as well as its intersection with the digital humanities, is discussed further in this chapter.

2.3 What are the humanities for?

Lisa Spiro (2012) reminds us that core humanities values include inquiry, critical thought, debate, pluralism, balancing innovation with tradition, as well as exploration and critique (p. 19). Furthermore, core methodological strengths of the humanities include crucial skills like attention to complexity, analytic depth, critique and interpretation (Schnapp & Presner, 2009). If the humanities have existed for so long, it is because they remain relevant by allowing us to pursue inquiry into significant questions about what it means to be human. They contribute to the needed perspective, training in complex human phenomena, and communication skills "needed to spark, understand, and make "human" the new discoveries" (4Humanities, n.d.). In that case, we should be asking, "what the digital does to our concept of the humanities, and, by extension, even our concept of the human" (Parry, p. 435). As Alan Liu eloquently argues, we need human*ism* and not necessarily the human*ities*. In fact, in this data-saturated neoliberal post-industrial techno-managerial academic environment, digital humanities are needed more than ever constantly to interrogate the technological, social and cultural trends shaping out experience. According to Liu (2012),

> There is not a single "grand challenge" announced by the Obama administration, the Grand Challenges in Global Health initiative, the U.S. National Academy of Engineering, and other

> agencies or foundations in the areas of energy, environment, biomedicine, food, water, education, and so on that does not require humanistic involvement. All these issues have a necessary cultural dimension, whether as cause or effect; and all, therefore, need the public service of humanists, and increasingly, digital humanist participants (p. 502).

In response to the question, *What are the humanities for?* Liu founded 4Humanities, a group created by the international community of digital humanists and educators "to assist in advocacy for the humanities" (4Humanities, n.d.) Perhaps what we need is not necessarily the digital humanities, but engaged public humanities, which are able, through digital and non-digital means, to connect to broader communities, communicate their mission, and contribute to thriving civic societies.

Furthermore, I believe it is also significant to outline my understanding of the relationship between the concepts of data curation and digital humanities. As this thesis limits its scope to examine data curation in the digital humanities alone, there is a risk to conflate the terms or assume both achieve the same goals. If we think of the concept of data curation in terms of its original Latin meaning (Oxford, 2014), *cura* – the care for – its relationship to the digital humanities and other terms, such as archiving and preservation, for example, becomes clearer. The role of data curation, then, is to care fore, to take care, to manage for future use. While some aspects of the digital humanities are certainly concerned with such activities, ultimately, the digital humanities are about seeking new understandings and deriving new meanings. Therefore, some DH work is data curation and some data curation work is DH in its intellectual contribution to research, but not all DH is data curation and data curation is not DH, anymore than it is not natural science or social science. For example, creating and sustaining a digital archive devoted to the idea of the current historical moment, as Yale's Historian's Eye project (2010) seeks to do, is a form of intellectual work, even where it stops short of analysis. In other words, organization of knowledge through efforts like archive building and creation of visual art APIs, is done in the context of the digital humanities, but it is not necessarily digital humanities work in and of itself, as it does not utilize digital methods of inquiry. Nevertheless, I would argue that this type of work is data curation as it contributes to sustaining and improving accessibility to knowledge through digital means.

2.4 Public humanities

Scholars agree that projects in the digital humanities may be particularly well positioned to reach audiences beyond the scholarly community, because they appeal to a broader public, "even if doing so was not part of their creator's intent" (Maron & Pickle, p. 9). For example, despite its significance to the improvement of human health, research in molecular cell biology does not necessarily have the same public appeal as studies of Ancient Roman life or the speeches of Abraham Lincoln. While fields like astronomy make use of public interest in the cosmos by asking for help from 'citizen scientists' in helping science researchers sift through massive amounts of data generated in experiments, such as those employed by the Large Hadron Collider at the European Organization for Nuclear Research, humanities are even more likely to be able to make use of public interest by engaging many 'citizen humanists' in large humanities projects. In other words, through multiple technologies, the digital humanities can stand for greater, public humanities, in their ability to bridge differences, foster dialogue, and enable lifelong learning in multiple user groups by providing access to the human record in the form of digital humanities data. Arguably, we need public humanities that serve as both a public-facing scholarship and are committed to a dialogue with broader publics through the tenets of research, teaching and service. In this sense, truly public humanities would ultimately contribute to engaged educational institutions that embrace communities as equal partners "who work with, not for, universities in a mutual exchange to discover new knowledge and promote and apply learning" (Stewart & Alrutz, p. 45).

In light of this discussion, I find David M. Berry's Six Principles of the Digital Humanities (Twitter, 2014) particularly useful in understanding the relationship between digital research and community engagement that might take place in the practice of truly public digital humanities. These principles include:

1. Engaging and sharing with the academic community, students, public and institutions
2. Building and making new things, new methods and meanings
3. Committing to a critical reflexive process in individual and collectivities
4. Being and becoming interdisciplinary, inclusive and connective

5. Fostering encounters with new and old objects, texts, humans and non-humans
6. Confronting matter and the materiality of things.

While not all scholars may share every single priority outlined above, I would argue that Berry articulates a compelling mission for humanistic scholarship in the digital age. He understands the avenues scholars need to pursue in order to create truly innovative, impactful and important contributions to the field of the humanities. These principles are therefore useful in considering any digital humanities project from its launch to evaluation.

3. Community-engaged scholarship

3.1 What is a community?

The topic of community has been discussed extensively in social science research literature, from sociological, psychological, anthropological to linguistics perspectives. Given the broad nature of this term, it is important to note that its use will reflect the assumptions and the disciplinary traditions of scholars who discuss the concept. For the purposes of this thesis, I am relying on a general definition of community focusing on the specific aspects of interest and interaction rather than aspects relating to geography or affect, as the former are elements evident in the subject of this study. Most broadly, a community can be defined a group of people sharing something in common, usually a geographic area or a common interest. Many definitions center around the notion of interaction and exchance. For example, Hillery, one of the foundational scholars on this issue, posits that community consists of "persons in social interaction within a geographic area and having one or more additional common ties" (1955, p. 111-123). Similarly, Freilich's pre-internet definition (1963), which may be more useful in analyzing some data curation projects than others, proposes that the essential ingredients of the concept of community include "people in relatively high-frequency interaction, exchanging information as a set of related centers, and practicing and developing local interaction culture based on past information shared" (p. 127). Clearly, in this model, Freilich strives

for an operational understanding of community based on the ideas of people interacting in a geographic area.

In the 1990's, social scientists began to redefine community within a postmodern context to strive toward an understanding that community should be based on interdependence between diverse individuals as well as highlight the role that differences play in this construction of community. Critiques from scholars like Stone (1990), Young (1990) and Phelan (1996), for example, emerged to underscore the notion of community being highly problematic "given that it was based primarily on creating connection through unity and sameness, which ultimately excludes as much as it might include" (Bettez, 2011, p. 8). Burns (1994), too, recognizes that the community is not a singular concept, and argues that in fact, "community" acts as an umbrella term under which many varying, competing and often conflicting interests come together.

One of the most common elements identified in the research literature on community is the notion of belonging and cohesion. Consisting of local, social relationships, belonging is particularly important because it is a "powerful, fundamental and extremely pervasive" human motivation to fulfill the need for attachment (Baumeister & Leary, 1995, p. 497). Similarly, for Block (2008), community is about the experience of belonging: "we are in community each time we fine a place where we belong" (p. xii). Further emphasizing the ideas of harmony, agreement and belonging is Pharr's (2010) definition, which conceptualizes community as "people in any configuration (geographic, identity, etc.) bonded together over time through common interest and concern, through responsibility and accountability to one another" (p. 594). Building on these definitions, Bettez (2011) proposes an understanding of the concept in terms of processes and goals, in which "continually shifting groups of people … actively listen to, and support each other, through reciprocal responsibility and accountability" based on a common interest or concern (p.10). Clearly, the discussion of belonging is related to the notion of affect, as the experience of feeling or emotion, which has been explored in psychological research (Zajonc, 1980; Lazarus, 1982; Brewin, 1989; Damasio, 1994; Griffiths, 1997; Lerner & Keltner, 2000).

Another facet of the concept of community has been the development of the notion of communities of practice, which can be understood as a "collection of people

who engage on an ongoing basis in some common endeavor" (Eckert, 2006, p. 683). Two conditions required to create a community of practice are shared experiences and shared understandings. In other words, community of practice is about making meanings or sense-making as well as mutual engagement. Outward-facing by virtue of relating to the world through conscious sense-making, communities of practice emerge in response to objective conditions. As a result, every community of practice "fits into a complex structure that connects individuals to each other and to the political economy" (Eckert & McConnell-Ginet, 2007, p. 32). Related to this conceptual framework, an information community can be understood as a "partnership of institutions and individuals forming and cultivating a community of interest around the provision and exchange of information, aimed at increasing access to that information or increasing communication, and thereby increasing that knowledge-base" (Durrance, 2001, p. 164). These rich understandings of the term community are useful in exploring the theme of community engagement and its relationship to digital scholarship.

3.2 Brief history of community-engaged scholarship

Much like the history of DH, the development of ideas surrounding community engagement in the academy forms its own trajectory, and I believe it is useful for the purposes of this discussion to summarize them here. While the establishment of American Land Grant universities in the Nineteenth Century carried with it the explicit mission of service to the community, "denoting primary research and teaching that would assist with the economic, social, and cultural developments of the city or region where the university was located" (Schuetze, 2012, p. 62), the dominant model of service learning in the United States has historically focused on the education of students, rather than the achievement of community goals (Stoecker, Loving, Reddy & Bollig, 2010, p. 281). Holland (2001) argues that with time, universities were pressured to demonstrate their role in contributing to the economic development of a region as well as "to the enhancement of social fabric and community capacity through campus-community partnerships" (Lockwood, Lockwood, Krajewski-Jaime & Wiencek, p.89). In fact, despite the declaration of service missions in most North American universities upon their establishment, this well-intentioned rhetoric does not translate into active

engagement "and serious commitment" to the idea of service on the part of the institutions (Schuetze, p. 64). In such cases, service to the community is often limited to such practices as knowledge transfer and service learning, as it continues to be the case in many academic institutions across North America today. Service learning is often framed as "a service or charity to those less fortunate, a sort of gift" in the form of outreach to the community (Stewart & Alrutz, p.45). Studies show that the concept of knowledge transfer has been framed as a passive relationship from expert – university – to beneficiary – community (Bringle & Hatcher, 2002). However, more recent discussions of community-based research and service learning as they fit into the community-engaged scholarship philosophy are based on the understanding that "new knowledge is generated by a collective process, and that the universities cannot generate such knowledge alone" (Schuetze, pp. 68-69). The new perspective on knowledge creation is therefore relational, reciprocal and participatory. "This situation generates a struggle between what universities affectively feel are incompatible purposes, but cognitively view as a civic responsibility. The crux of university-community engagement lies in operating from both the cognitive and the affective domains" (Lockwood, p. 89).

For Fenwick (2014), community engagement in higher education institutions means "deeper engagement with [concepts of] uncertainly, difference and responsibility" that is best examined through sociomaterial approaches to understanding communities and knowing, "where knowing is taken to emerge and be performed within the entanglements of materials with the social and personal: bodies, objects, technologies and places" (p. 121). Hoyt (2010), on the other hand, understands engagement as progression "from a technocratic to a democratic way of knowing" through five stages, which include pseudo-engagement, tentative engagement, stable engagement, authentic engagement, and finally sustained engaged relationships. She argues that each stage is vital to "our understanding of why people choose to participate in city-campus partnerships and how people and partnerships, as a result, evolve" (p. 79).

In *Challenging Knowledge*, Delanty (2001), too, argues for a new role and identity for the contemporary academic institution in the contemporary landscape of knowledge democratization. He states that since the university is no longer the only social institution responsible for the "reproduction of instrumental/technical knowledge,"

and therefore no longer the "codifier of a now fragmented national culture," the university can, and should, ally itself to civic society instead (p. 6). Borrowing Habermas's epistemological terms, Delanty calls for the university to "recover the public space of discourse that has been lost in the decline of the public sphere" and relink "knowledge and human interests" (p. 7). In other words, by reinventing their identity, universities can become sites where knowledge society and civic society can intersect. In this view, digital humanities projects, due to their mission and methods, are in fact well positioned to enable this kind of connectivity. I would add also that they could combine technical capacity with humanistic consciousness to foster the technical and cultural forms of citizenship for which Delanty calls in his book.

3.3 Why engage?

Practically, what might be some reasons for the academy to engage with its various communities? More specifically, why should the humanities be concerned with engagement at all, considering research and teaching are the primary goals upon which reward structures in the academic systems are based? In the historical context, the necessity for humanities to pursue public engagement emerged out of the so-called "culture wars" of the late 1980s and early 1990s, during which time, "neoconservative intellectuals, journalistic and academic, stage well-funded, reactionary attacks on humanities scholarship and curricula that revise Western canons, national narratives, and the relations among culture, society, and politics" (Bartha, p. 92). The 1990s saw the return of the plutocratic view of wealth as a societal good that does its best for society when endowed in private foundations and stewarded by the wealthy (Khoo, p. 24). Berman, for example, argues that the so-called crisis of the humanities "has very much to do with a problematic relationship to the public" and it lends itself well for analysis "in terms of the structure of the public sphere" (2012, p. 174). In many ways, this neoconservative perspective on the role of knowledge in contemporary society is precisely what Kit Dobson is responding to in his chapter of *Retooling the Humanities*, where he argues that the purpose of research is not solely to drive practical application of knowledge into products or services (2011). The relationship between public engagement

and the digital humanities is therefore crucial to understanding community-engaged scholarship.

The National Coordinating Centre for Public Engagement (NCCPE) provides a general definition of public engagement, which is applied across academia or higher education:

> Public engagement brings research and higher education institutions together with the public. It generates mutual benefit – with all parties learning from each other through sharing knowledge, expertise and skills. Done well, it builds trust, understanding and collaboration, and increases the institution's relevance to, and impact on, civil society (NCCPE, 2009).

Clearly, this understanding of public engagement framed in terms of relevance and impact reflects the constructivist theoretical approach to knowledge production. Other definitions reflect other facets of the concept, such as civic, public and social aspects of the community. For example, community engagement can be understood as "an activity where a combination of faculty, staff, students, and community members work collectively to address important, and sometimes urgent, societal problems that arise out of daily community life" (Rosaen, Foster-Fishman & Fear, 2001, p. 10). Ehrlich (2000), on the other hand, views it as the act of working "to make a difference in the civic life of our communities and developing the combination of knowledge, skills, values and motivation to make that difference. It means promoting the quality of life in a community, through both political and non-political processes" (p. *vi*). Finally, Kimball and Thomas (2012) argue that all types of community engagement, whether they are participatory and reciprocal or technocratic and linear, represent "a kind of place-building practice whose outcomes – economic and social relations, ethical conduct, construction and treatment of built and natural environments – embody a set of intrinsic beliefs and values motivating engagement strategies" (p. 19).

There are, of course, challenges with institutional commitment to community-engaged scholarship. According to Onciul (2013), for example, despite the positive assumptions it may generate, community engagement "has the potential to be both beneficial and detrimental" (p. 79). One of its detriments is the fact that it does not lend itself easily to quantifiable measurements, as easily as well-established metrics such as publication, citation counts and course instruction. Another challenge is that this type of

work is not "deemed to be core mandate[s]" (Schuetze, p. 71). For example, Lockwood (2011) emphasizes that many universities still view service as "largely unimportant and contrary to the advancement of knowledge through scientific research, which they see as the true purpose of higher education" (p. 89). After all, within reward and incentive structures of contemporary research universities, resources are arguably scarcer for "less prestigious outreach and service activities, especially when there is comparatively little recognition or other, more palpable benefits in return" (Schuetze, p. 66). As a result, this belief runs counter to the reinforced norms of academia where "autonomy reigns, a clear distinction is maintained between the researcher and research subjects, and community engagement as scholarship is either not valued, or only legitimized by inclusion in university tenure and promotion systems" (Lockwood, Lockwood, Krajewski-Jaime & Wiencek, p. 88).

3.4 Participation

Logically, the possibility of public humanities relies on the idea of participation: the ability of the public to take part in the process of knowledge production. The concept of participation stems largely from civic and political contexts in liberal societies, but it manifests itself in the digital networked sphere in practices such as crowdsourcing, social media use, and communication through platforms such as message boards, chat and interactive comments on news websites. In other words, the concept of participation is present in all areas of digital information networks, and I believe it is crucial to examine its role in the discourse of public engagement in the digital humanities as well.

It is useful to think about the notion of civic participation in terms of power. After all, historical research by Habermas, Foucault and others indicates that

> "power in non-authoritarian post-feudal societies and states is dispersed, and that the emergence of the public sphere is linked to the opening up of a political space not directly subordinate to state authority or economic function, thus reflecting the potential pluralism embedded in modern polycentrism" (Schecter, p. 35).

The principle of universal participation, by extending access to formerly marginalized groups, implied emancipatory action for the members of the public. However, the growth of increasingly commercial mass media had an effect on shaping the public, even as "voting and other political rights were extended to previously disenfranchised groups, expanding participation in public life, political debate in a commercialized public sphere lost its independent critical edge and became more sensationalized and trivialized" (Benson, 2009, p. 177).

Arnstein, a key scholar in this area, argues that citizen participation is the redistribution of power that enables the "have-not citizens, presently excluded from the political and economic processes, to be deliberately included in the future" (1969, p. 216). Ultimately, this understanding of participation is fundamental to the function of liberal democracies. Generally, more participation is considered to be better than less, and we see this theme played out in Arstein's conceptual framework of participation. This framework is articulated in terms of a ladder, establishing a hierarchy of concepts, as evident in Figure 2 below. Notice that the levels of the ladder fall under three main categories, representing various degrees of access to power: non-participation, tokenism and citizen power. From this model, I adopt the concept of a range or degrees of participation, by which I mean that participation is exercised to various extents rather than being completely present or completely absent. The concept of degrees of participation is useful in articulating the relationship between the academic institution carrying out the work of data curation and the broader community of users of that data.

Figure 2: Arnstein's Ladder of Participation (1969, p. 217)

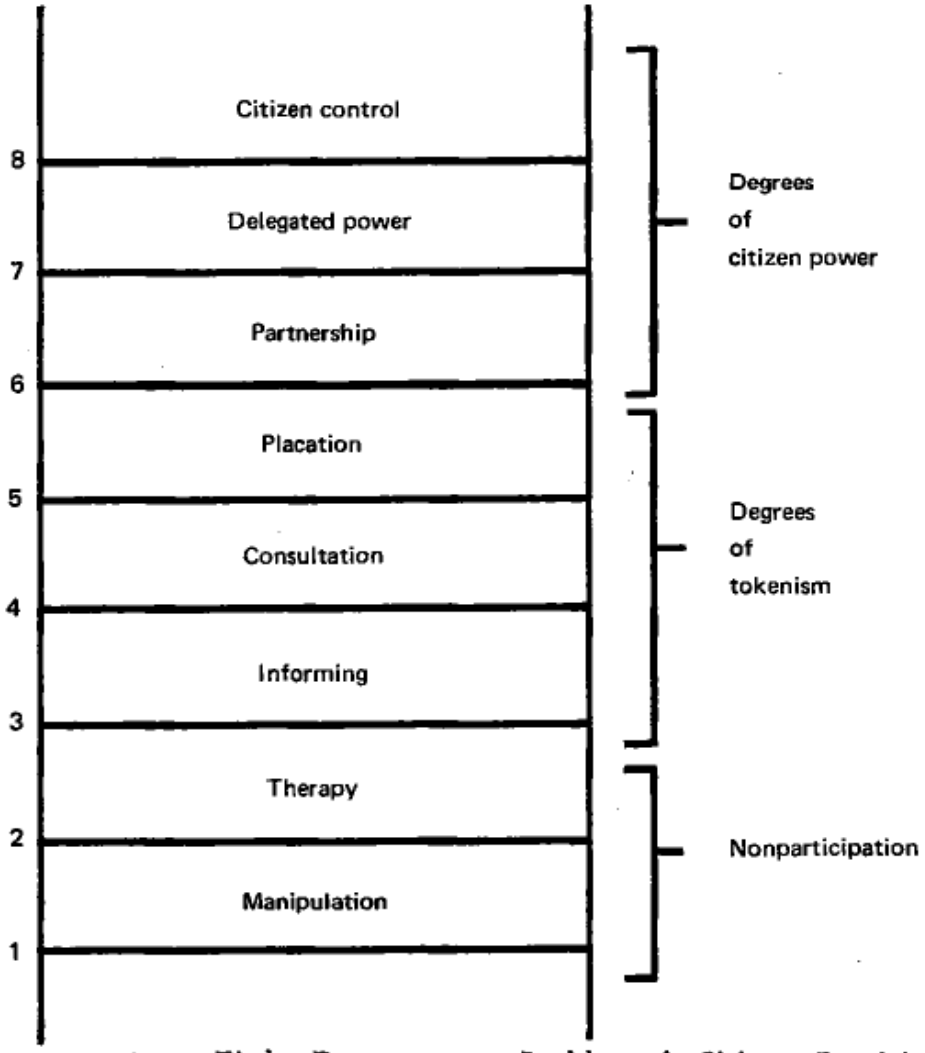


FIGURE 2 *Eight Rungs on a Ladder of Citizen Participation*

Building on Arnstein's Ladder of Participation, Wilson and Wilde propose an updated concept in their Dimensions of Community Participation (2003), which can be organized as demonstrated by Table 1 below.

Table 1: Wilson and Wilde's Dimensions of Community Participation

Influence How partnerships involve communities in the 'shaping' of regeneration plans/activities and in all decision-making.	**Inclusivity** How partnerships ensure all groups and interests in the community can participate, and the ways in which inequality is addressed.
Communication How partnerships develop effective ways of sharing information with communities and clear procedures that maximize community participation.	**Capacity** How partnerships provide the resources required by communities to participate and support both local people and those from partner agencies to develop their understanding, knowledge and skills.

