

## Data Assemblages: A Call to Conceptualize Materiality in the Academic Ecosystem

Nabeel Siddiqui <nasiddiqui\_at\_email\_dot\_wm\_dot\_edu>, College of William and Mary

### Abstract

As the diversity of digital humanities practitioners grows, the need to construct a framework allowing for equal acknowledgement to all involved has become more evident. In this article, I argue that the perceived immateriality of scholarship privileges conventional academic labor over similar pursuits, such as data curation, resulting in them being glanced over and ignored in tenure reviews and job evaluations. To counter this, I create a theoretical framework that places materiality at the forefront. More specifically, I draw on and expand Gilles Deleuze's notion of assemblages, as outlined by new materialist philosopher Manuel DeLanda, to posit the idea of "data assemblages," which are the result of digital humanities labor and consist of material parts contingent on their contextual relations and always in flux. I use the Digital Public Library as a case study and highlight how the reconceptualization of digital humanities labor moves beyond the merely theoretical to allow us to better understand the interdependency of individuals in the academic ecosystem and has broader implications for the nature of materiality in the digital age.

Despite the prominence of the digital humanities, discussions surrounding the assessment and nature of digital scholarship remain at the forefront of un/conferences, workshops, and other academic gatherings. Although some prominent organizations such as the Modern Language Association and American Historical Association have begun to discuss how to evaluate digital scholarship, many digital humanists continue to face backlash from peers, department chairs, and in many cases, scholarly organizations [Modern Language Association 2013] [Denbo2014]. According to Patrik Svensson, "For tenure-track scholars, there is often a sense that digital modes of representation may place you at a relative disadvantage. Indeed, this may be outright advice from senior faculty and administrators" [Svensson 2010].

1

At the same time, we have seen a shift in the types of people engaged in the digital humanities. While early practitioners largely came from traditional disciplines, digital librarians, data curators, alt-acs, and independent researchers are now just as likely to be seen at digital humanities conferences as tenured/tenure-track professors. Calls to reform scholarship in the digital age, however, have continued to privilege the research of traditional faculty over other practitioners, although both often generate similar products. As collaboration between different members of the academic ecosystem increases, I contend that it is imperative we create a new *theoretical* framework that adequately acknowledges the parallels between the work of conventional scholars and others, especially librarians and curators. In this essay, I hope to begin this reconceptualization by accentuating the shared materiality of both text-heavy digital humanities scholarship and more curatorial work through the notion of "data assemblages."

2

According to Kathleen Fitzpatrick, scholars "need to let go of some of our fixation on the notion of originality in scholarly production, recognizing that, in an environment in which more and more discourse is available, some of the most important work that we can do as scholars may more closely resemble contemporary editorial or curatorial practices" [Fitzpatrick 2011, 12]. The focus on originality has limited digital humanities scholars, and the work of some practitioners at the expense of others. As librarian Dorothea Salo notes, "academia privileges its notion of research to such a degree that it refuses to respect my praxis" [Salo 2011]. Similarly, in response to the failure of some faculty to recognize the work of curators and librarians as equal, Trevor Munoz argues that we should frame data curation as a form of publishing. According to Munoz, "Data curation as a 'publishing' activity is increasingly relevant to the working

3

lives of digital humanities scholars. Moreover, articulating connections between ‘publishing’ and data curation is important in the context of strategic decisions libraries might make and, in fact, are making about how to participate in ‘publishing’” [Munoz 2013]. Overall, calls like Munoz’s and Salo’s highlight how the separation between curation and conventional digital humanities scholarship has affected the ability of both to push research forward.

As scholars move back and forth between conventional and “alternative” academic positions, it becomes necessary to examine the equivalences in labor for future job evaluations, promotional reviews, and tenure requirements. A key reason why these evaluations sometimes downplay curatorial work is because of an enduring perception of research as disembodied practice. Jeanne Hamming and Helen Burgess make a similar contention in their analysis of materialism in new media studies and scholarship [Burgess 2011]. By drawing on the works of Bruno Latour, they highlight that the detachment of the material from the intellectual is continuing evidence of modernity’s inability to take hybrids and quasi-objects seriously. “This ongoing debate about new media’s functional (and in some cases even biological) difference from old media,” they claim, “contributes to a double erasure, for scholars working in multimedia, of both their intellectual contributions and their material labor” [Burgess 2011]. For Hamming and Burgess, it is important to examine the affordances and restraints of materiality in intellectual research. Still, despite their contentions, Hamming and Burgess maintain a rhetoric that continues to stress the importance of “traditional” academic arguments. For instance, in their discussion of Marcel O’Gorman’s *Dreadmill*, they point out that “he cites Nietzsche, Kittler, Virilio, Haraway, Ernest Becker, among other scholars” in order to lend credence to his academic authority [Burgess 2011]. Nonetheless, a significant portion of digital humanities scholarship does not contain a conventional thesis-driven argument, and this discourse continues to reject many non-traditional forms of academic labor.