They touch on such notions as agency of public players, relationships among them, social cohesion and exchange, all of which are dimensions of community engagement.

Related to the notion of participation and power distribution is Enos and Morton's (2003) notion of transactional and transformative relationships exhibited by institutions when engaged in community work. These notions are outlined in Table 2 below. Each type of relationship is appropriate and effective for specific purposes. From this perspective, data curation projects in the digital humanities represent both types of relationships, where neither is necessarily better than the other, but frames the goals and results of the relationship in a different manner.

Table 2: Transactional and transformative relationships

Criteria	**Transactional**	**Transformative**
Basis of relationship	Exchange-based and utilitarian	Focus on ends beyond utilitarian
End goal	Satisfaction with exchange	Mutual increase in aspirations
Purpose	Satisfaction of immediate needs	Arouses needs to create larger meaning
Roles played by partners	Managers	Leaders
Boundaries	Accepts institutional goals	Examines institutional goals
Support of existing institutional goals	Works within systems to satisfy interests and partners	Transcends self-interests to create larger meaning
Partner identity	Maintains institutional identity	Changes group identity and
Scope of commitment	Limited time, resources, and	Engages whole institutions and potentially unlimited exchanges

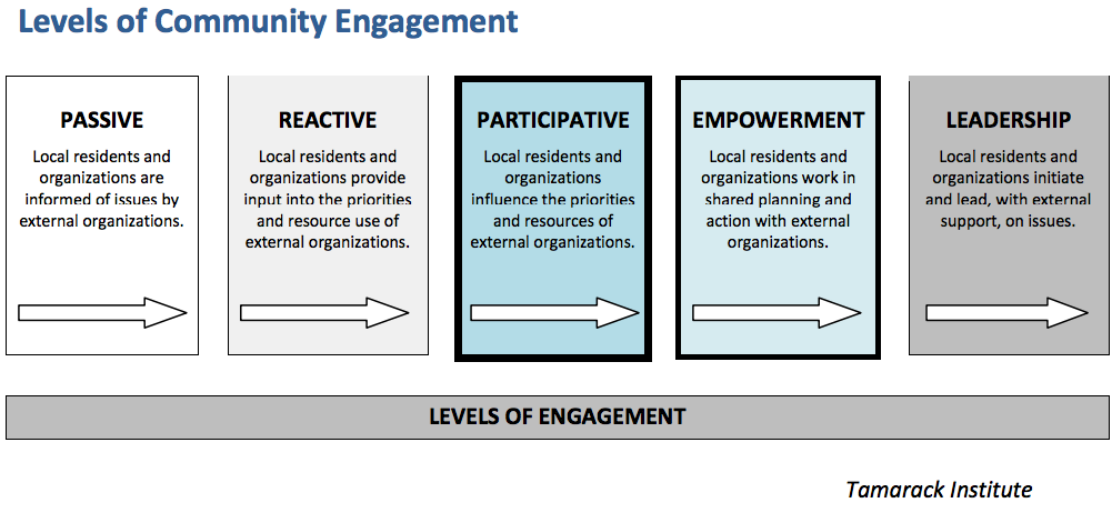
The criteria identified by Enos and Morton in this table are useful in evaluating digital humanities projects, as they also involve communities of interest as a user group for their data.

Further, in his taxonomy of 'emerging practices' of community-engaged scholarship, Barker (2004) sees the concept of civic engagement as much more than mere service to the community that implies a passive knowledge transfer relationship between the academy and the public. Instead, it is a set of processes that involve the communities and the wider public in the production of knowledge itself (Khoo, p. 26). Once again, community engagement is tied to the larger civic mission of the university as an institution fundamental to the function of a civic society.

These conceptual frameworks thereby get transformed into service philosophies, policy and strategic operations in public institutions, such as libraries, museums and universities. For example, in 2010, the Edmonton Public Library announced its commitment to the Community-Led Service Philosophy as an institutional mission. The

rationale for this service orientation is defined in its Community-Led Service Philosophy Toolkit document (2010, 2013). The Toolkit includes the following diagram outlining the various levels of community engagement in library service philosophies. Much like Arnstein's ladder of participation, which establishes degree of power given to the community to affect the institution, the Figure 3 below demonstrates various levels of engagement a community can have with an institution, such as a public library. In the context of Tamarack Institute's model, a digital humanities project may allow its community of users to lead all aspects of the project and therefore exhibit a leadership level of engagement, or, borrowing Arnstein's terms, an institution may inform the community through passive level of engagement. In the context of the Edmonton Public Library, for example, the diagram highlights the middle participative level of engagement adopted by the organization as part of its overall service model.

Figure 3: Tamarack Institute's Levels of Community Engagement



Finally, it might be useful to include in this discussion the theory of participatory culture, as developed by Jenkins et al (2006; 2010), which is defined according to the following criteria (2009, p. xi):

- low barriers to entry
- strong support for creation and sharing with others
- informal membership where experience is passed along to newbies
- members believe their contributions matter

- members feel some sense of social connection

Key concepts of this theory include: affiliation and membership, expressions and productions, collaborative problem solving and circulations of media (p. 9). Participatory culture is about jumping into learning and creation, establishing a sense of community, however informal it may be, as well as sharing knowledge through these communities. Whether it's fan fiction websites, massive online multi-player gaming communities or dedicated makerspaces in public libraries, participatory culture is thriving in the digital age, which is why Jenkins' theory is particularly effective in this discussion.

A growing body of research suggests that participatory culture provides multiple "opportunities for peer-to-peer learning, a changed attitude toward intellectual property, the diversification of cultural expression, the development of skills valued in the modern workplace, and a more empowered conception of citizenship" (p. xii). As a result, it enables the possibility of an engaged culture, though it does not determine it. Additionally, engagement in social media environments specifically has been argued to provide more opportunities for "self-representation, expression, reflection and more organized forms of collaboration and knowledge-building" (Warwick, Terras & Nyhan, 2012, p. 25).

I would like to relate the theory of participatory culture in the digital age to the socio-political aspects of civic participation. I would like to distinguish it from the interactive elements of digital technologies, however, because as Jenkins et al. argue, interactivity is "a property of technology, while participation is a property of culture" (2009, p. 8). In fact, participatory culture is emerging "as the culture absorbs and responds to the explosion of new media technologies" that allow individuals to consume, extend, repurpose and recirculate various forms of digital content in a new manner. The use of the concept of participation in Jenkins' theory "cuts across educational practices, creative processes, community life, and democratic citizenship" (p. 9) and is therefore effective in examining digital humanities scholarship for the purposes of this thesis.

Chapter 3: Research Design

Introduction

In designing my research, I chose a critical approach and method of inquiry to suit the exploratory nature of my thesis. I did not begin this work with a specific question about data curation in mind and sought, instead, to survey the field of data curation in the digital humanities in order to draw certain lessons from an evidence-based position. As a result, I decided upon a grounded theory approach because it permitted a flexibility to collect my data and analyze it without a specific theoretical lens. Nevertheless, I also relied on several theories related to the concepts of participation, power and community engagement to orient the themes that emerged from the data I gathered. The following chapter is organized into sections outlining my methodology and method, data collection techniques, data analysis procedures, as well as the discussion of the study's reliability and validity.

Methodology

This study uses an interpretive theoretical perspective, and its methodological approach can be best described as a grounded theory. It relies on an interpretivist paradigm, meaning that my ontological perspective is relativist and my epistemological commitment is subjectivist. In other words, I approach this study with the perspective that knowledge about the nature of social reality such as pursuit and organization of knowledge in the digital humanities is constructed intersubjectively "through the meanings and understandings developed socially and experientially" (Cohen & Crabtree, 2006).

My conceptual framework relies heavily on the notions of participation as an element of power in a relationship that is established between the researcher and the community. I used Arnstein's notion of participation as an expression of power as a foundational concept in my theoretical orientation in this thesis, because I found the notion of the degrees of participation – from non-participation to tokenism to citizen power – an extremely useful conceptual tool with which to approach my study. I also saw the idea of varied degrees of power in other theories, which also suggested that Arnstein

influenced many scholars in the field social science. For instance, the Tamarack Institute's levels of community engagement echoed a notion of progression in community relationship to an institution. Wilson and Wilde's notions of capacity and inclusivity further contributed to my burgeoning theoretical framework, as well as Enos and Morton's relationship types that exist between communities and institutions. Finally, Jenkins' participatory culture theory allowed me to view community-engaged data curation from the perspective of informal community membership and transfer or knowledge. All of these aspects of the theories covered in the previous chapter contributed to my theoretical approach to analyze and interpret the data collected in my sample population.

Fundamentally, I approach this study from a pragmatist perspective, with the understanding that knowledge exists to solve human problems, where the nature of knowledge, concepts, meanings, language and science are best viewed in terms of their practical use in solving human problems. Furthermore, from this theoretical orientation, data curation does not exist for the intellectual pure purpose of knowledge itself. Rather, in its focus to make knowledge discoverable, accessible and usable, data curation aligns well with a pragmatist perspective that human knowledge exists *for* other humans, to solve human problems and make the human world better. The inherent assumption of such a view is that preservation and access to knowledge is fundamentally beneficial to civic society.

From this theoretical stance, I employed the grounded theory approach to conduct my study. As Leedy and Omrod (2010) explain, grounded theory uses a "prescribed set of procedures for analyzing data and constructing a theoretical model from them" (p. 142). In this sense, the emerging theory of a particular process or behaviour being studied is derived from, and therefore, grounded in the data rather than taken from the research literature on the topic. In a typical grounded theory approach to research, data collection begins almost immediately, "at which point the researcher develops categories to classify the data" (Leedy & Omrod, p. 142). As data collection continues, the researcher aims to saturate the categories and learn as much as possible about various elements of the topic and obtain any findings that suggest that the categories need to be revised. While experts disagree about the most effective way to analyze the data (Charmaz, 2006; Corbin &

Strauss, 2008; Glasser, 1992), an approach proposed by Corbin and Strauss (1990; 2008) generally follows these stages:

1. Open coding – data are divided into segments and scrutinized for commonalities. Once the data are categorized, they are then examined for properties that characterize those categories.
2. Axial coding – interconnections are made among categories and subcategories.
3. Selective coding – the categories and their relationships are combined to form a narrative that describes the nature of the phenomenon being studied.
4. Development of a theory – a theory is offered to explain the phenomenon being studied. It may take place in the form of a verbal series of statements, visual model or series of hypotheses.

As Urquhart and Fernandez explain (2013), grounded theory was created as a general method with no explicit correct epistemology as a starting point for the research. This degree of epistemological neutrality makes grounded theory a highly useable research method, but according to Urquhart and Fernandez, qualitative researchers "have a responsibility to make their epistemological position clear, conduct their research in a manner consistent with that position, and present their findings in a way that allows them to be evaluated appropriately" (p. 229). Accordingly, Glasser and Strauss (1967) argue that researchers should not approach reality as a tabula rasa. Following this logic, I acknowledge my own biases when approaching this study, and rely on a theoretical perspective that helps me formulate abstractions from the data. The literature review and critical context covered in the previous chapter of this study helps define the problem domain (data curation in the digital humanities) as well as the theoretical sensitivity to approach this domain. Urquhart and Fernandez define theoretical sensitivity as the ability to understand the observations in a wider theoretical context (p. 230). While I began this study with no particular theory in mind, since few theories surrounding data curation in the digital humanities exist, I nevertheless rely on research literature on digital

humanities, community engagement and participation as critical concepts to help me make sense of the observations collected as part of this study.

Data collection techniques

In order to obtain a substantive sample of data to analyze, I created a list of identified institutions engaged in digital humanities by scanning relevant DH blogs, CenterNet – an international network of digital humanities centers – and Twitter. I also consulted with librarians working in data curation and with digital humanists. Many research universities across North America have established digital humanities centers, labs and dedicated units, and they were examined first through such methods as typing "University of California Digital Humanities" into a search engine. Not all DH centers have clearly organized websites, however, suggesting that the work carried out by members of those centers is distributed or perhaps experiential in nature with little data to share. For example, gaming and multimedia research is a growing area in the digital humanities, with certain projects like Lev Manovich's "Mining a Million Manga Images" and the gaming studies at the University of Alberta getting press recognition. However, the goals of such projects are not necessarily knowledge preservation, access and usability, as is the case in many data curation projects in the digital humanities. Instead, gaming and multimedia research is concerned with experiential and reflective modes of inquiry. As a result, projects of this nature were not selected for the study sample. For instance, the Massachusetts Institute of Technology has an established DH center called HyperStudio, yet I was unable to find any projects with unique websites and containing humanities data related to the HyperStudio from which to collect data for this study. Similarly, City University of New York's website for its DH center lacked any detail sufficient to conduct a data collection procedure for this study. I sought, instead, projects that offered potential lessons in data organization, community-engagement and/or innovative methods in the digital humanities. My main criteria for selection was a substantive body of data – in a variety of formats – as well as indication of reflective approaches to humanities research methods or evidence of engagement with broader communities of users of that data. For example, University of Victoria's MakerLab is very much an experientially-oriented project in the digital humanities with a focus on

pedagogy and physical computing, but it lacks an explicit set of data upon which to conduct analysis. While its website did not provide organized, curated data beyond project blog entries, it nevertheless offered potential observations about its capacity to engage the academic and public community of greater Victoria through social media and other means. I included this particular project in my study population for other aspects of evaluation, such as observations in digital humanities and community engagement components.

I consulted a list created by the Carolina Digital Humanities Initiative (2014) that features many digital humanities projects organized by discipline from various universities across the United States. I deemed this list extremely useful for selecting my study population, as it provided a range of disciplinary backgrounds in digital scholarship and projects of which I was unaware. One project was selected from each of the following categories included in the list: Archaeology, Arts, Classical Studies, Geography & Urban Studies, History, and Literature & Languages. I then browsed through the Carolina DH Initiative list to select one that was either representative of that Digital Humanities center, had a data curation component, or has its unique website outlining the details of the project. Five Canadian institutions were selected to provide balanced coverage of the range of digital humanities projects taking place in North America. Additionally, two projects carried out by publicly funded institutions and three projects carried out by non-profit organizations were selected, since data curation takes place of the academy. My intention was that inclusion of projects from a variety of geographic locations, disciplinary backgrounds and institutional affiliations, a more balanced representation of data curation in the digital humanities might emerge from the data gathered.

As a result of this broad search, twenty-eight digital humanities projects were selected as a study population. These projects represent a diversity of scholarship taking place in this field and therefore belong to a variety of disciplines in the humanities, including but not limited to literary studies, history and classics, modern languages, philosophy, art history, religious studies, or cultural and media studies. The selection of projects aims to capture a diversity of target user groups for the subjects of study, themes

and subject areas, content and data types, as well as geographical origin across North America. These project include:

1. New York Public Library Labs: **Building Inspector -** http://buildinginspector.nypl.org/
2. Scholar's Lab, University of Virginia: **Salem Witch Trials -** http://salem.lib.virginia.edu/
3. Maryland Institute for Technology in the Humanities: **Shelley Goodwin Archive** - http://shelleygodwinarchive.org/
4. Carolina Digital Humanities Initiative, University of North Carolina at Chapel Hill: **Digital Portobelo** - http://digitalportobelo.org/
5. Center for Digital Scholarship, Brown University – **Underground Rhode Island** - http://cds.library.brown.edu/projects/undergroundri/
6. McGill Center for Digital Humanities: **Early Modern Conversions -** http://earlymodernconversions.com/
7. University of Alberta: **Edmonton Pipelines -** http://edmontonpipelines.org/
8. University of California, Berkley: **Free Speech Movement Digital Archive**- http://bancroft.berkeley.edu/FSM/
9. Initiative for Digital Humanities, Media, and Culture, Texas A & M University: **Early Modern OCR Project** - http://emop.tamu.edu/
10. Emory Center for Digital Scholarship, Emory University – **Lincoln Logarithms** - http://disc.library.emory.edu/lincoln/
11. Stanford Humanities Lab: **Republic of Letters -** http://republicofletters.stanford.edu/
12. Center for Digital Research in the Humanities, University of Nebraska-Lincoln: **Omaha & Ponca Digital Dictionary** - http://omahaponca.unl.edu/
13. Center for Digital Humanities, Ryerson University: **Yellow Nineties -** http://1890s.ca/
14. **HistoryPin -** https://www.historypin.org/
15. Brooklyn Brainery: **A Handsome Atlas** - http://www.handsomeatlas.com/
16. University of Pennsylvania: **PhilaPlace** – http://www.philaplace.org/

17. University of California, Los Angeles: **Hypercities LA -** http://hypercities.com/LA/
18. Northeastern University: **Our Marathon -** http://marathon.neu.edu/about
19. Indiana University Center for Digital Arts and Humanities: **EVIA -** http://www.eviada.org/default.cfm
20. Yale University: **Historian's Eye** - http://historianseye.commons.yale.edu/
21. University of Washington: **Women Who Rock** - http://womenwhorockcommunity.org/
22. University of Toronto: **Culinaria** - https://utsc.utoronto.ca/digitalscholarship/culinaria/
23. Pratt Institute: **Linked Jazz** - https://linkedjazz.org
24. University of Arizona: **Digital Augustan Rome** - http://digitalaugustanrome.org/
25. University of Iowa – **Mapping Decline: St. Louis and the American City** - http://mappingdecline.lib.uiowa.edu/
26. Khan Academy: **Smarthistory** - http://smarthistory.khanacademy.org/
27. University of Nebraska-Lincoln: **History Harvest** - http://historyharvest.unl.edu/
28. University of Victoria: **Maker Lab in the Humanities** - http://maker.uvic.ca/

Data collection techniques consisted of observations in the form of detailed notes. Initially, I collected free-form notes about five projects, noting presence of any conceptual category on the project site, essentially describing by impressions of what the project aimed to do. Following a sample of 4 projects with free-form, unstructured observation notes, I then collected data in a more systematic fashion, using consistent vocabulary, such as "no evidence of data sharing" or "no unique social media presence" to help identify trends and patterns for analysis.

Following an approach typical in a grounded theory study, a data collection sample was gathered from one large multi-institutional and one small academic project, as well as one public and one non-profit project. Based on the detailed notes outlining issues related to the topics of data curation, digital humanities and community engagement, I further expanded my evaluation criteria for my study population. The criteria I constructed reflect the focus on three main areas of this research project:

analysis of data curation practice, digital humanities methodologies and community-engagement aspects of various digital projects taking place in the humanities across North America. As such, the evaluation criteria developed through the data collection pilot aimed to allow me to gather as much meaningful data about data curation work in the digital humanities as possible.

The data gathered for this study were collected in Edmonton, Alberta between November 1 and December 5, 2014. This study's findings are based on data gathered only from the public-facing websites of the projects sampled, scanning any social media posts shared by the projects. The findings are not based on any extra information obtained from such sources as interviews with project directors, project websites requiring registration or access to the project's internal statistics and analytics. As a result, some of the project directors may share information about projects evaluated in this study in research articles or presentations that are not publically available on their websites, which may create a discrepancy between numbers and statements recorded in this study, but since I have no access to internal information of these digital projects, all data gathered is subject to the changing nature of the web. In fact, since collecting the data, I noticed that the Khan Academy changed the layout for Smarthistory's website around December 4, 2014, adding some additional features not reflected in the study.

Data analysis procedures

I used the constant comparative method in data analysis and coding procedures in this study. As Urquhart and Fernandez explain (2013), constant comparison is the driving technique of data analysis in grounded theory methodology as well as the facilitator of theoretical sampling and therefore the means to reach what Glaser and Strauss (1967) term theoretical saturation – "the point at which data gathering stops and substantive grounded theory begins to emerge" (p. 225). They emphasize, however, that while following grounded theory coding procedures is necessary, "slavish adherence to those procedures is not on its own sufficient to produce good theoretical outcomes" (p. 225). To this extent, my coding procedures included tracking recurring themes such as use of the same terms both on the project websites and in the data they shared. For example, while I started with a vague notion of "policy and standard" evident in most projects

examined through presence of privacy policies, use of specific metadata schemes or licenses governing data sharing and reuse on the sites, I disambiguated the broad concept into specific themes of copyright statements, metadata standards and licenses. Most of the labels used in my coding procedures were therefore derived from the data itself, often using the language found on the websites describing digital humanities projects.

Reliability and validity

Reliability and validity of data and the generalizability of results of this study can be difficult to evaluate. Given that there are thousands of digital humanities projects taking place across academic, public and non-profit institutions across North America, it is unreasonable to expect that this study's sample size of 28 projects will be generalizable to all data curation projects in the digital humanities. However, the themes derived from the data collected as part of this study are likely to be present in any digital humanities project that has data. I believe that the categories and criteria outlined above are useful in this area of research, and likely reflect key principles in data curation. As a result, despite the small population size, one way to test the reliability and validity of the study would be to sample a wider range of projects, including increasing the sample size to 100 projects, representing more institutional backgrounds and even extending the geographic range to Europe and other parts of the world. Another way to account for possible bias in the selection of the study population would be to make a list of 300 digital humanities project based on the key criterion of those projects having data (rather than being projects of experiential or procedural nature such as gaming or philosophy of software) and select a sample of those projects at random. Overall, my findings are generalizable in that they draw upon theories and concepts established in related fields, such as those of community engagement, participation, power, for example. They also have the possibility to inform practice of conducting digital humanities projects involving aspects of data curation. As mentioned in the Research Problem portion of this study, my findings make a contribution to a general understanding on the emerging role of community engagement in the field of data curation, and propose useful criteria that can be applied to future practical situations.

One way to check the validity of my findings consisted of making connections to theories and research on data curation, digital humanities and community engagement. Doing so situates the study in a wider context of knowledge on these topics as well as accounts for potential errors in judgment that might occur on the part of the study investigator. Similarly, a way to check the credibility of my study includes measuring persistent observation, which Bitsch (2005) defines as "whether the researcher [has] done an in-depth study to gain detail" (p. 83) In other words, the credibility of results can be measured by evaluating extent to which persistent observation has been used to identify the most relevant characteristics of the problem. I would argue that the comparative nature of my analysis, and the construction of multiple specific criteria in each major category of my model of data curation in the digital humanities address this measurement. I have attempted to capture as much detail and depth in each aspect of major categories observed in the data to check for potential errors in credibility of results. Finally, one way to check for reliability of my findings consisted of explication where I explain and document my interpretation of the data by providing examples. This method also includes quoting directly from the observations collected from the project websites as well as including screenshots demonstrating concepts reflected in my conclusions. As a result, the Findings and Analysis chapter includes quotations taken from project websites represented in *italics* and screenshots labeled as numbered figures.

Chapter 4: Findings

Introduction

As the purpose of a grounded theory approach is to seek general theories, patterns and trends, I made a conscious decision not to identify individual projects when evaluating them based on specific criteria. Details from individual projects are used as examples, but individual projects are not discussed for their unique properties or compared to other individual objects. Recognizing that some identified thematic categories listed in this section feature only one or two projects with identified criteria, it will therefore be possible to interpret to which specific project I am referring in the discussion. Nonetheless, it is not my intention to isolate some projects over others, nor to suggest that some are better than others. However, for those interested in learning about the specific projects, a Venn diagram mapping the placement of all 28 projects on the conceptual map constructed based on the findings of this is included as an appendix to this thesis. A digital version of the complete data set collected over the course of this study has been uploaded and can be accessed at http://hdl.handle.net/10402/era.38716.

Overview of projects studied

The following section provides a general overview of the variety of projects represented in this study based on their institutional affiliation, geographic location and disciplinary background. Figure 4 below demonstrates the range of host institutions to which projects in this study belong. Academic here refers to publically and privately funded higher education institutions like universities, colleges, and research institutes. Public here refers to publically funded institutions, such as public libraries, historical societies, and archives. Non-profit refers to neither publically funded nor higher education academic institutions, but private, non-profit organizations that share the values such as education and access to information. Table 3 further outlines the types of institutions included in the analysis along with names of individual organizations. While academic institutions such as universities and colleges featured prominently in this study, my attempt to provide a balance analysis of data curation work in the digital humanities

is represented by the number and names of public and non-profit organizations that also demonstrate evidence of data curation projects.

Figure 4: Institutional backgrounds of projects studied

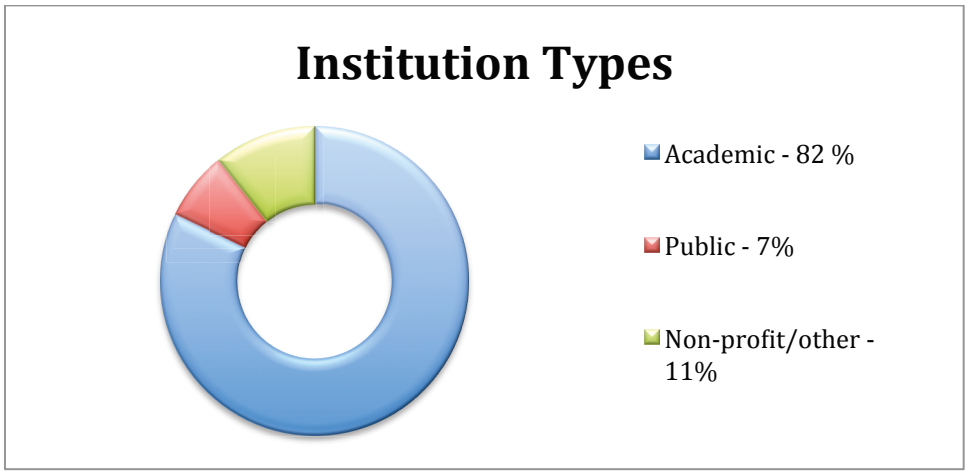
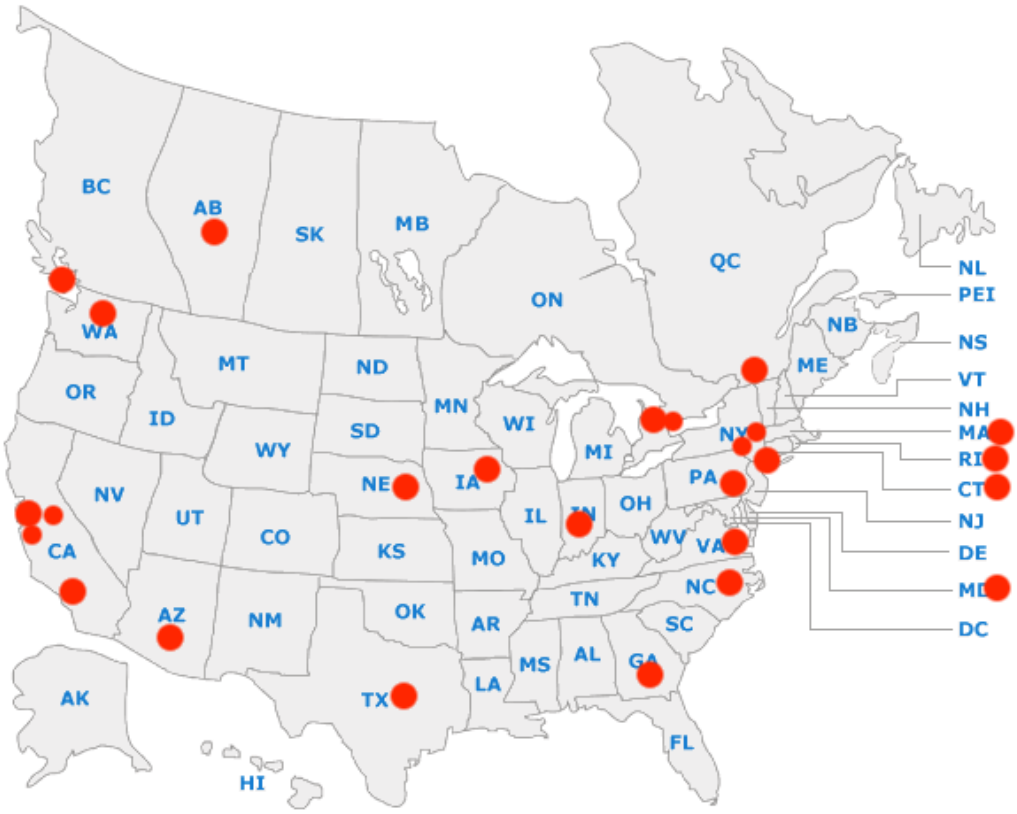


Table 3: Institutional backgrounds of project studied

Type of institution	**Number of projects**	**Institution type**	**Host institution names**
Public	2	Public library, historical society	NYPL Labs, Pennsylvania Historical Society
Non-profit	3	Private non-profit company, educational organization	Brooklyn Brainery, Shift, Khan Academy
Academic	23	Universities, research institutes	Brown University, University of Arizona, University of Alberta, University of California -Berkeley, University of California - Los Angeles University of Maryland, University of Virginia, University of Iowa, University of North Carolina at Chapel Hill, McGill University, Texas A & M University, Emory University, Ryerson University, University of Nebraska-Lincoln, Stanford University, Northeastern University, University of Indiana, Pratt Institute, University of Toronto, University of Victoria, University of Washington, Yale University

To help articulate the distribution of projects included as part of this study, I created a map with each point representing an institution or organization from my list of project, as demonstrated by Figure 5 below.

Figure 5: Geographic distribution of projects studied



The interdisciplinary nature of digital humanities and data curation work means collaborations between units within universities as well as among institutions, often around the world. While many of the projects sampled in this study include content contributed from both American and Canadian institutions, the specific inclusion of projects lead by Canadian researchers was deemed important to representation of diversity of digital scholarship across North America. For the purposes of the study, each project is organized according to where its headquarters are located. Often, this is the location of the team that initiated the project. For example, Early Modern Conversions is

a sprawling multi-university effort spanning researchers, graduate students, partners and affiliates from Canada, United States and Europe, but the project is led by McGill University and is included in the list of Canadian institutions in this study. The following tables demonstrates the range of geographic locations included in this study's population

Table 4: Geographic distribution of projects studied

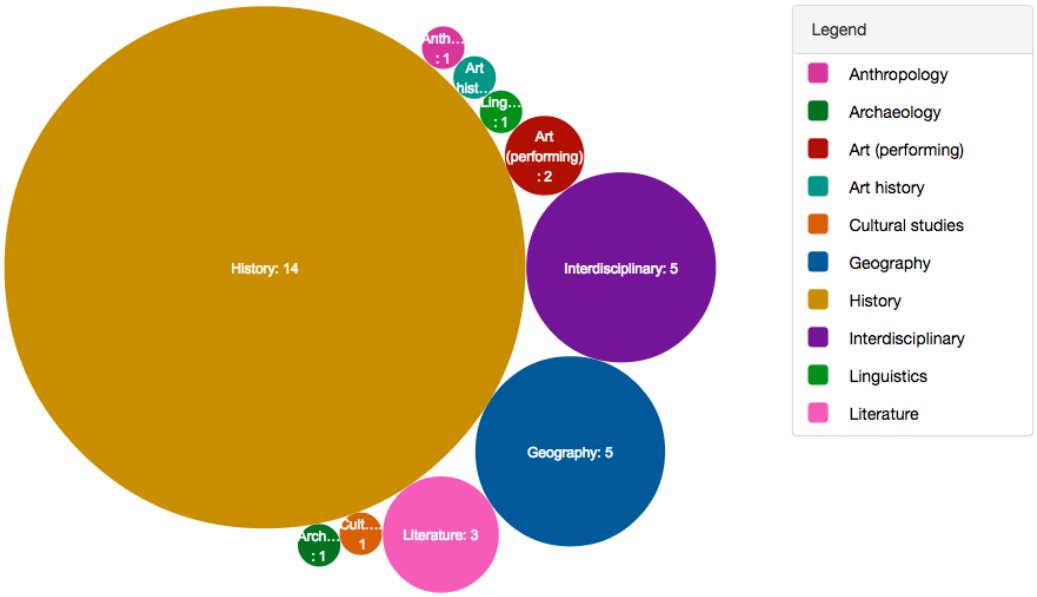
Country of project's headquarters	**Number of projects**	**State or province of host institution**	**Host institution name**
Canada	5	Ontario, Quebec, Alberta, British Columbia	University of Toronto, McGill University, University of Alberta, Ryerson University, University of Victoria
United States	23	Arizona, California, Connecticut, Georgia, Maryland, Massachusetts, Rhode Island, Indiana, Virginia, North Carolina, Texas, Washington, Iowa, Nebraska, Pennsylvania, New York.	Brown University, University of Arizona, University of California -Berkeley, University of California - Los Angeles University of Maryland, University of Virginia, University of Iowa, University of North Carolina at Chapel Hill, Texas A & M University, Emory University, University of Nebraska-Lincoln, Stanford University, Northeastern University, University of Indiana, Pratt Institute, University of Washington, Yale University

Table 5 and Figure 6 below aim to capture another broad facet of this study's population: the range of disciplines featured in the list of projects selected for analysis. They outline the names of disciplines I was able to identify from the project websites as well as the number of projects that feature those disciplines.

Table 5: Disciplinary breakdown of projects studied

Discipline	Sub-disciplines or specializations	Number of projects
Anthropology	Cultural	1
Archaeology	Ancient Roman	1
Art (performing)	Music, dance	2
Art history	World	1
Cultural studies	None identified	1
Geography	American, Canadian	5
History	American, Intellectual (History of Science), Early modern, European	14
Interdisciplinary	Modern and Ancient Languages, Philosophy, Political Science, Sociology, Religious Studies, Media Studies, Gender Studies, Musicology, Drama, Design, Computing Science, Cognitive Science	5
Literature	English, Modern Languages	3
Linguistics	North American Indigenous	1

Figure 6: Disciplinary breakdown of projects studied



At first glance, it appears that history projects dominate the data sample selected for this study, suggesting that the study investigator has a preference for projects of historical nature. However, upon further reflection, the high number of historical projects in the study sample also reflects the disciplinary environment of digital humanities. History research, along with literary studies, tends to be more text-oriented in terms of its primary data sources. As a result, more data in the discipline would be available to study, and would thereby generate more research projects. It is likely that since more data is available to access, more text-processing and mapping tools have been developed, shared and reused by history scholars, which enable further inquiry into the digital historical record. As well, since the majority of projects selected for this study are hosted by American institutions, they also benefit from a stronger archival and digitization infrastructure than those in Canada, for example. Arguably, American History documents are more likely to be preserved, digitized, transcribed and shared in open digital collections, which would allow greater use of those collections by researchers. Finally, historical archival material that becomes data for digital humanities research also tends to be in the public domain and therefore less restricted by copyright and other data use policies compared to other forms of data, such as contemporary art images, video, audio recordings and even literary texts. Admittedly, large volumes of literary heritage from around the world have been digitized and shared in such digital libraries like Project Gutenberg and HathiTrust. However, more contemporary literary texts remain out of reach for most digital humanists as they are still protected by copyright and/or have no usable digital versions of those data. In addition, it is possible that historians more so than scholars from other disciplines are inclined to organize, preserve and make accessible the digitized historical record. In this sense, data curation activities may already be part of their research process and thus, evidence of this curatorial activity may be reflected in this study's sample population.

1. Data curation themes

Given a general introduction to the variety of digital projects taking place in the humanities, the themes included in this section of the findings chapter aim to represent the nature of activities performed around the data lifecycle, from creation of humanities

data to its description and enrichment to long-term preservation, access and reuse. This multi-faceted examination of data curation trends addresses such elements as data size, types of data, use of metadata standards, data sharing and most common information architecture features and functionalities evident in the projects reviewed.

1.1 Data types

Despite the generalization that most digital humanists are heavily text-oriented in their work, the variety of data types observed in the population of this study suggests that humanists employ many data types and document formats, including a higher use of video than expected. As Table 6 below demonstrates, the web environment allows digital humanists to participate in a truly media-rich scholarly domain, being able to incorporate various data formats into a single project.

Table 6: Data types used in projects studied

Data type	**Used in number of projects**	**Percentage of all projects**
Text	28	100
image	27	96
GIS/map	9	32
video	12	43
audio	7	25
code	3	11
numeric	2	7
SMS	1	4

1.2 Content types

While formats used in digital projects reveal only the broad picture of what kind of data humanities are employing by capturing the physical media and dimensions of digital objects, a more detailed look into specific types of content or resources used in the project demonstrates the richness of information landscape in emerging digital scholarship. According to the Metadata Object Description Schema (MODS), a resource

type specifies “the characteristics and general type of content of the resource” (2014). The Schema defines these object types as follows:

- **Text** – resources that are basically textual in nature.
- **Cartographic** materials, including maps, atlases, globes, digital maps, and other cartographic items.
- **Sound recording** – includes a “mixture of musical and non-musical sound recordings occurs in a resource or when a user does not want to or cannot make a distinction between musical and non-musical”. Since I have not listened to all audio recordings samples in this study and cannot say for certain that all audio files are non-musical only, I am using this vocabulary to describe this type of audio content.
- **Still image** – including two-dimensional images and slides and transparencies.
- **Moving image** – items like motion pictures and video recordings, as well as television programs, digital video, and animated computer graphics—but not slides and transparencies. It does not include moving images that are primarily computer programs, such as computer games or computer-oriented multimedia; these are included in "software, multimedia".
- **Software, multimedia** is appropriate for “any electronic resource without a significant aspect that indicates one of the other type of resource categories. It includes: software, numeric data, computer-oriented multimedia, and online systems and services.”

Table 7 below illustrates the variety of types of content found in digital humanities projects examined in this study. Clearly, textual content is very common, but still images and moving images are also evident in digital humanities projects.

Table 7: Resource types in projects studied

MODS resource type	**Specific content types present in projects studied**	**Present in number of projects**
Text	Speeches, interviews, essays, literary works (novels, poems, drama), dictionaries, legal documents, forms, reports, software documentation, biographies, addresses, news	28
Cartographic material	Digital maps, scans of historic maps, atlases	9
Sound recording	Interviews, oral histories, micro history, speeches, testimonials	7
Still image	Scans of text, digitized photographs, digital photographs, posters, visual art	27
Moving image	Interviews, oral histories, traditional performance, music performance, testimonials	12
Software, multimedia	Code for software application	3

1. 3 Data sizes

The theme of data size proved problematic in the study population. Of the 28 projects studied, only 11 make any specific mention about data size used in the course of the project. Of those 11 that mention data size, the following variation in reference to data size is found in the data collected:

- 500,00 maps, 200,000 books and atlases digitized
- 350 paragraphs, 350 addresses plotted on a map
- 45 million pages of text scanned and processed
- 383,890 images, 59,671 users
- 3 maps, 3 census data sets plotted
- 200,000 images, 10,000 words
- 2,600 items total (images, video, paragraphs of text)
- 2,037 images used
- 605 videos, 310 articles/essays
- 1,206 tweets, 466 users, 278 photos
- 320 images, 1,472 texts, web websites

This type of variation makes it difficult to determine the scope of data size measurement. Should a data size measurement therefore be limited to a single item, such as a single image or video, a megabyte of data, a single word, a page of text, a paragraph? After all, a single "item", such as an atlas has many pages, all of which can be broken down and counted as individual data points, as they often are for digitization purposes. For the purposes of this study, I limited my definition of a digital humanities data "item" as a single page of text, a single image, a single digitized map, a single audio or video recording, a single address, and a single page of code. As a consequence, I have had to estimate the size of projects where no specific size was listed. For example, some project included multiple volumes of digitized magazines as part of their collection. Sampling a couple of such digitized objects, I recorded the average number of pages contained in each volume, and multiplied them by the number of volumes included in the collection, adding them together, as well as any other extraneous digitized material of multiple pages, such as posters, essays, reviews, biographies and bibliographies that made up the digital collection of the project. In other cases, while the project website did not list the total number of items in the digital collection, the same items (recordings of oral history interviews, for example) were houses in the institutional repository, which would give a total number of objects belonging to the collection by the name of the project. I would record this number as representative of the data size in that project, knowing that it alone did not necessarily reflect the total number of items involved in the project. Other material, such as critical essays or additional photographs describing the interviews were also part of the collection. While such estimates had high margin of error, they nonetheless helped separate projects dealing with a million pages of text, images or lines in excel files from those projects that really only included a thousand or so digital objects.

Based on this admittedly flawed approach, I created the following classification of data sizes present in the 28 projects studied. While the categories seem arbitrary initially, they generally follow a relationship to the size of the project. In other words, small projects led by one or two principal investigators typically deal with data between several hundred items (images, pages of text, audio recordings) to a couple thousand items. Furthermore, this amount of data is possible to process without use of data analysis tools. As a result, the small data category was capped at 5,000 items. In contrast, big data has

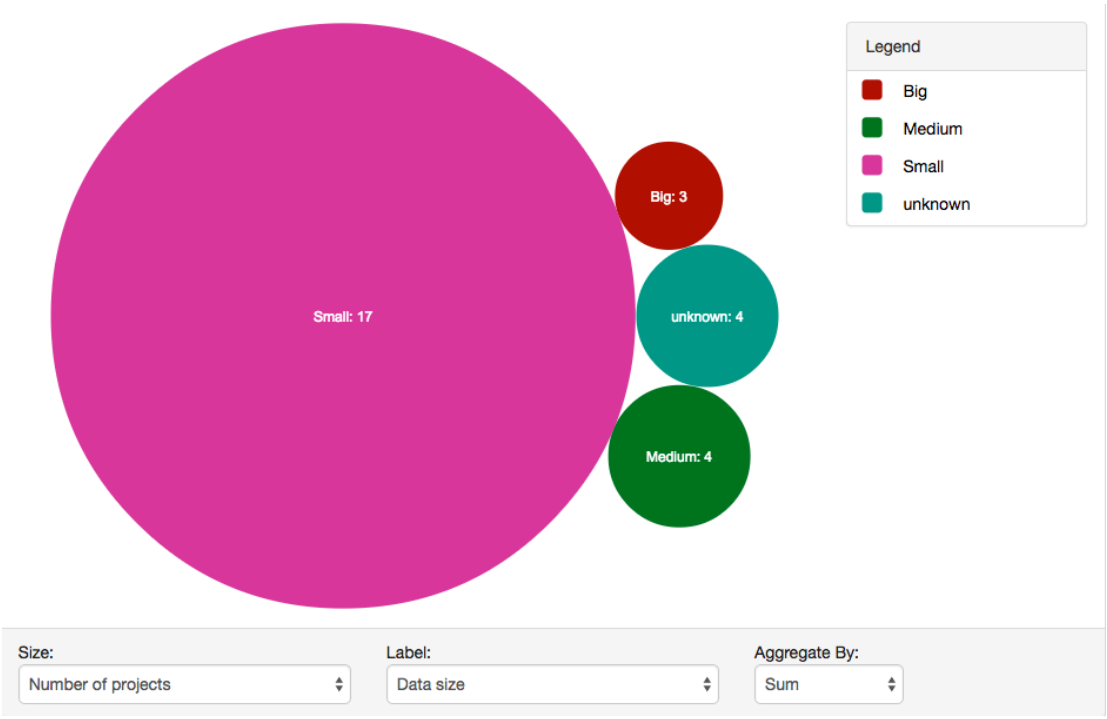
received much attention in the recent years from the sciences where technical instruments easily generate a million data points over the course of the project's time span. While big data does exist in the humanities, there are only a handful of projects that deal with data sizes in the millions in the sample of this study. As a result, the big data size was capped at one half of a million items, as that was the biggest difference between projects of large scope and medium scope in the sample studied. Arguably, one PDF of text contains more data than a single data point generated by a scientific measurement instrument, so perhaps big data in the humanities is just as big as it is in the sciences. Finally, medium data was deemed everything in between these two measurements. Not surprisingly, it is almost as frequent as big data projects in this sample, demonstrating that despite digital humanities developing original software that enables them to create and process data with computational means, the majority of projects taking place across North America today remain relatively small. We might argue that most researchers "have small data and that's okay." Big data projects, despite getting much attention, remain a minority of work underway in the academy and beyond.