The perceived immateriality of scholarship is even more insidiously damaging to the digital humanities because it reestablishes the privileging of conventional analogue scholarship above curatorial pursuits. In this essay, I seek to counter this by positing the notion of “data assemblages” as a partial solution. I begin by briefly surveying the changing nature of research in the era of big data and the attempts by both curators and others to make sense of its impact on the academy. Next, I provide an overview of Gilles Deleuze’s notion of assemblages as outlined by new materialist scholar Manuel DeLanda. By expanding on this concept, I detail the way that “data assemblages” allow us to more easily observe the analogous labor of traditional scholars, digital humanists, and curators. I define data assemblages as the resulting end product of curatorial and digital humanities labor that consist of material parts that are contingent on their contextual relations and always in flux. Following this, I use the Digital Public Library of America as a case study for my notion of data assemblages. Finally, I conclude by demonstrating how this reconceptualization moves beyond the merely theoretical to understand the interdependency of individuals in the academic ecosystem, and more broadly, permits us to reexamine the nature of materiality in the digital era.

## The Changing Nature of Scholarship in the Digital Era

The impact of widespread data on academic research has yet to be seen. Still, numerous groups have contended that this research will impact not only scientific but also humanities research. In this section, I briefly examine how scholars and librarians have viewed the emergence of big data and highlight how, despite numerous attempts, a comprehensive foundation remains unavailable that takes seriously the materiality of digital humanities labor *and* its contextual and fluctuating nature.

According to Chris Anderson of *Wired* Magazine, we are now living in the Petabyte Age [Anderson 2008]. Anderson polemically argues that this mass quantity of available data is especially important for scholars because it requires us to reexamine our methodological assumptions. According to him, we do not need to know why numbers allow us to predict behavior. Instead, “the point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves” [Anderson 2008]. The sciences have led the charge in reevaluating their epistemological assumptions in light of big data. Early attempts at examining data in context emerged in informatics, but a more extensive movement arose in the 2000s known as E-science. E-science provided a lens to deal with the massive influx of data and the increased necessity of collaborative scholarship. According to one group of early practitioners, E-science “enables a new order of collaborative, more inter-disciplinary research, based on shared research expertise, instruments and computing resources, and, crucially, increasing access to collections of primary

research data and information” [Lord 2004]. Major scientific research now exists within this E-science paradigm. One example is the Sloan Digital Sky Survey that has resulted in an opening up of scholarship and a reassessment of the role of mass computational methods in astronomical research. The team that led the project regarded it as the beginning of a “fourth paradigm” in scientific research [Gray et al. 2002]. Similarly, large computational methods have begun to dominate the fields of bioengineering and physics.

Data curation, a growing faction in library and information science, takes seriously the claims of these scientific big data movements and seeks to address the “data deluge” more thoroughly. The movement’s most direct precedent is in the notion of digital curation. Digital curation, according to Elizabeth Yakei, is “the active involvement of information professionals in the management, including the preservation, of digital data for future use” [Yakei 2007]. As Allen Renear and Trevor Munoz note, however, digital curation fails to engage with the research needs of scholars. In contrast, they seek to reframe *digital* curation as *data* curation. For them, “data curation addresses the challenge of maintaining digital information that is produced in the course of research in a manner that preserves its meaning *and* [emphasis mine] usefulness as a potential input for further research” [Munoz 2013]. By taking the needs of researchers as critical to their own praxis, data curators make apparent the need for scholars to find a way to recognize one another’s work through a shared lens.

More broadly, library science has also taken up the question of context in the gathering of resources through the concept of “collections.” Collections are important for research and scholarship because they recognize that “the totality of the records provides information that no individual record can” [Duff and Johnson 2002]. In other words, they represent the need of understanding how material develops in a context. Still, the perceived materiality of collections problematizes them as a framework for digital libraries. For instance, Hur-Li Lee highlights, “The long history of the library being associated with a physical building may have ... made imagining virtual collections difficult” [Lee 2005].