This type of division may appear arbitrary, but has real-world application for any data curation project in the digital humanities involving digitization. In order to fully understand the size and scope of the project, its principal investigator needs to evaluate how much "stuff" is involved, thereby affecting how long the entire project will take to complete. Since libraries are often involved in the digitization, curation and preservation aspects of digital humanities work, they rely on vendors to provide digital scans of the data at hand, which would affect the overall cost to the institution. Having a reliable measurement of data size is extremely useful to all parties involved in this type of work. As Figure 7 and Table 8 demonstrate below, no consistently reliable measurement of data size in the digital humanities exists at this point, with the majority of projects not being to communicate clearly the amount of data involved in their particular case.

Table 8: Data sizes in projects studied

Data size	Definition	Number of projects	Percentage of total
Small	Under 5,000 items	17	61
Medium	5,000 to 500,000 items	4	14
Big	Over 500,000 items	3	11
unknown	No data size given	4	14

Figure 7: Data sizes in projects studied



1.4 Data curation activities: enrichment and usability

Several data curation activities were identified in the data sample collected. The definitions provided in Table 9 are articulated by me and reflect my understanding of the concepts, as seen in the data rather than referring to an external source, which may define them differently. These activities show a range of web usability practices standard to organization of information and access requirements in contemporary web environment.

Table 9: Data curation activities in the digital humanities

Data curation activity	**Definition or example(s)**	**Used in number of projects**	**Percentage of all projects**
Browsing	Ability to view all items in the collection through displays, page-breakdowns, titles, links, etc.	18	64
Searching	Ability to find a specific item through a search box function	11	39
Tagging	Tags assigned to objects to help find all items in that category; ability for users to add their own tags	8	29
Faceted organization of information	Multiple aspects to the same category, ie. multi-level navigation menu	18	64
Mapping or plotting data points	Data assigned to particular points on a map	5	18
Thematic organization of information	Data organized according to a theme, topic or subject	24	86
Related content to aid discovery	Links to "similar" items, displays of "recently added" or "recently reviewed" items incorporated into the website to encourage browsing through the collection	13	46
Annotation	Additional information provided about the item on the side; ability to add additional information to the item by users	3	11
Zoom-in/detailed view	Ability to enlarge the item, such as make font bigger or zoom into an image	8	29
Multiple formats for download	Multiple file formats provided for users to download, such as CSV file, PDF, Word document, XML view, etc.	2	7
Transcriptions	Text version of video or hand-written archival documents available	6	21
Data analysis built into the browser	Ability to manipulate data right in the project site, without downloading data or applying other analysis tools	2	7
Interactive functionality	Ability to save, favourite, highlight or share data	3	11

This theme reflects a wide range of activities that make up the concept of "data curation" in digital humanities practice. Humanists, therefore, manipulate the data they

have in many ways, and consequently, add a lot of value to that data by improving its access, organization and context. Inclusion of any number of these data curation activities benefits the project overall, by providing additional functionality and user engagement with digital content in the project.

1.5 Data sharing

In the age of transparency and open access, the thematic category of data sharing aimed to capture the extent to which digital humanities projects 'walk the talk' of openness and sharing of data for use and repurposing. It was important to examine whether or not they made available not just the results of their projects, but the data itself. Unfortunately, as Figure 8 and Table 10 below demonstrate, the majority (79%) of digital humanities projects do not share their data in any systematic or accessible manner – that is, either in machine-readable formats or in formats fit for human consumption.

Figure 8: Data sharing

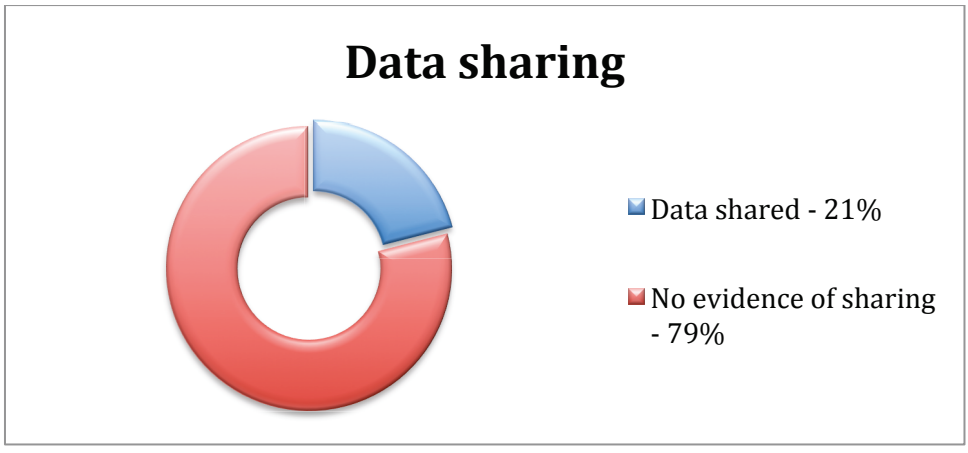


Table 10: Data sharing

Data sharing status	Number of projects	Percentage of total
No evidence of data sharing	22	79
Data shared	6	21

Of those projects that do make their data openly available, common ways that data is shared include:

- multiple file format export, including CSV, JSON, atom, dcmes-xml, omeka-json, omeka-xml, rss2
- GitHub account with code, documentation and associated files
- Application Programming Interface (API) – this includes currently implemented or future plans for one

This list is important because it demonstrates that few projects are committed to access and sharing of their research data. However, those projects that do provide a data export, demonstrate the awareness of digital humanists' needs for information access both in machine-readable and human-readable format. The variety of options for data access and sharing increases interoperability, use and engagement with the project, and is in line with the growing demand for open data in digital scholarship.

1.6 Data storage and preservation

Related to data sharing, the theme of systematic data storage evident in the study population provides insight into the institutional relationship of the digital humanities projects and their host institutions. Specifically, this section observes whether digital objects featured in the project are housed in an institutional repository or on a custom site. As Table 11 below outlines, the majority of digital humanists (82%) create their own digital ecosystems to house their data rather than partnering with their digital library services to place the objects in a repository. However, even those projects that did not link their data from the institutional repository or digital archive often cite collaboration with their institutional digital scholarship services unit – whether that is a library, a DH center, or other department. This facet of data curation activity is further explored in the community engagement section of this analysis.

Table 11: Data storage location in digital humanities projects

Location of data	**Number of projects**	**Percentage of total**
Repository or archive	5	18
Custom	23	82

1.7 Metadata

Most broadly, metadata can be defined as data about data. Often, metadata describes objects and provides further context for objects or concepts. For example, bibliographic information such as the title, author, date of publication, International Standard Book Number and subjects categories assigned to a published work all form the metadata about a particular book. By virtue of description, metadata also aids the discovery of relevant information. For example, the provision of a subject about a particular literary work allows other works of similar nature to be discovered, used and shared. This very crude definition of metadata is provides to contextualize the thematic category of metadata as outlined in this portion of the findings chapter. The theme of metadata reflects the identified efforts by project leads to reference existing data organization schemas designed to improve access and interoperability of data, especially since many cite this very reason as one of the project goals. Metadata is important for digital humanists because it provides well-documented benefits to anyone dealing with information, such as:

- Facilitating description and presentation of physical artifacts, such as books, historical documents or material objects
- Increasing interoperability among information systems and encouraging sharing of information
- Improving search and retrieval of information
- Aiding discovery of content throughout and across websites
- Creating a context for each object, which can be considered a kind of scholarship in itself

- Creating associations and relationships among data, as well as among data and users
- Increasing efficiency and reliability of browsing and search functions
- Providing consistency to information systems, such as project websites or internal databases of information about digital objects
- Improving accessibility for users of that information

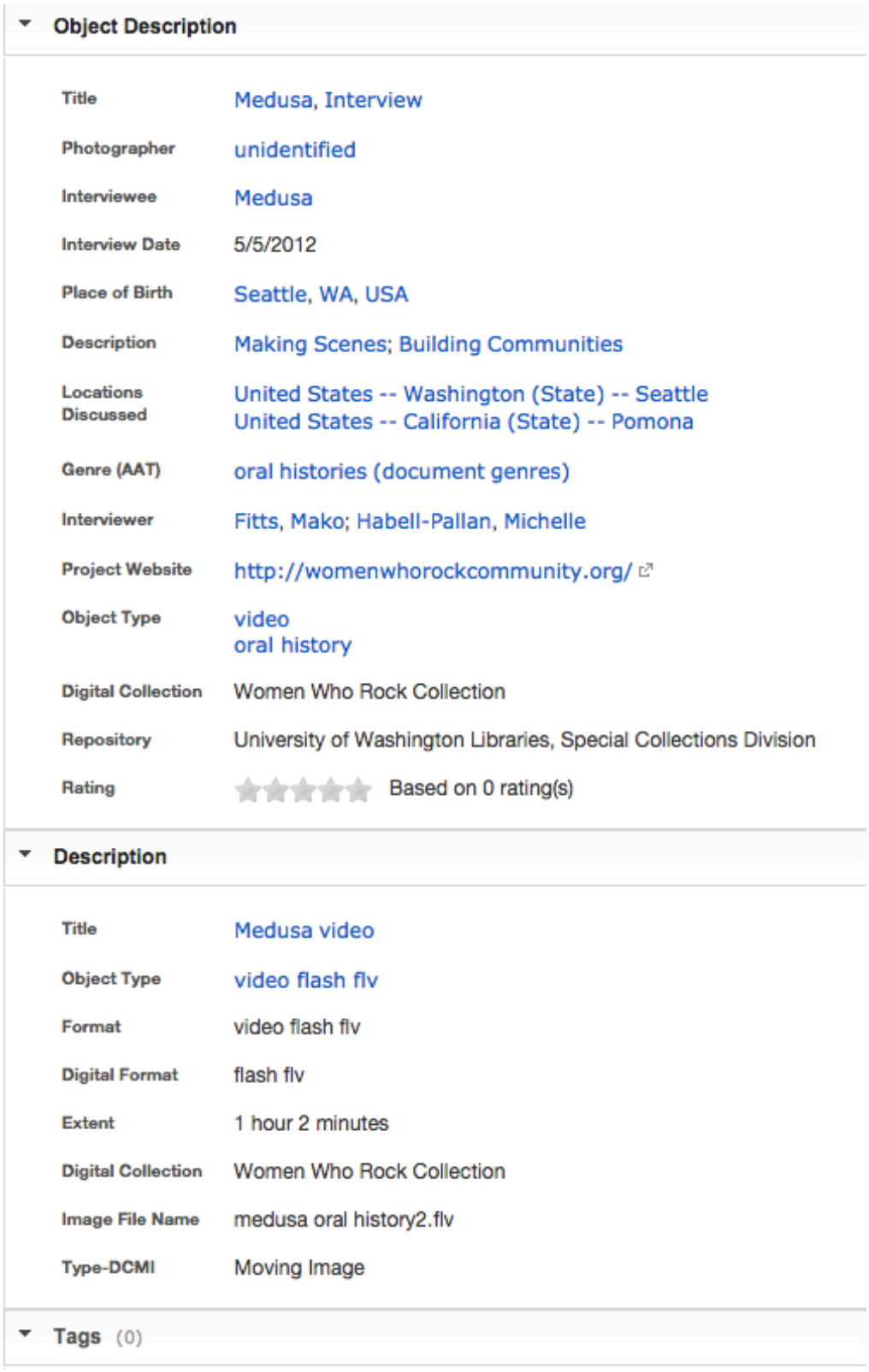
Table 12 below represents the breakdown of metadata schemas, as seen in the data sample collected. As the table demonstrates, no particular metadata standard is favoured over any other in data curation work in the digital humanities. Most of projects examined create their own custom metadata fields, with little reference to a standard, or have no metadata describing their digital objects at all.

Table 12: Metadata use in digital humanities projects

Type of metadata used	**Number of projects**	**Percentage of total**
Basic custom	12	43
Formal standard (METS, Dublin Core, MARC)	10	36
No standard/no evidence	6	21

This theme is important to the notion of data curation because the use of metadata, especially a formal standard developed by Library and Information Science professionals, such as Dublin Core standard, can provide a great benefit to discovery, browsing, searching, sharing and accessing digital information within the project site and on the web. In this context, use of formal metadata standards includes METS, Dublin Core, MARC as well as other types of data modeling techniques like Linked Open Data standards and XML schemas. What I am referring to as a "basic custom" metadata schemes means that metadata fields were created uniquely for the project, including modification or simplification of formal standards such as Dublin Core Metadata Standard. For example, Figure 9 below features a custom and extensive metadata scheme that is not modeled after any particular standard, and is adapted to suit the needs of the project.

Figure 9: Custom metadata scheme example



Object Description

Title	Medusa, Interview
Photographer	unidentified
Interviewee	Medusa
Interview Date	5/5/2012
Place of Birth	Seattle, WA, USA
Description	Making Scenes; Building Communities
Locations Discussed	United States -- Washington (State) -- Seattle United States -- California (State) -- Pomona
Genre (AAT)	oral histories (document genres)
Interviewer	Fitts, Mako; Habell-Pallan, Michelle
Project Website	http://womenwhorockcommunity.org/
Object Type	video oral history
Digital Collection	Women Who Rock Collection
Repository	University of Washington Libraries, Special Collections Division
Rating	Based on 0 rating(s)

Description

Title	Medusa video
Object Type	video flash flv
Format	video flash flv
Digital Format	flash flv
Extent	1 hour 2 minutes
Digital Collection	Women Who Rock Collection
Image File Name	medusa oral history2.flv
Type-DCMI	Moving Image

Tags (0)

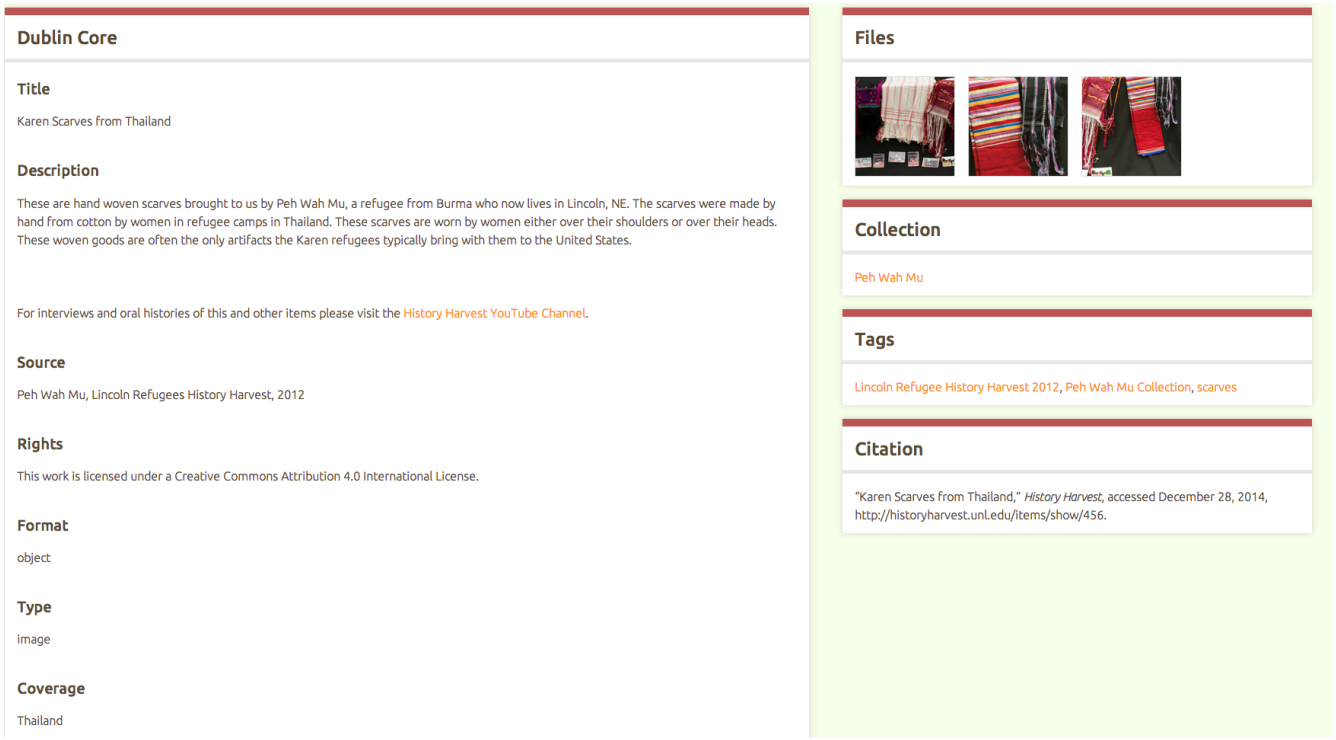
On the other hand, here is an example of the Metadata Encoding & Transmission Standard (METS) used to describe objects in the Underground Rhode Island project. Each metadata record is available in XML format from a digital document map:

METS:mets xmlns:METS="http://www.loc.gov/METS/" xmlns:mods="http://www.loc.gov/mods/v3" xmlns:xlink="http://www.w3.org/1999/xlink"xmlns:brownmix="http://dl.lib.brown.edu/md/brownmix/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xs="http://www.w3.org/2001/XMLSchema"xmlns:rights="http://cosimo.stanford.edu/sdr/metsrights/" xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-0.xsd http://dl.lib.brown.edu/md/brownmix/ http://dl.lib.brown.edu/md/brownmix/brownmix.xsd http://cosimo.stanford.edu/sdr/metsrights/ http://cosimo.stanford.edu/sdr/metsrights.xsd" LABEL="Ana Flores interview" OBJID="1126056124251972" TYPE="text.oral histories">
<METS:metsHdr CREATEDATE="2005-09-06T00:00:00.001" LASTMODDATE="2005-09-06T00:00:00.001">
<METS:agent ROLE="CREATOR">
<METS:name>
Brown University Library, Center for Digital Initiatives
</METS:name>
</METS:agent>
</METS:metsHdr>
<METS:dmdSec ID="DM1">
<METS:mdWrap MDTYPE="MODS">
<METS:xmlData>
<mods:mods ID="lost000271">
<mods:titleInfo>
<mods:title>Ana Flores interview</mods:title>
</mods:titleInfo>
<mods:name type="personal">
<mods:namePart>Flores, Ana</mods:namePart>
<mods:role>
<mods:roleTerm type="text">interviewee</mods:roleTerm>
</mods:role>
</mods:name>
<mods:name type="personal">
<mods:namePart>Siew, Angela</mods:namePart>
<mods:role>
<mods:roleTerm type="text">interviewer</mods:roleTerm>
</mods:role>
</mods:name>
<mods:originInfo>

As this snippet of the metadata record shows, two schemas are used simultaneously, where METS (highlighted in red) is used as a wrapper to encode the structure of the object description and where MODS (highlighted in blue) is used to describe the objects. This example is important because it illustrates an understanding of metadata's purpose by the project director(s). In this case, each object is given an extensive description, helping contextualize its meaning. The metadata is also consistently applied across the entire data set of the Underground Rhode Island project. It shows value of the project by demonstrating that resources were committed to make such detailed description possible. Finally, each record is encoded in extensible mark-up language (XML) for human and machine-readable consumption, which encourages indexing of each record for improved search and retrieval on the web. Overall, this record is an example of why metadata use is beneficial for digital humanities research.

Similarly, here is an example of a Dublin Core metadata record, as provided on the History Harvest project's website:

Figure 10: Metadata record for Karen Scarves from Thailand



Another interesting finding in the data consists of very detailed, structured document maps, which make up a kind of metadata record about the digital object as well. For example, Figure 11 below represents one such document map found in the Underground Rhode Island project. As one can observe, the map consists of a unique URL to a METS metadata record encoded in XML at the top as well as multiple sizes of the digital image of the object (in this case, a scanned newspaper article), each with their own unique URL. The project clearly invested time into such detailed treatment of each digital object, not only describing them in depth using a detailed metadata schema such as METS, but also uniting all relevant URLS about the object into a document map, forming another layer of metadata records. While it is unreasonable to expect every digital humanities project to go to such length, I would argue that evidence of such commitment to description, preservation and access of digital humanities data also demonstrates the benefits of partnering with organizations specializing in such activities, namely libraries. From the URLs, one can see that each object is housed in the Brown University Digital Library, showing a partnership between the project team and the academic library. Arguably, the result of such a partnership leads to more stable, accessible and contextualized data.

In the Figure 11 below we can see that the project provides multiple sizes of each image, which shows commitment to ensuring user accessibility. It also addresses current and future needs of digital humanists, who may wish to store high-resolution images for future use in their own project or reference them back to this project. While the data used in URI project is not available for systematic download, such as through an API or even a large zip file package, the provision of this type of detailed document map demonstrates commitment to best practices in data curation in the digital humanities and can be modeled across the field.

Figure 11: Document map for ‘Jar on the Floor’

Jar the Floor

Watertown, MA: Boston Globe, September 24, 1993

Document Map

You can use this page to obtain the URLs or filenames for any file that this object contains. Click on the file information button for technical information about that file.

metadata record		
http://library.brown.edu/metsrecords/1125337370851235.xml		
thumbnail files		
#	Type	Location
1	image/jpeg	http://dl.lib.brown.edu/jpegs/1125337380725433.jpg
lowres files		
#	Type	Location
1	image/jpeg	http://dl.lib.brown.edu/jpegs/1125337379443590.jpg
medres files		
#	Type	Location
1	image/jpeg	http://dl.lib.brown.edu/jpegs/1125337377971473.jpg
highres files		
#	Type	Location
1	image/jpeg	http://dl.lib.brown.edu/jpegs/1125337375778320.jpg
master files		
#	Type	Location
1	image/jpeg	1125337337314777.jpg

Of the 10 projects that make use of formal metadata standards, whether they are standards developed by library and information professionals for object description like MARC and Dublin Core, emerging data models like linked data, or data representation schemes like XML, only 7 include any statement about metadata use on their project website.

Upon closer review of the use of metadata in those 7 projects, only 3 describe how the metadata was assigned and by whom. In one case, project members are cited as sources of metadata creation, while in two other projects, the University Library and/or University Archives are responsible for the metadata used as part of the project. In all cases, the metadata was used to aid discovery, relate concepts together and provide contexts for the digital objects included in the projects. For example, while Pennsylvania's Historical Society's PhilaPlace project makes no mention about origin, purpose or use of metadata, it is nonetheless obvious that the metadata enriches the understanding of the image included in the "Books and Pamphlets" collection, as demonstrated in the Figure 12 below.

Figure 12: Metadata record for ‘Morning Light’

"Morning Light," Clipper Ship; Gross Tonnage, 850 tons

Neighborhood
Fishtown/Kensington

Related Story
Cramp Shipyard — "A Community Organization and a Community Institution"

Street Address
Richmond St. and Norris St.

Date Created
1910

Description
Illustration depicts masted ship in the ocean.

Contributing Institution
Historical Society of Pennsylvania

Collection
Historical Society of Pennsylvania: Books and Pamphlets Collection

Source (Bibliographic)
Cramp's Shipyard: the William Cramp & Sons Ship & Engine Building Company, 1830 : the I.P. Morris Company, 1828; the Kensington Shipyard Co., 1900. Philadelphia, PA, 1910.

Tags
Cramp Shipyard

Topics
Landscape & Architecture, Public & Social Life, Work

This Item Is Unavailable For Purchase

Finally, only 8 projects out of 28 sampled in this study show any indication of use of library resources or services in the process of the project. For example, they integrate the objects in the project's collection from the institutional repository, make use of Library of Congress Name Authorities to create concept networks, cite the Library as a publication partner or explicitly thank the Library for its support in making the project happen.

I was particularly interested in observing how digital humanists understand the role of metadata in their projects, and I would like to summarize their reflections, in their own words, as evidence for why other digital humanists embarking on data curation projects should consider this category early on. The following themes have been identified, as illustrated by excerpts from project websites I was able to gather:

1. Metadata helps address current and future needs of digital humanists:

- *"to facilitate the description and presentation of physical artefacts—usually textual—in the emerging linked open data ecosystem"*
- *"encoding machine-readable texts according to the needs of scholarly communities in the humanities"*
- *"enabling researchers, editors, and students to pursue a variety of scholarly investigations"*
- *"facilitate interoperability within the wider digital humanities community"*

2. Metadata improves accessibility of content:

- *"In addition to bibliographic and iconographic metadata… [we] include prose descriptions that are not only marked up for searching but are also available as descriptors of images for the visually impaired"*
- *"The purpose of any controlled vocabulary is to create consistency across collections and to improve the quality of searching. Ambiguity is reduced by using a single term to refer to a concept, object, place, etc."*
- *"With key input from librarians, we have incorporated standard MARC record cataloging and controlled vocabularies into our implementation."*

- *"By integrating library standards into our metadata, we provide a very high quality search experience, enabling users to effectively locate resources from many different angles and play back desired sections. We ensure that the organization of our data is compatible with a host of other institutions and that it will be part of long-term strategies for metadata schema design and interoperability."*

3. Metadata can always be improved, especially in crowd sourced efforts:
   - *"If you find content that has metadata (i.e. information on the location and date) that you think you can improve, click on "Dispute" underneath the content. "*

4. Metadata aids digital methods:
   - *"For centuries, scholars have used annotation to describe and interact with documents of all kinds. Writing in the margins of ancient manuscripts or creating new explanatory editions of older works has long been part of the academic enterprise."*

5. Metadata is a form scholarship in itself:
   - *"modeling ambiguity and interpretation"*
   - *"Descriptions and indexes made by the scholar who created a given recording often are the most important pieces of information for archivists and users."*

Based on the data observed in the study population, the theme of metadata discussed above demonstrates a thorough understanding and application of metadata best practices in implementing data curation projects in the digital humanities.

1.8 Copyright, data use and other licenses

The theme of copyright and data use summarizes extent to which data curation projects in the digital humanities plan for data use and repurposing by using formal means such as copyright statements, privacy policies, terms and conditions and other rules outlined by the project director(s). As Table 14 below demonstrates, the

examination of the data revealed a variety of licenses, use statements and policies present in digital humanities projects, often in multiple permutations. More than half of the projects reviewed included some sort of statement about copyright, sharing or terms of use. About half of those projects used a Creative Commons license, while the rest created their own custom statement related to the use of data found on the project site. Here are some examples of statements related to copyright, use, sharing or repurposing of data, as collected from the project websites gathered for this study:

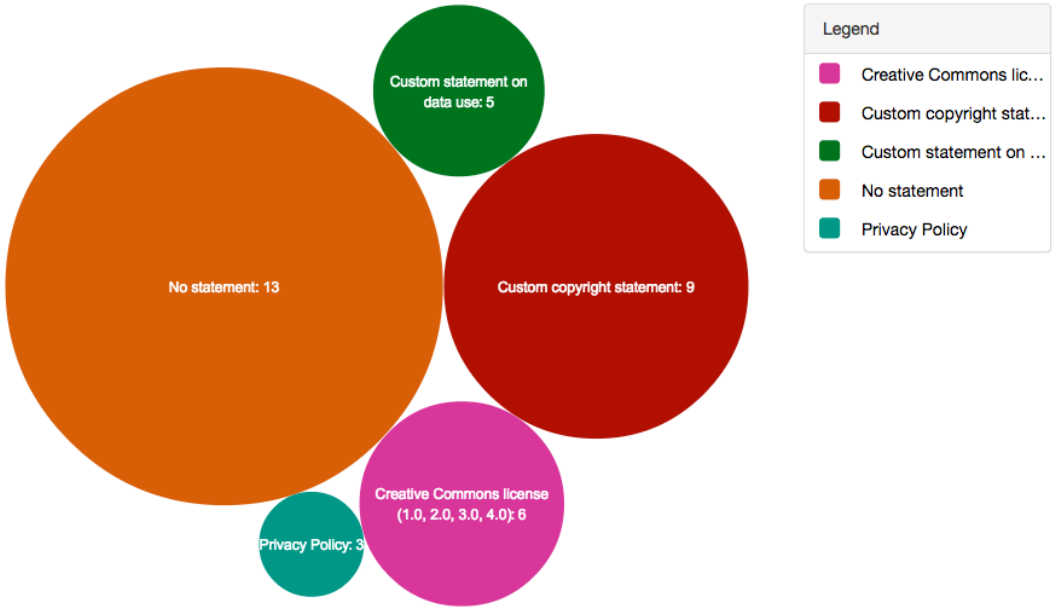
- *"The material presented in the Salem Witch Trials Documentary Archive is provided freely for non-commercial educational purposes. All other uses require advance permission from the project originators."*
- *"Copyright 2014 The Regents of the University of California.*
  All rights reserved. Document maintained by The Bancroft Library.
  Last updated 07/30/12."
- *"The data produced through this project is covered by a*
  Creative Commons CC0 1.0 Universal Public Domain Designation. designation.
  By participating in this project, you agree that any copyrightable content you submit shall be subject to a Creative Commons CC0 1.0 Universal Public Domain Dedication."
- *"The material you submit must have been created by you, wholly original, and shall not be copied from or based, in whole or in part, upon any other photographic, literary, or other material, except to the extent that such material is in the public domain. Further, submitted material must not violate any confidentiality, privacy, security or other laws. We reserve the right to mark any submission private that BMA staff identify as offensive."*

Table 13 and Figure 13 below further summarize the range of policies and licenses found in the study population.

Table 13: Copyright policies and statements

Type of license or policy	Number of projects	Percentage of all projects
Creative Commons license (1.0, 2.0, 3.0, 4.0)	6	21
Custom statement on data use or repurposing	5	18
Custom copyright statement	9	32
Privacy Policy	3	11
No statement on copyright, data use, privacy or other	13	46

Figure 13: Copyright policies and statements



The findings summarized in this section suggest that data use, repurposing and sharing can only be expected to grow in the future, as scholars, students and the general public increasingly show desire to access quality humanities data. Having standardized policies and procedures in place surrounding access, use and sharing of that data is extremely important to avoid copyright infringement, intellectual property disputes and to

manage expectations around the data. Having an established policy, license or statement of this nature helps both digital humanists and users of their project.

2. Digital humanities themes

This thematic category aims to evaluate the extent to which data curation projects in the digital humanities demonstrate commitment to digital methods of inquiry by engaging in computational methods such as data mining and analytics, visualization, digital mapping, digital narratives and others. As well, responding to David M. Berry's call for critical theorizing of the digital, this category specifically examined the awareness of the role digital methodologies play in humanities research by seeking evidence of critical reflexivity as part of the projects' missions.

2.1 Common digital humanities activities

Several digital humanities activities have been identified in the study population. Table 14 below outlines the most common of such activities carried out with data by digital humanists. This theme demonstrates further intellectual engagement with data in the digital humanities. Whereas data curation activities focus primarily on selection, organization, enrichment and preservation of data, the activities outlined below focus more on intellectual and interpretive nature of work with data in the digital humanities. Arguably, many individual defining characteristics of digital humanities activities identified here also intersect with data curation priorities, indicating that curation is an intellectual activity in its own right.

Table 14: Common digital humanities activities

Digital humanities activity	**Definition, example**
Collection or corpus building	Gathering digital objects together under common themes or under functional categories, such as creating digital dictionaries
Text analysis or data mining	Performing a computer-assisted analysis based on a body of text related to a single individual, such as sentiment analysis in a literary novel

Digital humanities activity	Definition, example
Transcription	Converting one representation of information into another format. For example, turning the content of a video or audio recording into a textual equivalent; decoding scans of digitized historical documents and creating their text-based versions
Translation	Transfer of information from one language to another, such a Spanish interview into an English version
Description	Assignment of relevant details and properties of digital objects, such as assignment of descriptive metadata in a "history harvest" gathered from a community group
Annotation	Adding information about an entity in the text, such as providing historical context, describing physical documents, commenting upon the content of the text or offering a critique
Preservation	Ensuring persistent access to documents to allow for their use, whether digitized or born-digital
Micro/oral history recording	Creating individual records for personal histories and narratives in audio, video or textual format
Visualization	Graphic representation of abstract concepts, such as relationships, processes or changes. Figure 14 provides an example of visualization activity as found in the study population.
Entity extraction	Determining the identity of entities, disambiguating between names, places or things
Creating linked open data	Publishing structured, machine-readable data that allows metadata about the object or concept described to be connected to other related resources about that object or concept
Machine-learning algorithmic applications	Creating software that operates by building a model based on the data provided and uses the model to make predictions rather than rely on specific instructions given by the software creator
Creating timelines	Presentation of chronological sequence of events along a line that allows the user to browse in either direction. Figure 15 below provides an example of timeline from a digital humanities project in the study population.

The range of DH activities identified as a result of this study suggests that digital humanists are preforming multiple complex and interdisciplinary tasks with their data. From organization to interpretation, many skills are required in manipulating the data in this field.

Figure 14: Example of visualization of a text analysis

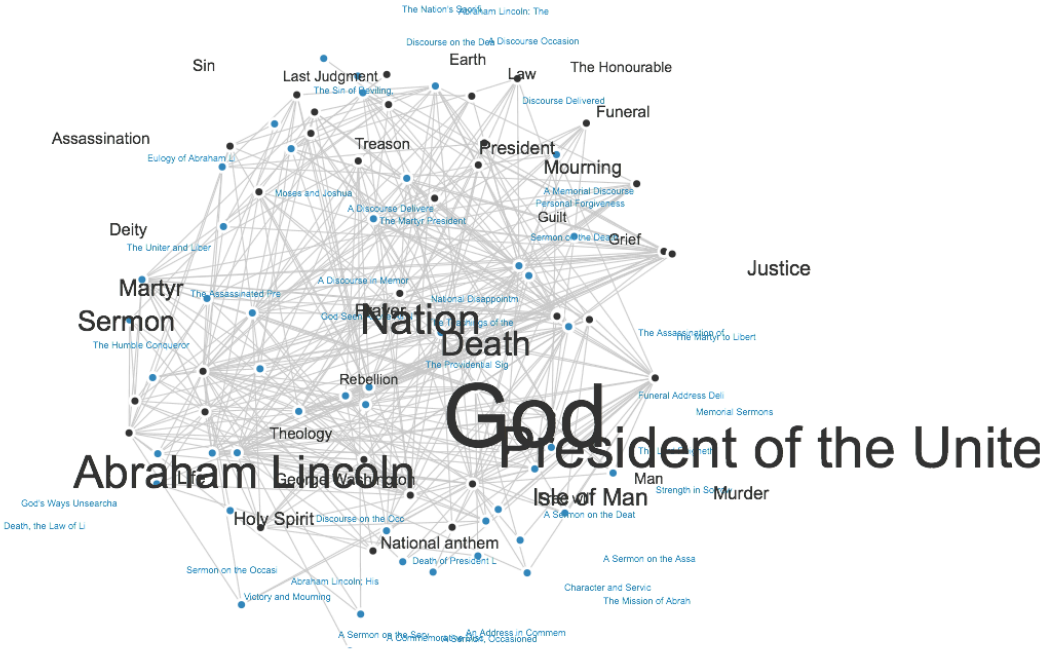
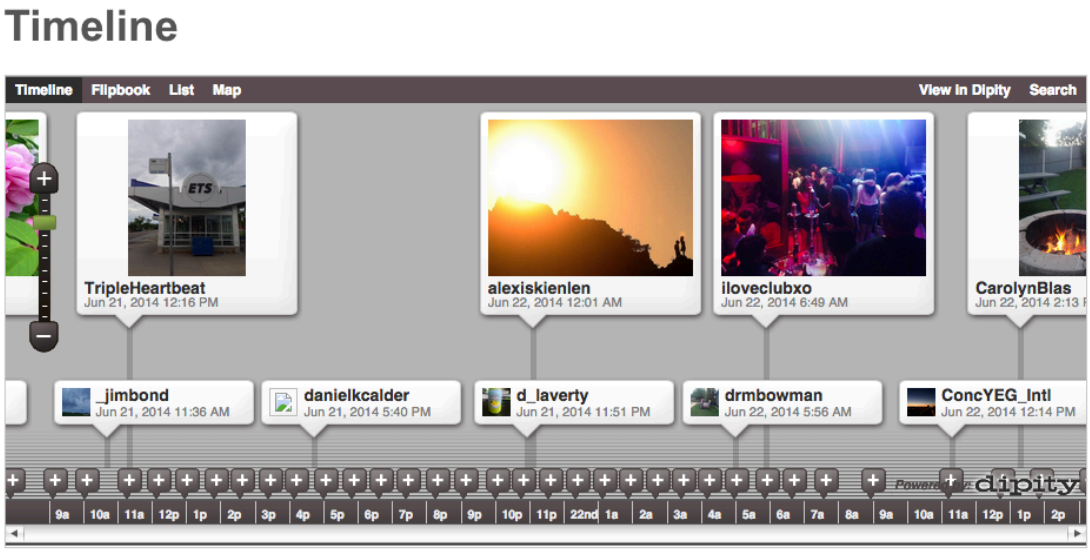


Figure 15: Example of a digital timeline



2.2 Critical reflection in digital humanities

The theme of critical reflection in the digital humanities projects has been identified in the study population. Critical reflection can be defined as the explicitly articulated sense of reflection on the role of digital processes in scholarly inquiry. In other words, I observed the capacity on the part of the data curation project leaders to express engagement with the idea of what digital scholarship means and how it impacts their work. As Table 15 below demonstrates, the majority (68%) of the projects in this study demonstrate evidence of critical awareness of the importance of digital methods, tools and approaches in the evolving landscape of humanities scholarly practice.

Table 15: Critical reflection on digital scholarship

Evidence of critical reflection	**Number of projects**	**Percentage of total**
Yes	19	68
No	9	32

In addition, many other forms of critical reflection are documented in the "about" pages of the projects studied. Here are the most common topics of reflection carried out by principal investigators and their teams:

- Micro and oral histories: commentary on the role of individual lives in the greater body historical scholarship.
- Collective memory: the importance of communal forms of historical documentation.
- Artifact-based approaches to knowledge creation: construction of historical narratives based on physical objects and cultural artifacts.
- Experiential learning: reflection on learning styles and new approaches to pedagogy.
- Cultural, historical, political, social, and other contextualization of information objects.

- Theory of space: theorizing the relationship between space and time through digital technologies.
- Digital pedagogy processes: thinking about benefits and limitations of using digital tools and methods for pedagogical purposes.

This list demonstrates that directors of the digital humanities projects selected for this study consciously position their work in the wider context of emerging modes of digital scholarship and in the long tradition of humanistic pursuit of knowledge.

2.3 New methodologies and methods in the digital humanities

Digital humanists leading data curation projects in the study population have also demonstrated evidence of participation in emergent research methodologies. A list of such methodologies includes such unique approaches to inquiry in the humanities as:

- Construction of multi-layer view of history and space through digital mapping, such as "thick mapping"
- Combining algorithmic (computational) and human processes to apply to humanities data sets
- Creating networks among data entities (people, places, ideas) - Figure 16 below is an example of one such network, for example
- Mining, organizing and visualizing data - Figure 17 illustrates one example of such processes
- Distributed quality control procedures related to data entry, description, or curation of humanities data
- Creating software to aid management and analysis of humanities data sets
- Creating mobile applications that allow users to interact with humanities data in virtual and spatial dimension
- Tracking data changes, such as textual variation over time, allowing for a kind of "digital palimpsests" to be documented and preserved for future humanities research

Clearly, these types of activities are fundamentally tied to the process of digital scholarship, and can only be pursued by working with digital methods, tools and data. This finding suggests a deep level of reflection on the process on humanities scholarship and a commitment to innovation in terms of seeking new avenues for knowledge creation. For example, by sharing a visualization of the project's Twitter network, as represented by Figure 16 below, or by summarizing the relationship of the #yeglongday Twitter campaign, as captured by Figure 17 below, the Edmonton Pipelines project encourage new perspectives on the pursuit of humanities scholarship while capitalizing on the massively popular platform.

Figure 16: Network representation of Twitter relationships

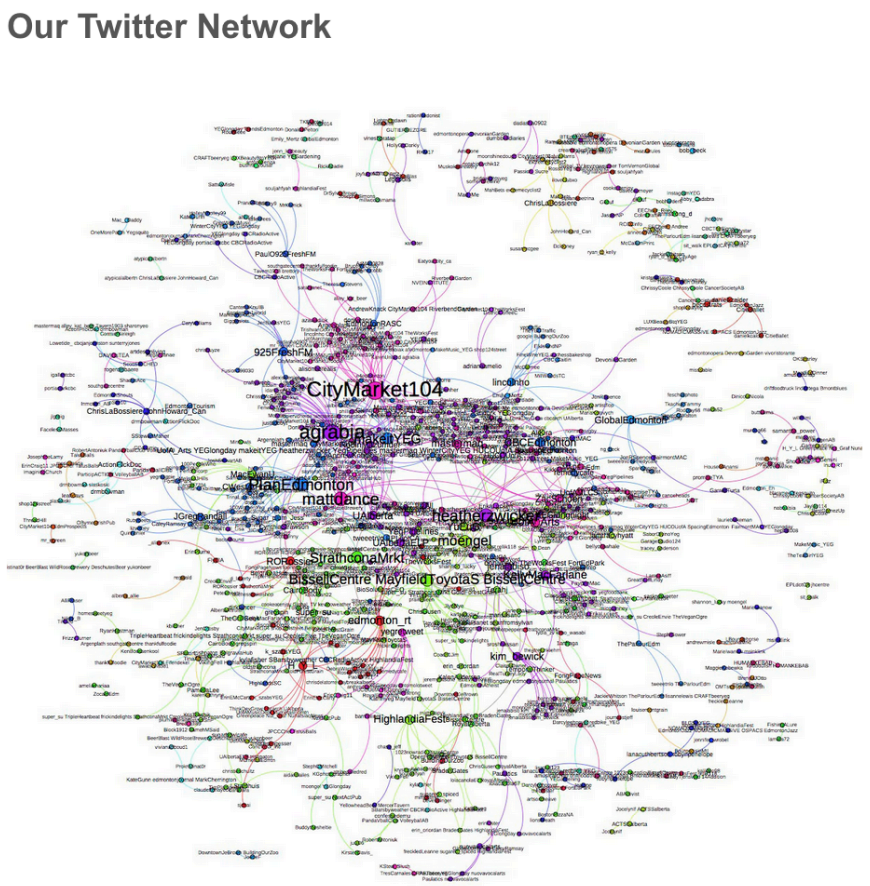
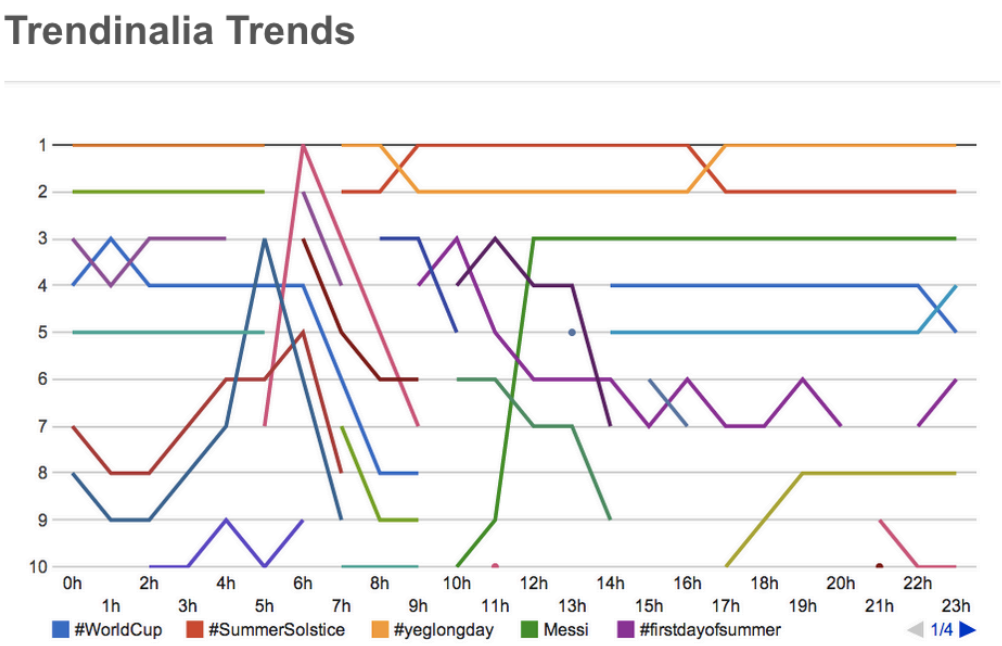


Figure 17: Twitter trends on #YEGlongday, Edmonton Pipelines project



3. Community engagement themes

The themes discussed in this section all relate to the multi-faceted notion of community engagement observed in the study population. The analysis of this thematic category provided detailed insights into the practice of community-engaged scholarship from the "ground-level" of data curation projects in the digital humanities. Notably, the themes described below emerged from the documented elements of community engagement in digital scholarly work, such as written statements, tweets, project documentation and social media pages related to the data curation projects. As I had no access to inner documents related to the concepts of impact of the projects on their target user communities, any discussion of such complex ideas as impact, value or relevance of the projects remains beyond the scope of this study. Nevertheless, the themes discussed below provide a deeper understanding of the concept of community-engaged digital scholarship in the humanities.

3.1 Primary user group

A clear theme emerged from the data gathered which suggested that the majority (57%) of digital humanities projects identify an academic audience as their main "user base." This finding is perhaps not surprising, given that most of projects selected for this study are affiliated with academic institutions, and are therefore likely to focus on the needs of academic audiences. Nevertheless, as Table 16 below demonstrates, there is reason to believe that a significant number (44%) of digital humanities projects demonstrate engagement with the general public. It can be argued that targeting both academic and public audiences supports the general mission of community-engaged data curation as a practice that distributes power and invites participation from the members of the civic society to some extent. This number is derived by combining the number of projects primarily focused on public audiences with the number of projects focused on both public and academic users.

Table 16: Primary target communities in digital humanities projects

Primary user community	**Number of projects**	**Percentage of total**
Academic	16	57
Public	8	29
Both public and academic	4	14

3.2 Division of power

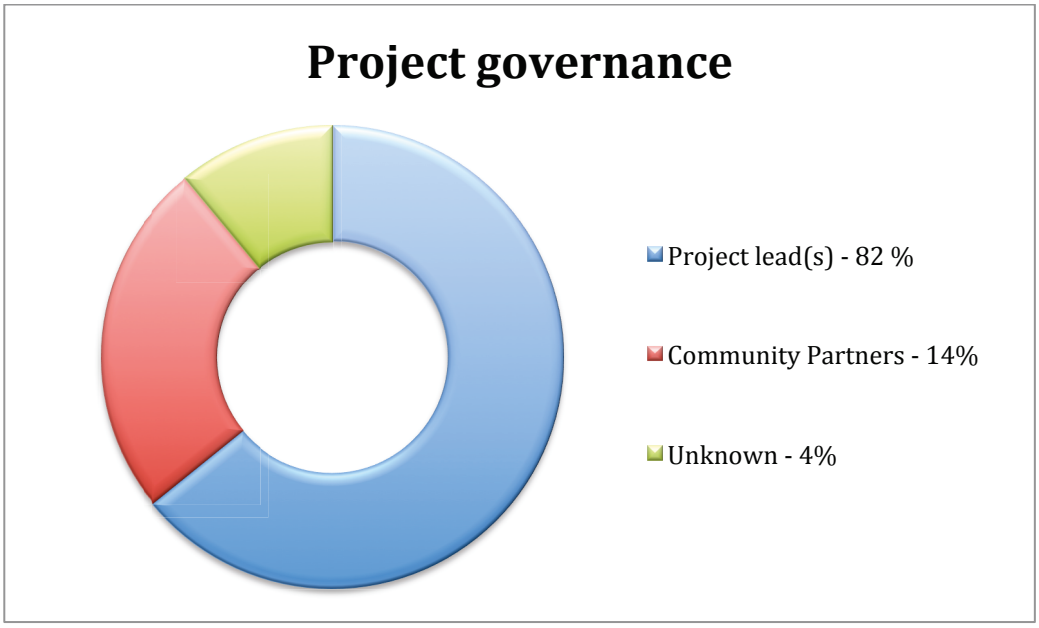
This theme reflects the observed division of power in the context of data curation project management, with reference to Arnstein's articulation of participation as a political concept. I was interested in seeing who sets the agenda in digital humanities projects. In other words, how are the priorities decided and how is the project governed? Table 17 and Figure 18 below are unequivocally clear: in the majority of cases (82%), project directors set the priorities that govern the direction and management of data curation projects in the digital humanities. In other words, despite openness to

participation from the community, project leads remain key decision-makers in the process. This is significant because by retaining control over the directions of the work, they frame the methodological and theoretical approach to the construction of knowledge rather than allowing a reciprocal, communicative exchange process to occur. At the same time, given that the majority of projects in the study population are affiliated with academic institutions, the notion of institutional authority in project governance is unlikely to be challenged or replaced in favour of other governance models. This finding suggests that the notion of institutional authority is strong in digital humanities projects, even where the broader goals of the project address the need for public knowledge transfer and engagement with the public at large.

Table 17: Project governance in digital humanities

Who sets the agenda	**Number of projects**	**Percentage of total**
Research team/project lead(s)	23	82
Community groups, partners	4	14
Unknown	1	4

Figure 18. Project governance in digital humanities



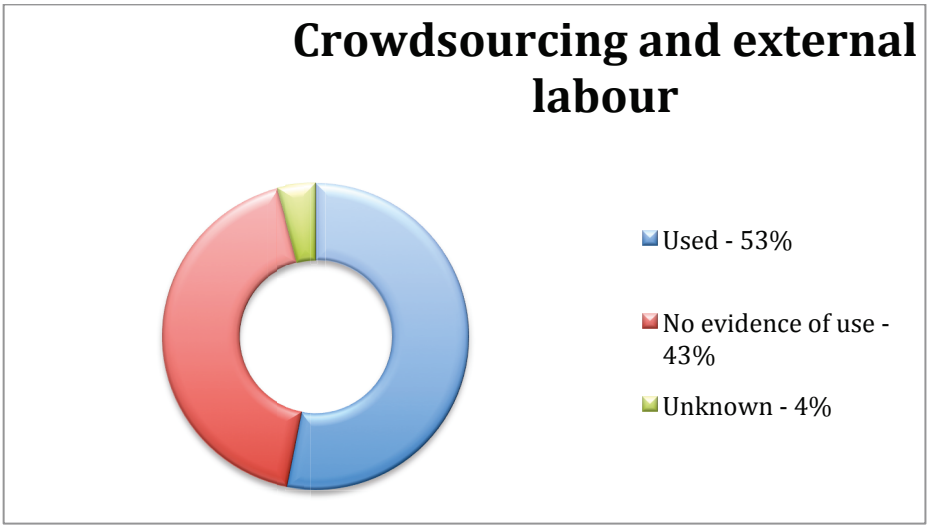
3.3 Crowdsourcing and use of other external input

A clear theme of reliance on crowdsourced labour in digital humanities data curation projects emerges out of the data gathered. Howe (2006) defines crowdsourcing as "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call' for help". For the purposes of this study, I included such activities as the use of volunteers for physical activities (workshops, events, lectures), review help, as well as activities like gamification of data quality control processes, requests for annotation, verification and edition of texts, as well as other data management tasks into the definition of crowdsourcing. Table 18 and Figure 19 below clearly demonstrate that a solid half of all projects studied (53%) rely on some form of crowdsourcing or external help into the operations of the project.

Table 18: Crowdsourcing and external labour use

Evidence of crowdsourcing?	**Number of projects**	**Percentage of total**
Yes	15	53
No	12	43
Unknown	1	4

Figure 19: Crowdsourcing and external labour



This theme addresses the participatory culture theory developed by Jenkins and references to some degree the Ladder of Participation developed by Arnstein as well. Crowdsourcing is a problematic issue in digital scholarship. It provides individuals from all over the world with the opportunity to contribute to large humanities projects, thereby allowing the possibility to learn, reflect and participate in the scholarly process. At the same time, it clearly relies on free labour of the anonymous individuals to do so, at least in cases involving digital input rather than in-person volunteer opportunities. On the one hand, it is perhaps surprising to see the extent to which data curation projects in the humanities engage their communities through this means, demonstrating the need for labour beyond the project team to achieve the goals of the project. On the other hand, crowdsourcing can also be seen as a form of citizenship if it allows individuals to contribute to a project that fosters knowledge creation and sharing, and as a result, does not frame their contribution as a form of manipulation but rather, a manifestation of personal power. This theme is highly contentious, but is presence in the data gathered suggests the need to address issues associated with crowdsourcing use in digital humanities scholarship because reliance on crowdsourced efforts is likely to increase.

3.4 Social media use

The use of social media in digital scholarship has been noted in the research literature, and, not surprisingly, is also evident in the data collected for this study. Consequently, this section aims to gain a better understanding of how and why social

media use relates to the broader category of community engagement in the digital humanities data curation practice. To do so, I observed the number and type of social media platforms used in the digital humanities projects selected for this study's population. I then analyzed in what types of activities the projects' teams were engaging through these platforms and what they hoped to achieve as a result of this use. The conclusions of this observation are summarized in Table 19 and Figure 20 below. Overall, the findings related to this theme suggest that 50% of digital humanities projects employ one or more types of social media platforms to connect with broader communities of users while the other 50% of digital humanities projects show no evidence of using social media in their online presence. Blogging and Twitter appear to be the most commonly used ways to engage users online for the digital humanities sampled in this study. Sharing and highlighting unique archival content, such as images from digital collection appears to be the most common activity achieved using social media in this population. This finding is in line with expectations about community engagement through digital means as explored in the literature on digital humanities and social media. Notably, the lack of social media use in half of the study population can be partially attributed to the face that several digital humanities projects included in the study population launched before the rise of social media on the web and terminated without ever employing social media. For example, Salem Witch Trials and The Underground Rhode Island projects are pioneers in the digital humanities, and show no evidence of using social media because Twitter, Facebook and YouTube did not exist at the time of the projects' development.

Table 19: Social media use in projects studied

Number of platforms used	**Used in number of projects**	**Percentage of all projects**
none	14	50
1-2	7	25
3 or more	7	25

Figure 20: Social media platforms used

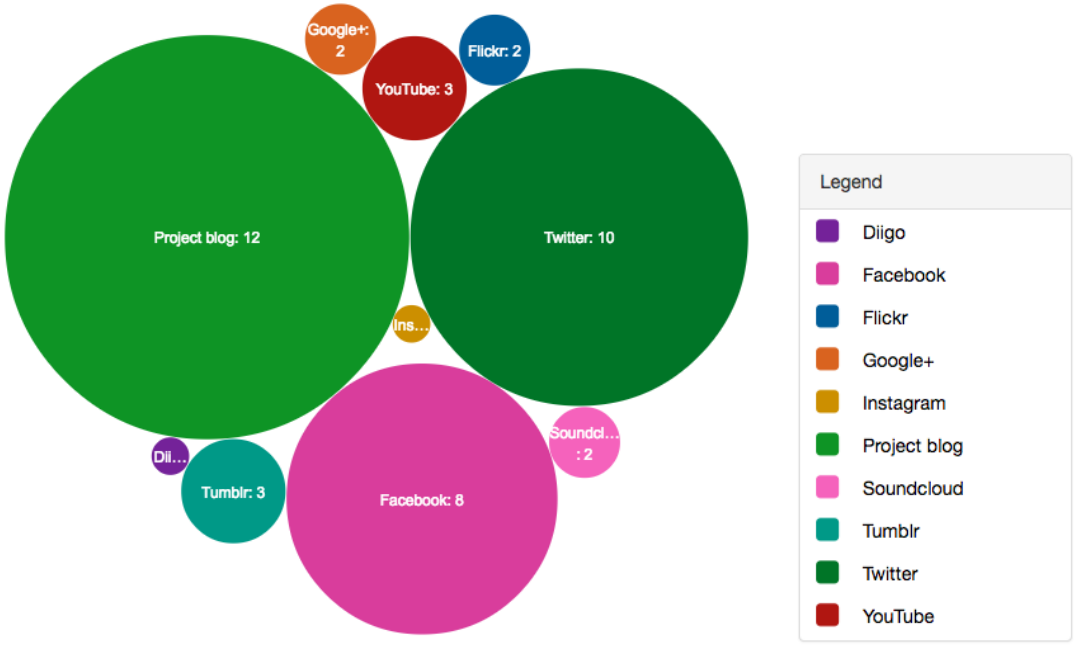


Table 20 below outlines in detail specific platforms used in data curation projects in the digital humanities, their frequency of use in the study population, as well as most common purposes for which these platforms are used by the projects.

Table 20: Social media platforms used

Social media platform name	**Used in number of projects**	**Demonstrated uses for social media platform**
Facebook	8	Sharing press coverage about project Sharing photos, videos, and other content from the project Sharing articles and websites about content related to the project by topic, including contemporary events on the same topic Sharing news articles about subjects interviewed in the project Sharing news about project's host institution Announcements about people involved in the project Announcements for upcoming and past events related to the project, including photos and videos from the events

Social media platform name	Used in number of projects	Demonstrated uses for social media platform
Twitter	10	Posts about project progress Calls for volunteers needed for the project Sharing links to articles and websites related to the project Posts about project's host institution Answering questions from Twitter followers Sharing images, videos, audio and other content from the project as highlights Sharing images, video and other content related to the project Announcements about people involved in the project Call for papers, journal submission deadlines and other scholarly communication announcements
YouTube	3	Sharing video recordings of guest speaker lecture series from the project, interviews, cultural and art performance Sharing video recordings of performance art Sharing videos of interviews with the subjects of the project Posting introductory and orientation videos describing the project Sharing testimonials and reflections from project participants about participation in the project
Tumblr	3	Sharing photos, videos and other content from the project, including public lectures Sharing quotations from related literature Updates from principal investigator
Google+	2	Highlighting content from the project, such as photos and their descriptions Following Google+ members Sharing articles about people involved in the project
Flickr	2	Sharing archival and contemporary photos to be used in the project's collection Discussions on topics proposed by members of the Flickr group Posting Flickr group description and code of conduct

Social media platform name	Used in number of projects	Demonstrated uses for social media platform
Instagram	1	Sharing photos of the project's subject, such as images of the contemporary cities, real people or activities discussed in the project, ie. landscapes of Portobelo, Panama
Diigo	1	Sharing links to articles and websites related to the project - organized by subject, i.e. "maps, geography, dh, urban"
Soundcloud	2	Audio recordings of interviews and musical performances from the project site Liking and following other channels Sharing radio spots promoting the project on local radio stations
Project blog (no specific platform)	12	Updates on project's progress News about new and upcoming projects from the same project lead or institution Sharing photos, videos and other content related to the project, such as popular culture events, films, books, etc. Upcoming event notices, news related to the project Discussion of research process, such as reflections of the principal investigator Sharing photos, videos and other content about the project Sharing photos, videos and other content about the press coverage of the project Sharing photos, videos and other content from community events and outreach program carried out by the project Posting critical essays about themes reflected in the project's collection and methods used

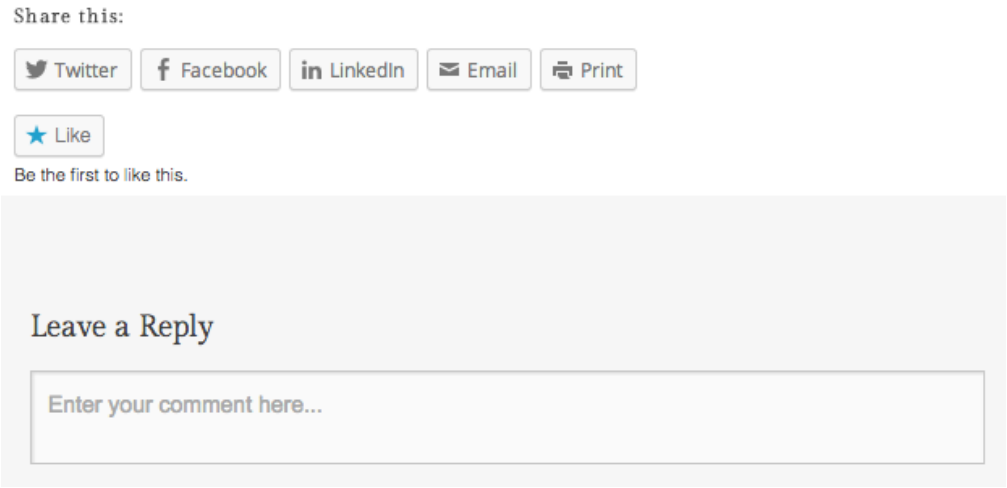
While using social media alone cannot be considered to make a data curation project engaging in and of itself, the relative popularity of this form of community engagement can be attributed to the key features of social media. In other words, both researchers and community groups involved in data curation work are likely to use social media as part of their community-engagement strategies because:

1. Social media platforms are largely free and flexible – it can be used as often or at little as needed, it has minimal access barriers in terms of registration, and it is interoperable with many other communication technologies. Admittedly, using social media has its costs for projects, as it involves the use of resources. Nevertheless, barriers to adoption of social media for digital research projects are fairly low. For example, many social media platforms integrate well with each other, by allowing users to log in with one service to use another. Content sharing is encouraged, and to, a certain extent, supports the very notion of accessibility and usability that is core to data curation.
2. It is text- and image-oriented, as the findings of this study suggest are the digital humanities as well.
3. It permits interaction and exchange with users of the project or target community.
4. It speaks the language of the general public – it is open, popular, and democratic. In many ways, social media is a common denominator for users of various ethnic backgrounds, cultures, and geographic regions.
5. It provides an immediate communication channel as opposed to the slow process of scholarly communication, such as presentations, publications and workshops.

3.5 User engagement tracking

Another identifiable and recurrent theme in the data gathered in this study includes the seemingly endless tracking of user engagement with the data available on the digital humanities projects sites, through such processes as usage statistics, clicks, shares, likes, views and other measurements. For example, in contemporary digital environment, hardly a user exists today who has not seen the Twitter, Facebook, or "Like" buttons, as represented in Figure 21 below.

Figure 21. Sharing buttons and usage tracking



Overall, about half of all digital humanities projects studied demonstrate no evidence of recording and sharing such measurements of engagement, whereas about 43% of projects do track various types of social media use on their sites. These findings are summarized in Table 21 below.

Table 21: Tracking of user engagement with data in digital humanities projects

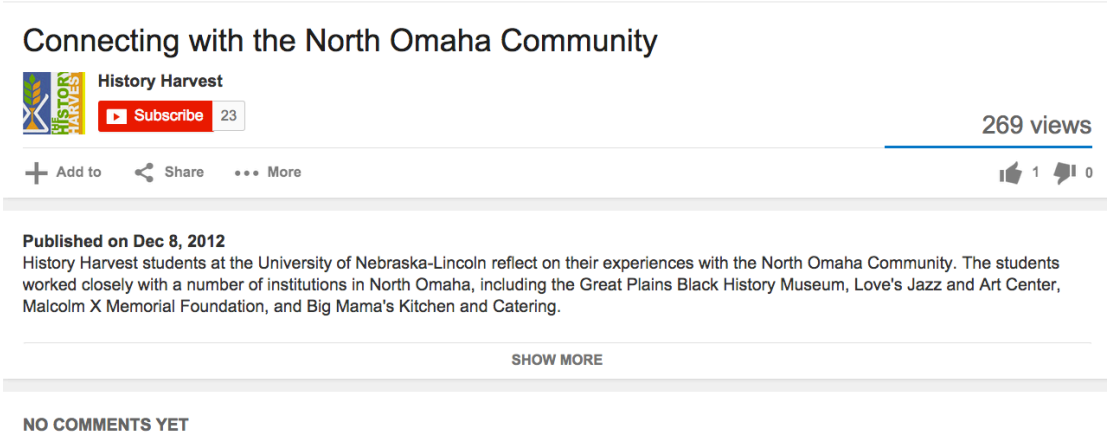
User engagement tracking evidence	**Evident in number of projects**	**Percentage of total**
Usage tracked	12	43
No evidence of tracking	14	50
Unknown	2	7

The following activities related to user engagement tracking have been identified in the study population:

- The ability for users to register and log into the project site to track usage and display statistics related to the content they shared through the project's website.
- Counting likes, views, comments, favourites and shares on the project site or through various web platforms, including social media. For example, the right hand corner of the Figure 23 below features a popular function of tracking the

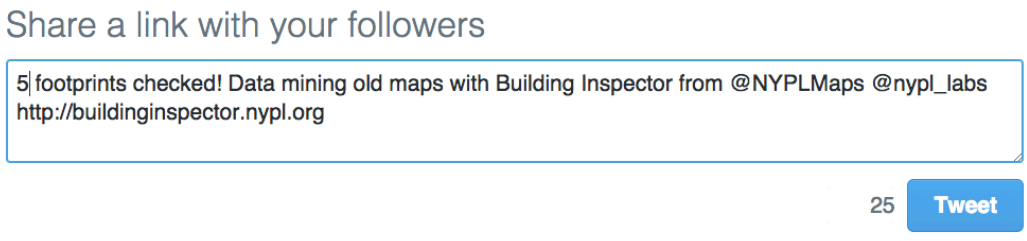
number of views on YouTube, along with positive and negative user votes on the content shared by the History Harvest project.

Figure 22: User engagement tracking on YouTube, History Harvest project



- The ability for users of the project site to tweet the number of contributions made. For example, Figure 24 below demonstrates the way New York Public Library Labs' Building Inspector project encourages its contributors to share their crowdsourced labour input to the project, thereby promoting the project, increasing awareness of its goals, and, presumably, encouraging others to do the same.

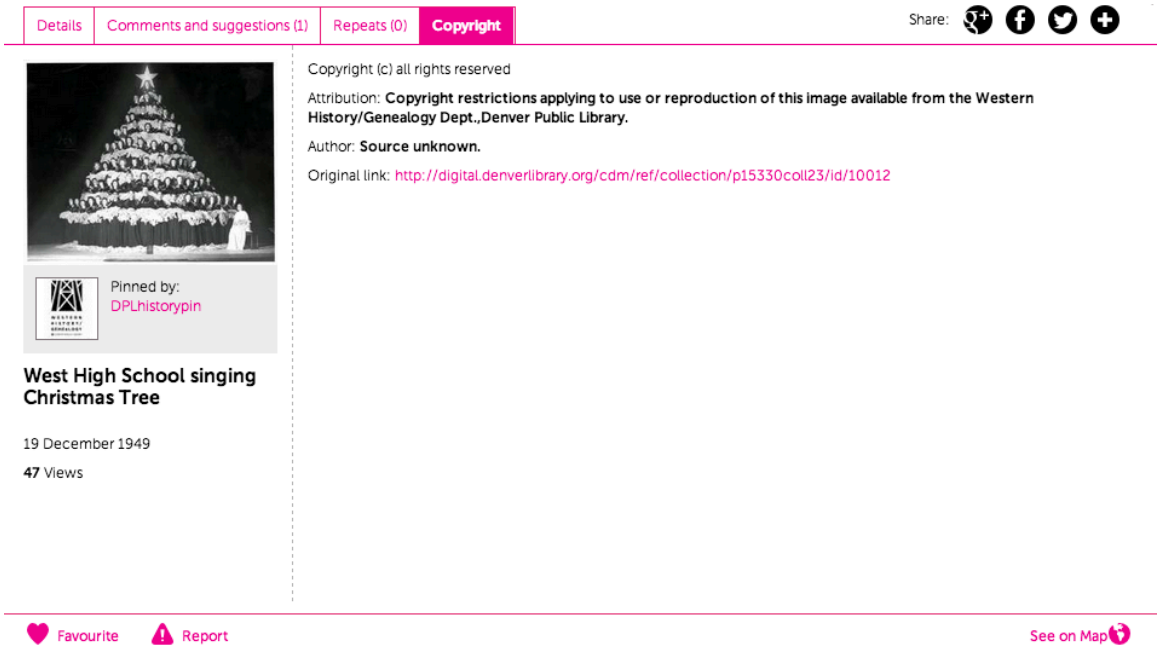
Figure 23: Example of usage tracking promotion through Twitter



- The ability to see counts on the number of views of each image. For example, Historypin displays the number of unique views on each piece of content contributed to the project site, as seen in Figure 25 below. Doing so presumably allows the contributing institution to measure more accurately the popularity of

their collections and account for impact by using Historypin as a platform for digital collections.

Figure 24: View count tracking in Historypin



- The ability to comment on a specific digital object, such as an image, video, or even map, as seen in Figure 26 below. This screenshot, for instance, is taken from the Handsome Atlas project containing digitized historical maps and statistical atlases.

Figure 25: Comment, Twitter and Facebook likes counts

3.6 Community-engagement strategies

Social media is not the only way to engage the community, however, and other forms of community engagement have been identified in the study population. These include such diverse activities as:

- Giving presentations and sharing presentation slides on the project site
- Hosting publicly accessible workshops or seminars
- Delivering outreach programs, tours, and scavenger hunts
- Sharing educational resources, such as curriculum guides and related school material on the project site
- Blogging to communicate the project's progress and exchange ideas with its users base
- Creating training videos and tutorials based on the content of the project
- Collecting feedback on the project through comments and email
- Gamifying the research process that rewards contribution and interaction with the data available through the project site
- Site membership
- Providing discussion forums
- Encouraging users to tag project's data and to create user-generated folksonomies

The identifiable theme of community-engagement strategies through digital and non-digital means suggests that a certain number of data curation projects in the digital humanities seek to establish relationships with a wide range of users, both within the academic environment and beyond it. For example, by holding tours of Philadelphia and relating such community outreach programs like walking tours and scavenger hunts to the content provided on its website, the PhilaPlace project demonstrates a commitment to active engagement with the local community. Furthermore, by creating mobile apps that employ the content organized and shared through their project sites, PhilaPlace and Building Inspector projects combine digital and analog ways of experiencing history. This clearly supports the notion that data curation can act as a form of community

engagement by encouraging active, free and open use of knowledge for individuals and groups. Similarly, by seeking content contribution from the general public to build and grow the archive of the "contemporary historical moment", projects like Yale's Historian's Eye indicate that social knowledge has a place in the academy.

3.7 Articulation of mission

The final theme identified in the data collected in this study includes the articulation, on the behalf of data curation projects, of a broader mission beyond the immediate creation of knowledge. Specifically, while 5 projects (18% of total) make no mention of anticipated impact beyond the immediate scholarly community in which they exist, the other 82% of projects studied demonstrate evidence of situating their work in the local or even global context. The perceived impact factors through which this theme can be identified include statements by the project leads about such topics as:

- Fostering open access and accessibility of knowledge as an important factor in contemporary society
- Giving priority to long-term preservation of knowledge
- Making history "alive" and bridging the past/present and the space/time divide through such technologies as mobile applications, lectures and outreach programs.
- Raising awareness about a particular issue through the project site and events relating to the project, such as local events, meetings, and advocacy efforts
- Challenging dominant narratives by highlighting the project's subject.
- Creating contexts for understanding the subject at the core of the project
- Bringing communities together and enhancing "community pride" through the digital project
- Striving to make knowledge "useful" and relevant by increasing interaction with knowledge created
- The desire to empower communities through shared knowledge, especially in marginalized populations

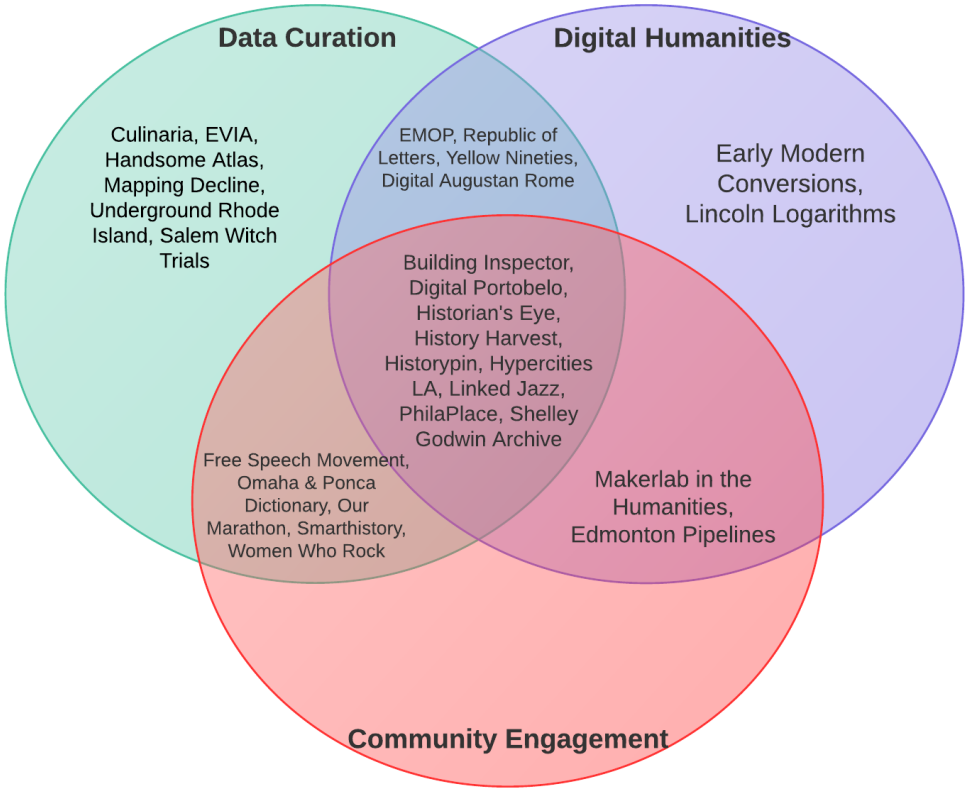
The recurrence of such highly abstract aspirations on the part of digital humanities project suggests that digital humanists clearly see beyond the immediate priorities of the scholarly environment. The desire to improve human lives, support local and global communities, as well as to ensure that future generations have access to the products of the research conducted all suggest that community engagement is a multi-faceted aspect of data curation practice in the digital humanities. It is practical and conceptual; it is passionate and reflective. This finding suggests that data curation has the capacity to advocate for the users of the humanities data it seeks to make accessible and for the digital humanities as the discipline that makes this type of work possible.

Chapter 5: Discussion

Upon review of all projects based on the presence of the themes identified in the previous chapter, each project was assigned a point in the categories of Data Curation, Digital Humanities and/or Community Engagement if it was judged to demonstrate evidence of the majority of the themes in that category. For example, if a project was deemed to show evidence of building new things, being critical of the digital, pursuing new methodologies and new methods in the humanities, and articulating a clear mission that saw itself as part of larger Humanities context, it was given a point in the DH thematic category. If the same project then also showed evidence of allowing participation from the external community, using social media to reach its audience, shared its data, tracked data usage and showed awareness of impact on the community, it was also given a point in the community engagement category. However, if that project did not have very much data, did not indicate its size, type or content, nor did it express clearly how it enriched the data and made it accessible, usable and sharable, it was not given a point in the data curation category. Consequently, such a project was given 2 points out of 3 and placed on the lower right-hand side of the thematic map, as expressed by the Venn diagram in Figure 28 below.

As evident in Figure 26, 9 out of 28 projects (32%) show evidence of all 3 categories identified as a result of this study. Furthermore, 11 out of 28 projects (39%) show evidence of any 2 of the 3 categories. Finally, 16 out of 28 projects (57%) show evidence of community engagement in one of the categories. This means that community engagement is alive and well in data curation. This finding is pleasantly surprising, because it shows that, despite having a predominantly academic focus in their projects, digital humanities do engage with broader communities to a certain extent.

Figure 26: Thematic map of projects studied



Upon positioning each project on this conceptual map, I wanted to analyze further if those projects found at the very center of the diagram – projects demonstrating evidence of all three thematic categories described in this study – shared something in common, and what those similarities might be. By reviewing these nine projects from this perspective, I intended to determine the characteristics of a truly community-engaged data curation project in the digital humanities. Based on the data collected as part of this study, a "portrait" or "profile" of a typical community-engaged data curation project in the digital humanities is likely to include the following characteristics:

- It probably has small data, which consists of digitized text, image audio and/or video. The project uses a digital tool or develops one for its purposes.
- It describes its data according to a customized metadata standard to suit the needs and processes of the project.
- It is likely to data in a human or machine-readable form, as it is aware of accessibility issues of its users.
- The project team conducts autonomous work, but sometimes partners with libraries and other organizations when necessary.
- The project team is aware of the role of digital methods and approaches in the scholarly process.
- The project uses one or two main innovative methodological approaches.
- It identifies a user group as its community and engages with it through one or two social media platforms.
- It articulates a sense of mission beyond the project itself with ambitious and audacious goals to further knowledge and improve lives.

In addition, several gaps in digital humanities data curation practice have also been identified as a result of this study. These gaps include:

1. Defining data size – it is clear that at this point, no consistent method to measure and compare data size exists in the digital humanities. Defining data size in the project site and related documentation can greatly improve curation and preservation of the project content for future use.
2. Data sharing – data sharing is not widely evident, often made difficult and in some case, prohibited. The demand for accessible data in the digital humanities is likely to increase in the future, as suggested by the research literature. Therefore, ensuring that data is organized, accessible and openly shared can greatly increase the impact of the digital humanities project.
3. Metadata – there is a wide gap in approaches to treating metadata of digital objects in the digital humanities. While some project teams recognize its importance, the majority of projects appear to treat metadata as an afterthought,

with little articulation of purpose, processes and standards. The benefits of metadata use have been highlighted in the findings chapter. Clearly, planning for metadata at the beginning of the project provides the digital humanities project with multiple benefits, such as accessibility, use and discovery of content for the future.

4. Partnership with library for storage and preservation of data – few projects appear to partner with their local libraries, such as academic digital library service units, even though doing so can benefit the stability and persistence of the data they collect, organize, analyze and share.
5. Data use statements – few projects clearly state the rules governing data use, repurposing and sharing, which makes engaging with the project data problematic for the end-user. Inclusion of such statement, and even outlining reasons for not being able to share data, help clarify the project's mission and demonstrate a thorough, systematic approach to knowledge creation.

Addressing the research questions

In this thesis, I set out to address the following research questions:

1. What are the key elements in a conceptual model for a community-engaged data curation in the digital humanities?
2. Based on these key elements, can an evaluation framework for community-engaged data curation digital humanities projects be proposed?
3. How can data curation support the goals of public humanities to foster the pursuit, development and sharing of knowledge beyond the academy?

Based on my analysis of the data gathered as part of this study, I would like to articulate my vision for a community-engaged data curation in the digital humanities. This vision includes a data curation practice that is systematic, reflective and

participative. Specifically, the key elements in a conceptual model for community-engaged data curation in the digital humanities can be described as being:

- **Systematic:** data curation that is committed to ensuring the sustainability of knowledge through the use of appropriate standards, knowledge organization structures and evidence-based practices. This includes adherence to established metadata standards, information architecture principles and technical capacity to share and preserve knowledge produced through the research process. Specifically, this element addresses having evidence-based processes in place to ensure that data is conceptually and technologically organized and preserved. For example, technical capacity to preserve digital humanities data relates to using platforms, tools or approaches to ensure that data are available for future long-term use and access.

- **Reflective:** data curation is critical of its role in ensuring knowledge preservation. It highlights the goals of the research and is aligned with the methodology of the research. It articulates the mission of the work and seeks to enable new intellectual work by ensuring access to the product of research. The choice of subject declares its political and conceptual orientation. For example, by choosing to create an accessible, online dictionary of the Omaha and Ponca language, the Omaha & Ponca Digital Dictionary project articulates a political stance that places value on preservation of North American indigenous languages and cultures. Its target user groups, then, include both the academic community of linguists and the Omaha and Ponca communities.

- **Participative:** data curation does not give up the notion of authority inherent in the role of the academic-curator, but recognizes the capacity of the broader community of users to inform and improve the curation process. This relationship to the target user community reflects a participatory model of knowledge creation, in which the community contributes, critiques and transforms the authority of the research project, which represents in many ways the larger institution that supports the work. Within this participative model, data curation is relational: it

articulates a clear sense of mission, defines a target audience and seeks to maximize the potential of the relationship with the community to make the project meaningful and impactful. The act of power distribution through various degrees of participation recognizes the expectations and possibilities of the contemporary digital environment in which such data curation work exists.

Based on the findings of this study, I would like to argue that well-organized, purposeful and participatory forms of data curation allow the research project to become an advocacy piece that highlights the existing resource gap in the field of digital humanities, and to demonstrate the value of scholarship to a broader public community. Community engagement is not a placation, to use Arnstein's vocabulary, but a value statement on the role of scholarship in contemporary civil society. In this light, community-engaged scholarship becomes another way to conceptualize the process of research. Fundamentally, community engagement is about power distribution and is an inherently political act.

The second goal of this thesis was to construct an evaluation framework for community-engaged data curation in the digital humanities. My proposed contribution in this area can be summarized in Table 22 below. It is organized around the three defining characteristics established earlier: systematic, reflective and participative capacities for community engagement in data curation. Specific evaluation criteria that make up my proposed framework are articulated as a series of statements, aiming to function as a set of guiding principles for project leads undertaking data curation work in the field of digital humanities. Examples are included in the framework to situate the criteria in the real-world context and are sourced from the study population (quoted in italics).

Table 22: Evaluation framework for community-engaged data curation in the digital humanities

Key characteristic	**Specific evaluation criteria**	**Examples**
Systematic	The mission of the project is well articulated.	*"To investigate the potential application of Linked Open Data technology to enhance discovery and visibility of digital cultural heritage materials."*
	The content of the project is identified. Its size is measured and communicated.	*"45 million pages of text have been scanned"*
	Knowledge organization principles are employed to increase data access, use and discovery.	"*by integrating library standards into our metadata, we provide a very high quality search experience, enabling users to effectively locate resources from many different angles and play back desired sections"*
	The project establishes clear policies governing use, sharing and repurpose of its data.	*"Creative Commons 1.0 license on all data available through this site"*
	The project defines, assesses and evaluates impact through quantitative and qualitative means.	*"This project's Facebook page includes 631 likes"*
Reflective	The project seeks to connect its goals and mission to the operation and management of data.	*"we're training computers to do the heavy lifting, and then distributing the remaining quality control tasks to smart, motivated citizens"*
	The projects comments on the role of digital methods in the pursuit of knowledge in the humanities.	*"the digital permits what the physical cannot"* *"working together to 'teach' digital tools how to organize and render our projects with the complexity we intend demands that we critically examine and explain the micro-steps we use to make the material mean and matter in the particular ways they do in our written and embodied scholarship"*

Key characteristic	Specific evaluation criteria	Examples
	The choice of project's subject/topic/theme is explained.	*"To create a visual synopsis about what is known about the city of Rome during a key period of Rome's transformation into an imperial capital."*
	User experience of the data is prioritized through intuitive and accessible design.	*"We want to make history accessible, interactive, intergenerational, user-friendly."*
	The project articulates a perceived impact on the community of users.	*"to empower the community to preserve and communicate their own cultural practices in the digital age"*
Participative	The project prioritizes access to information and preservation of knowledge.	*"challenge and expand notions of 'open-access'"*
	The project provides channels for feedback and dialogue with its community of users.	*"We welcome any suggestions, questions or comments in the field below or by email"*
	The project balances reliance on crowdsourced labour with provision of meaningful and rewarding opportunities for contributors.	*"The Linked Jazz API outputs the relationships discovered through Transcript Analysis, as well as any crowdsourcing contributions."*
	The project shares data in human- and/or machine-readable formats.	*"The API outputs in JSON, RDF Triples and Gephi GEXF files."*
	The project seeks collaboration with community members on its governance, operations, and/or curation efforts.	*"create partnerships between computer scientists, humanities scholars, web designers, program developers, and community members"*

The third goal of this thesis was to investigate ways in which data curation can support the goals of public humanities to foster the pursuit, development and sharing of knowledge beyond the academy. To recall David M. Berry's (2014) goals of the digital humanities include:

1. Engaging and sharing with the academic community, students, public and institutions
2. Building and making new things, new methods and meanings
3. Committing to a critical reflexive process in individual and collectivities
4. Being and becoming interdisciplinary, inclusive and connective
5. Fostering encounters with new and old objects, texts, humans and non-humans
6. Confronting matter and the materiality of things.

Based on the review of research literature on the topic and the close analysis of the data gathered as part of this study, I would like to argue that community-engaged data curation in the digital humanities can support the goals and mission of public humanities, therefore moving toward public digital humanities through:

- Understanding the purpose and values of public humanities
- Being critical of the contribution of digital methods to the humanities research process
- Sharing and communicating with scholars and the public
- Seeking partnerships and finding ways to involve individuals and groups, including both public and academic libraries, students, colleagues, community groups, and project users from around the world
- Advocating for the humanities through action
- Taking a political stance on access to the results of research and research data
- Supporting the notion of openness and accessibility
- Fostering new knowledge by encouraging data use and repurposing
- Providing meaningful ways to contribute to knowledge

- Employing quantitative and qualitative measurement strategies for digital humanities research – by collecting clicks and downloads as well as stories, photographs and testimonials from the users of the data

While all of the proposed ways to support the mission of greater public digital humanities are not without their challenges, in this thesis, I have attempted to demonstrate that a number of data curation projects in the digital humanities are already taking strides in the direction of engaging public communities. They are representative of the possibilities that digital humanities provide through their use of emergent methods and tools. Such projects also highlight the importance of systematic data curation approaches in allowing public and scholarly communities to participate in the research process. It is clear that growing engaged public humanities is an ongoing, long-term effort. However, access to digital humanities data can benefit both public and academic audiences. It is also one of the core values of the kind of humanities for which I am advocating in this thesis. Therefore, the connection I would like to draw between data curation in the digital humanities and community engagement is that open, accessible and sustainable projects that contain digital humanities data demonstrate value and become forms of advocacy that highlights the purpose and benefit of the humanities. Such projects enable data reuse and generate new ideas.

Chapter 6: Conclusion

Key findings

The research objective for this study was three-fold: to identify critical elements in a conceptual model for a community-engaged data curation in the digital humanities, to propose a set of evaluation criteria that would act as guiding principles in pursuing such work in the future, and to explore ways in which community-engaged data curation can further the mission of public digital humanities. In order to answer these research questions, I decided to study 28 data curation projects taking place in the digital humanities and carefully scrutinize their websites for evidence of emergent themes related to the categories of data curation, digital humanities and community engagement. In my study, I adopted a grounded theory methodological approach, because of its flexible nature and ability to apply the methodology to various social phenomena. Having approached my study from a pragmatist theoretical perspective, I sought to find ways in which data curation can act as a form of community engagement on communities beyond the scholarly environment. Using a constant comparative data analysis method typical of a grounded theory method study, I observed many themes, which organized into the categories of data curation, digital humanities and community engagement. I examined these themes further, and made connections between the findings that emerged from the data. The key elements of community-engaged data curation in the digital humanities therefore included practice that was systematic, reflective and participative. Having articulated the key elements of the conceptual model, I constructed an evaluation framework for conducting projects of this nature, outlining specific criteria that would serve as guiding principles of community-engaged data curation. These criteria include adherence to established metadata standards, prioritization of data accessibility, and connecting the goals and mission to the project's operation and data management practices. Finally, based on the examples observed in the study population, I argue that community-engaged data curation can advocate for public digital humanities by fostering partnerships with community groups, committing to making research data accessible and providing meaningful ways for the public to contribute to the research projects.

Further research

Future research in this area includes the following possible trajectories for analyzing data curation in the digital humanities in more depth by:

1. Conducting a study with larger sample size and including digital humanities projects from all around the world, whether Europe, Australia, New Zealand, South America, Africa and Middle East. It is possible that the very notions of "community" and "engagement" begin to break down in parts of the world not influenced by neoliberal attitudes toward knowledge creation and dissemination, and therefore their relationship to concepts outlined in this study may be entirely different.
2. Comparing institutional policies on community engagement with policies, goals and missions of DH projects carried out by researchers belonging to those institutions. In other words, a study of this nature might involve comparing to what extent DH projects live out the larger service missions of their universities.
3. Conducting a study that evaluates the degree of user engagement in data curation projects in the digital humanities by collecting both observations and interviews to compare project missions with actual and/or perceived impact from the perspective of the community of users of those projects.

There are many other ways to approach this topic, but these types of studies might further our understanding of the changing nature of data curation in the digital humanities. As more training is delivered by librarians, more collaboration occurs between libraries and DH researchers, as universities rethink their commitments to service as their institutional mandates, this field will continue to evolve and transform. Nevertheless, a general understanding of what actually takes place in data curation projects in the digital humanities is a useful starting point to explore this field.

The humanities have a great potential to question, critique, challenge and contextualize the digital in contemporary society, which is why critical theory has an incredibly powerful and useful role to play in that endeavor. Just as linguists have discussed the linguistic turn in social and political theory, so too a 'computational' or a digital turn needs to be discussed in the contemporary social sciences and humanities

domains (Berry, 2014, p. 49). As Berry (2014) argues, instead of teaching computer programming as a skill required for function in a new digital economy, we need to explore instead "the historical, philosophical, theoretical and critical context for particular kinds of the various forms of digital praxis" (p. 209). My study contributes to this critical assessment of digital scholarship in the humanities in practice by situating it in a larger context of community engagement. Data curation can play a significant role in the area of public digital humanities by fostering preservation, access and sharing of knowledge by embracing the theory and practice of community-engaged digital scholarship.

Final thoughts

The practice of data curation that has tended to be discussed in library and information studies research literature without a theoretical stance, as if the organization of knowledge is an apolitical act that exists in an environment untouched by political and social forces. I was interested in exploring how data curation can be seen as a form of community engagement, and how the participative aspects of community-engaged scholarship reflect the division of power between the institution and the community of users. Data curation can further the values and goals of community-engaged scholarship. Arguably, data curation in the digital humanities has a high potential to support the mission of public humanities by capitalizing on the public interest in the subject of digital humanities data and by sharing the results of that research with the public as well. It was my intention with this thesis to bring a pragmatist perspective to the conceptual discussion of both the notions of data curation and digital humanities, and to gain a deeper understanding of the field from an evidence-based approach.

Bibliography

4Humanities. (n.d.) Mission. Retrieved from http://4humanities.org/mission/ Accessed December 3, 2014.

Alemu, G., Stevens, B., Ross, P. & Chandler, J. (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World*, *113*(11-12), 549-570.

Anderson, S., & Blanke, T. (2012). Taking the Long View: From e-Science Humanities to Humanities Digital Ecosystems. *Historical Social Research / Historische Sozialforschung, 3*(141), 147-157.

Arnstein, S. R. (1969). Ladder of citizen participation. *Journal of the American Institute of Planners*, 35*,* 216-224.

Barker, D. (2004). The scholarship of engagement: a taxonomy of five emerging practices. *Journal of Higher Education, Outreach and Engagement, 9*(2), 123-37.

Bartha, M. (2010). Serious Work: Public Engagement and the Humanities. *Western Humanities Review*, *64*(3), 85-104.

Baumeister, R. F. & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529.

Bengtson, J. (2012). Preparing for the age of the digital palimpsest. *Library Hi Tech, 30*(3), 513-522.

Benson, R. (2009). Shaping the Public Sphere: Habermas and Beyond. *The American Sociologist, 40*(3), 175-197.

Berman, R. A. (2012). Humanities and the Public Sphere: Scholarship, Language, Technology. *Telos,* 159, 173-186.

Bernardou, A., Constantopoulos, P., Dallas, C., Gavrilis, D. (2010). Understanding the Information Requirements of Arts and Humanities Scholarship. *The International Journal of Digital Curation, 1*(5), 18-33.

Berry, D. M. (2012). Introduction: Understanding the digital humanities. In D. M. Berry (Ed.), *Understanding digital humanities* (pp. 1-20). Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.

----------. (2014). *Critical theory and the digital*. New York: Bloomsbury Academic.

----------. (2014, May 8). Our working group's six principles of #digitalhumanities for #teachDH. [Tweet]. Retrieved from https://twitter.com/berrydm/status/464339281285513216 Accessed October 22, 2014.

Bettez, S. C. (2011). Critical Community Building: Beyond Belonging. *Educational Foundations 25(*3-4), 3-19.

Bitsch, V. (2005). Qualitative research: A grounded theory example and evaluation criteria. *Journal Of Agribusiness, 23*(1), 75-91.

Block, P. (2008). *Community: The structure of belonging.* San Francisco: Berrett-Koehler.

Borgman, C. L. (2007). *Scholarship in the digital age: information, infrastructure, and the Internet.* Cambridge, MA: MIT Press.

Borgman, C.L. (2009). The digital future is now: a call to action for the humanities. *Digital Humanities Quarterly 3*(4).

Brewin, C. R. (1989). "Cognitive Change Processes in Psychotherapy". *Psychological Review* **96** (45): 379–394.

Bringle, R. G., & Hatcher, J. A. (2002). Campus-community partnerships: the terms of engagement. *Journal Of Social Issues*, 3, 503.

Burns, D., Hambleton, R. & Hoggett, P. (1994). *The Politics of Decentralization: Revitalizing Local Democracy.* London: Macmillan.

Chang, C.-C., Lin, C.-Y., Chen, Y.-C., & Chin, Y.-C. (2009). Predicting information-seeking intention in academic digital libraries. *Electronic Library*, *27*(3), 448-460.

Carolina Digital Humanities Initiative (2014). Retrieved from http://digitalhumanities.unc.edu/resources/projects/ Accessed November 5, 2014.

Charmaz, K. (2006). *Constructing grounded theory.* Thousand Oaks, CA: Sage.

Choudhury, G.S. (2008). *Case Study in Data Curation at Johns Hopkins University. Library Trends, 57*(2), 211-220.

Chowdhury, G. (2009). From digital libraries to digital preservation research: the importance of users and context. *Journal of Documentation, 66*(2), 207-223.

Cohen, D. & Crabtree, B. (2006). Qualitative Research Guidelines Project. Retrieved from http://www.qualres.org/HomeInte-3516.htm Accessed January 18, 2015.

Collie, A.W. & Witt, M. (2011). A Practice and Value Proposal for Doctoral Dissertation Data Curation. *The International Journal of Data Curation, 2*(6), 165-175.

Corbin, J. & Strauss, A.C. (1990). Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology 13*(1), 3-22.

Corbin, J. & Strauss, A.C. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam.

Data Curation Guide. (2012). Retrieved from http://guide.dhcuration.org/index.html Accessed February 18, 2013.

DeLugan, R., Roussos, S., & Skram, G. (2014). Linking Academic and Community Guidelines for Community-Engaged Scholarship. *Journal Of Higher Education Outreach And Engagement, 18*(1), 155-168.

Delanty, G. (2001). *Challenging knowledge : the university in the knowledge society*. Buckingham [England]; Philadelphia, PA: Society for Research into Higher Education: Open University Press.

Dillon, C. (2013). The Research Library as Digital Curator at Virginia Tech. *College & Undergraduate Libraries, 20*(2), 232-237.

Digital Curation Center. (2014). What is digital curation? Retrieved from http://www.dcc.ac.uk/digital-curation/what-digital-curation Accessed September 27, 2013.

Dobson, K. (2011). Mining the valley of its making: culture and knowledge as market commodities. In Coleman, D., & Kamboureli, S. (Eds.), *Retooling the humanities : the culture of research in Canadian universities*. Edmonton: University of Alberta Press.

Dunn, S., & Hedges, M. (2013). Crowd-sourcing as a Component of Humanities Research Infrastructures. *Journal Of Humanities & Arts Computing: A Journal Of Digital Humanities, 7*(1/2), 147-169. doi:10.3366/ijhac.2013.0086

Durrance, J. C. (2001). The vital role of librarians in creating information communities: strategies for success. *Library Administration & Management*, *15*(3), 161-168.

Eckert, P. (2006). Communities of practice. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 683-85). Elsevier. Retrieved from http://web.stanford.edu/~eckert/PDF/eckert2006.pdf Accessed March 11, 2015.

Eckert, P. & McConnell-Ginet, S. (2007). Putting communities of practice in their place. *Gender & Language 1*(1), 27-37.

Edmonton Public Library. (2010, Revised 2013). Community-Led Service Philosophy Toolkit. Retrieved from http://www.epl.ca/sites/default/files/pdf/CommunityLedServicePhilosophyToolkit.pdf Accessed December 15, 2014.

Ehrlich, T. (Ed.). (2000). *Civic responsibility and higher education.* Phoenix, AZ: Oryx Press.

Enos, S., & Morton, K. (2003). Developing a theory and practice of campus-community partnerships. In B. Jacoby (Ed.), *Building Partnerships for Service-Learning* (pp. 20-41). San Francisco: Jossey-Bass.

Fenwick, T. (2014). Futures for community engagement: a sociomaterial perspective. In Gibbs, P. & Barnettm R. (Eds.), *Thinking about Higher Education* (pp. 109-122*)*. Cham: Springer.

Freilich, M. (1963). Toward an Operational Definition of Community. *Rural Sociology 28*(2), 117-127.

Ge, X. (2010). Information-seeking behaviour in the digital age: a multidisciplinary study of academic researchers. *Digital Age*, 435-456.

Glasser, B. G. (1992). *Basics of grounded theory analysis.* Mills Valey, CA: Sociology Press.

Glasser, B. G. & Strauss, A. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Griffiths, P. E. (1997). *What Emotions Really Are: The Problem of Psychological Categories*. The University of Chicago Press: Chicago

Harvey, R. (2010). *Digital Curation: A How-To-Do-It Manual.* New York: Neal-Schuman Publishers, Inc.

Hayles, N. K. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. Chicago and London: University of Chicago Press.

Heidorn, P. B. (2011). The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration, 51*, 662-672.

Hillery, G. (1955). Definitions of community: areas of agreement. *Rural Sociology, 20*, 111-123.

Holland, B. A. (2001). From the Editor. *Metropolitan Universities: An International Forum, 12*(4), 5.

Howe, J. (2006) Crowdsourcing: a definition. Wired Blog Network: Crowdsourcing. Retrieved from http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html Accessed December 2, 2014.

Hoyt, L. (2010). A City-Campus Engagement Theory from, and for, Practice. *Michigan Journal Of Community Service Learning, 17*(1), 75-88.

Hswe, P., Furlough, M. J., Giarlo, M. J. & Martin, M. (2011). Responding to the Call to Curate: Digital Curation in Practice at Penn State University Libraries. *The International Journal of Digital Curation, 2*(6), 195-207.

Jenkins, H. (2009). *Confronting the challenges of participatory culture: media education for the 21st century*. Cambridge, MA: The MIT Press, 2009.

Juola, P. (2008). Killer Applications in Digital Humanities. *Literary and Linguistic Computing 21*(1): 73-85.

Khoo, S. (2013). Between Engagement and Citizenship. Engaged Scholarship, 21. doi:10.1007/978-94-6209-290-7_3

Kimball, M. J., & Thomas, D. F. (2012). Place-Building Theory: A Framework for Assessing and Advancing Community Engagement in Higher Education. *Michigan Journal Of Community Service Learning, 18*(2), 19-28.

Lazarus, R. S. (1982). Thoughts on the Relations between Emotions and Cognition. *American Physiologist 37*(10), 1019–1024.

Lee, C.A. & Tibbo, H.R. (2007). Digital curation and trusted repositories: steps toward success. *JODI: Journal of Digital Information 8*(2).

Leedy, P. D., & Ormrod, J. E. (2010). *Practical research: planning and design*. Upper Saddle River, NJ: Merrill.

Lerner, J. S.& Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion 14*(4), 473–493.

Lester, Geoffrey. (2005). Atlas of Alberta Railways. Edmonton: University of Alberta Press. Retrieved from http://railways.library.ualberta.ca Accessed November 3, 2012.

Liu, A. (2011). The state of the digital humanities: a report and a critique. *Arts and Humanities in Higher Education, 11*(1-2), 8-41.

--------. (2012). Where is cultural criticism in the digital humanities? In M.K. Gold (Ed.), *Debates in the digital humanities* (pp. 490-502). Minneapolis: University of Minnesota Press.

Lockwood, D. K., Lockwood, J., Krajewski-Jaime, E. R., & Wiencek, P. (2011). The Community Engaged Faculty: An Education Process. *International Journal Of The Humanities, 9*(1), 87-95.

Manovich, L. (2013). *Software takes command: extending the language of new media*. Gordonsville, VA: Bloomsbury USA Academic.

Maron, N. L., Yun, J. & Pickle, S. (2013, January 29). Sustaining our Digital Future: Institutional Strategies for Digital Content. Ithaka S+R. Retrieved from

http://www.sr.ithaka.org/research-publications/sustaining-our-digital-future Accessed November 29, 2014.

Maron, N. L. & Pickle, S. (2014, June 18). Sustaining the Digital Humanities: Host Institution Support Beyond the Startup Phase. Ithaka S+R. Retrieved from http://www.sr.ithaka.org/research-publications/sustaining-digital-humanities Accessed November 29, 2014.

Metadata Object Description Schema (MODS). (2014). Retrieved from http://www.loc.gov/standards/mods/userguide/typeofresource.html Accessed January 12, 2015.

Munoz, T. (2012). Digital humanities in the library isn't a service. Retrieved from https://gist.github.com/3415438 Accessed March 3, 2013.

NCCPE: National Co-ordinating Center for Public Engagement. (2009). Draft Definition of Public Engagement, www.publicengagement.ac.uk.

Neuroth, H., Lohmeier, F. & Smith, K.M. (2011). TextGrid – Virtual Research Environment for the Humanities. *The International Journal of Digital Curation, 2*(6), 222-232.

Newton, M. P., Miller, C. C. & Stowell Bracke, M. (2011). Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force. *Collection Management*, 36, 53-67.

Noonan, D., & Chute, T. (2014). Data Curation and the University Archives. *American Archivist, 77(*1), 201-240.

Ogburn, J. (2010). The Imperative for Data Curation. *Libraries and the Academy, 10*(2), 241-246.

Onciul, B. (2013). Community Engagement, Curatorial Practice, and Museum Ethos in Alberta, Canada." In Golding, V., & Modest, W. (Eds.). *Museums and communities: curators, collections, and collaboration*. London: Bloomsbury.

Oxford Dictionaries. (2014). Retrieved from http://www.oxforddictionaries.com/definition/english/curate. Accessed February 5, 2015.

Pasanek, B. (2010). The Mind is a Metaphor. Retrieved from http://metaphors.lib.virginia.edu/metaphors Accessed September 13, 2012.

Parry, D. (2012). The digital humanities or a digital humanism. In M.K. Gold (Ed.), *Debates in the digital humanities* (pp.429-437). Minneapolis: University of Minnesota Press.

Pharr, S. (2010). Reflections on liberation. In M. Adams, W. J. Blumenfeld, C. Castaneda, H. W. Hackman, M. L. Peters & X. Zuniga (Eds.), *Readings for diversity and social justice* (2nd ed.) (pp. 591-598). New York: Routledge.

Phelan, S. (1996). All the comforts of home: The genealogy of community. In N. J. Hirschmanii & C. DiStefano (Eds.), *Revisioning the political: Feminist reconstructions of traditional concepts in western political theory* (pp. 235-250). Boulder, CO: Westview Press.

Poole, A. H. (2013). "Now is the Future Now? The Urgency of Digital Data Curation in the Digital Humanities." *Digital Humanities Quarterly, 7*(2).

Posner, M. (2013). No Half Measures: Overcoming Common Challenges to Doing Digital Humanities in the Library. *Journal of Library Administration, 53*, 43-52.

Prom, C. (2011). Making Digital Curation a Systematic Institutional Function. *International Journal of Digital Curation,* 6(1), 139-152.

Raboin, R., Reznik-Zellen, R.C. & Salo, D. (2012). Forging New Service Paths: Institutional Approaches to Providing Research Data Management Services. *Journal Of Escience Librarianship, (3)*, 134. doi:10.7191/jeslib.2012.1021

Ray, J. (2012). The rise of digital curation and cyberinfrastucture. *Library Hi Tech, 30*(4), 604-622.

Rice, R. & Haywood, J. (2011). Research Data Management Initiatives at University of Edinburgh. *International Journal of Digital Curation, 2*(6), 232-244.

Rosaen, C. L., Foster-Fishman, P. G., and Fear, F. A. (2001). The citizen scholar: Joining Voices and Values in the Engagement Interface. *Journal of Metropolitan Universities, 12*(4), 10-29.

Sandy, M. (2011). Practical Beauty and the Legacy of Pragmatism: Generating Theory for Community-Engaged Scholarship. *Interchange: A Quarterly Review Of Education, (3),* 261.

Schecter, D. (2013). *Critical theory in the twenty-first century.* New York : Bloomsbury.

Schnapp, J. & Presner, P. (2009) Digital Humanities Manifesto 2.0. Retrieved from http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf Accessed November 17, 2013.

Schreibman, Susan, Ray Siemens and John Unsworth. "The Digital Humanities and Humanities Computing: An Introduction." In Schreibman, S., Siemens, R. & Unsworth, J. (Eds.). *A Companion to Digital Humanities*. Blackwell Publishing.

Schuetze, H.G. (2012). Universities and their communities – engagement and service as primary mission. In McIlrath, L., Lyons, A. & Munck, R. (Eds.), *Higher Education and Civic Engagement.* New York: Palgrave Macmillan.

Spiro, L. (2011). Why the Digital Humanities? GLCA New Directions Digital Humanities Workshop. Retrieved from https://digitalscholarship.files.wordpress.com/2011/10/dhglca-5.pdf Accessed March 4, 2013.

Spiro, L. (2012). This is why we fight: defining the values of the digital humanities. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp.16-35). Minneapolis: University of Minnesota Press.

Star, L. (1998). Grounded Classification: Grounded Theory and Faceted Classification. *Library Trends, 47*(2), 218.

Stewart, T., & Alrutz, M. (2012). Meaningful relationships: cruxes of university-community partnerships for sustainable and happy engagement. *Journal Of Community Engagement And Scholarship (1)*, 44.

Stoecker, R., Loving, K., Reddy, M. & Bollig, N. (2010). Can community-based research guide service learning? *Journal Of Community Practice, 18*(2/3), 280-296. doi:10.1080/10705422.2010.485878

Stone, L. (1992). Disavowing community. In H. Alexander & R. Arcilla (Eds.), Philosophy of Education, 1992. Retrieved from http://www.ed.uiuc.edu/eps/PES-Yearbook/92_docs/STONE.HTM

Sukovic, S. (2011). E-texts in research projects in the humanities. *Advances in Librarianship, 33*, 131-202.

Sula, C.A. (2013). Digital Humanities and Libraries: A Conceptual Model. *Journal of Library Administration, 53*, 10-26.

Tahir, M., Mahmood, K. & Shafique, F. (2008). Use of electronic information resources and facilities by humanities scholars. *The Electronic Library, 28*(1), 122-135.

Talja, S., Vakkari, P., Fry, J. & Wouters, P. (2007). Impact of research cultures on the use of digital library resources. *Journal of the American Society for Information Science and Technology, 58*(11), 1674-1685.

Tibbo, H. R. (2012). Placing the Horse before the Cart: Conceptual and Technical Dimensions of Digital Curation. *Historical Social Research, 37*, 187-200.

Toms, E. G. & O'Brien, H. L. (2006). Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation, 64*(1), 102-130.

Urquhart, C., & Fernández, W. (2013). Using grounded theory method in information systems: The researcher as blank slate and other myths. *Journal of Information Technology, 28(*3), 224-236. doi:http://dx.doi.org/10.1057/jit.2012.34

Vandergrift, M. & Varner, S. (2013). Evolving in Common: Creating Mutually Supportive Relationships Between Libraries and the Digital Humanities. *Journal of Library Administration, 53,* 67-78.

Vershbow, B. (2013). NYPL Labs: Hacking the Library. *Journal of Library Administration, 53*, 79-96.

Vinopal, J. & McCormick, M. (2013). Supporting Digital Scholarship in Research Libraries: Scalability and Sustainability. *Journal of Library Administration, 53,* 27-42.

Ward, C., Freiman, L., Jones, S., Molloy, L. & Snow, K. (2011). Making Sense: Talking Data Management with Researchers. *The International Journal of Digital Curation, 2*(6), 256-273.

Warwick, C., Terras, M. M., & Nyhan, J. (2012). *Digital humanities in practice*. London: Facet Publishing in association with UCL Centre for Digital Humanities.

Wilson, J. A. J., Martinez-Uribe, L., Fraser, M. A. & Jeffreys, P. (2011). An Institutional Approach to Developing Research Data Management Infrastructure. *The International Journal of Digital Curation, 2*(6), 274-287.

Wilson, M. & Wilde, P. (2003). Benchmarking community participation: developing and implementing the Active Partners benchmarks. Work: Joseph Rowntree Foundation. Retrieved from http://www.jrf.org.uk/sites/files/jrf/1859351026.pdf Accessed January 20, 2015.

Wise, M., Spiro, L., Henry, G. & Byrd, S. (2007). Expanding roles for the institutional repository. *International digital library perspectives, 23*(2), 216-223.

Witt, M., Carlson, J., Brandt, S. & Cragin, M.H. (2009). Constructing Data Curation Profiles. *International Journal of Digital Curation, 3*(4), 93-103.

Yakel, E. (2007). Digital curation. *OCLC Systems & Services: International digital library perspectives, 23*(4), 335-340.

Young, I. M. (1990). *Justice and the politics of difference.* Princeton, NJ: Princeton University Press.

Zajonc, R. B. (1980). Feelings and Thinking: Preferences Need No Inferences. *American Psychologist 35*(2), 151–175.