Others have pushed to understand the relations between different materials through the notion of “contextual mass.” Contextual mass is a principle that places an “emphasis on collecting materials that work together as a system of sources, with meaningful interrelationships between different types of materials and subjects, to support research inquiry” [Palmer, Zavalina, and Fenlon 2010]. Collections based on the principle of “contextual mass” can be of any size as long as their contextual relations remain. The main problem with contextual mass, however, is that it perpetuates an image of a non-fluctuating collection that serves as a singular totality. As a result, the focus is on the materiality of the collection itself rather than the larger fluctuating infrastructural needs of research.

Despite changes in the library and science communities, humanities scholars, as a whole, have been less likely to view themselves as requiring large-scale infrastructure or needing to examine data contextually. One reason for this is because of the nature of data that humanities scholars find important. Christine Borgman notes, “the humanities and arts are the least likely of the disciplines to generate their own data in the forms of observations, models, or experiments” [Borgman 2009]. Humanities scholars are more likely to rely on data created by others — i.e. diaries, literature, and films — than their own. Yet, this may be changing. For instance, large-scale projects in corpus linguistics have gained traction in the digital humanities community. Organizations, such as the HathiTrust, and projects, such as Google Books, now require humanities scholars to seriously examine the infrastructure encompassing their data. Furthermore, national and international funding agencies are now taking data seriously before financing humanities projects. The National Endowment of the Humanities, for instance, has begun to require Data Management Plans in order to secure grant funding [National Endowment for the Humanities]. The often-ambiguous guidelines and regulations of these plans demonstrate the embryonic nature of assessing digital materials. Yet, with a large organization such as the NEH leading the charge, other granting agencies will follow suit, and there will be an increased requirement for digital humanities scholars to reassess how they share and store their data. Still, most humanities scholars, outside of the digital humanities, are not trained in computational methods, and the nature of humanities research often discourages collaborative efforts that may yield new insights.

All this is not to say, however, that humanities scholars have completely neglected to examine changing infrastructural needs in the digital age. Digital history pioneers Dan Cohen and Roy Rosenzweig noted the difficulty of large-scale information preservation [Cohen and Rosenzweig 2006, 1–17]. Later, Rosenzweig argued that the problem of digital

infrastructure was no longer the question of scarcity but of abundance. He feared that the massive amount of information on the Internet would be too large to preserve and a large record of cultural production would be lost. Rosenzweig implored historians to think “simultaneously about how to research, write, and teach in a world of unheard of historical abundance and how to avoid a future of record scarcity” [Rosenzweig 2003]. In other words, because of the mass of information that was created on the web there were fundamental issues involved in assessing what exactly we should archive.

Along with grappling with how to save digital data, humanities scholars have also sought to create new degree programs that address the need for graduate training in digital methods. Regarding an attempt to form a Master's Degree in Digital Humanities at the University of Virginia, John Unsworth noted that the catalyst for the degree was based on:

13

The simple observation that our culture and our cultural heritage are migrating very rapidly to digital forms, and in order to manage that migration and take advantage of the new intellectual and creative possibilities it offers, we will need trained professionals who understand both the humanities and information technology, and we will need them in a number of different areas — in museums, libraries, teaching, scholarship, publishing, government, communications, entertainment, and elsewhere [Unsworth 2001]

Despite these outliers, however, few humanities practitioners are deeply engaged with the digital humanities. According to Shelia Anderson and Tobias Blanke, “Except for a small minority the humanities do not have a tradition of dealing with machine algorithms, with the graphs produced from statistical analysis, and the maps, trees and other forms of visual representation that arise from big data analysis” [Anderson 2012]. As a result, humanities scholars, for the time being, will likely continue to rely on curators and information science professionals trained in handling large data structures, and a theoretical structure that acknowledges the often-equal input of all camps is imperative.

14

## Data Assemblages

Christine Borgman contends that one of the fundamental questions that digital humanities needs to answer is that of “what are data?” [Borgman 2009]. For many, data exists in the immaterial space of the digital. The notion of cyberspace as immaterial is a common one in science fiction, and popular movies, such as *The Matrix*, perpetuate an image of the computer realm as something dream-like [Wachowski 1999]. This association of cyberspace with the immaterial comes from pioneers of modern science fiction, such as William Gibson. According to Gibson's definition of cyberspace in *Neuromancer*, cyberspace is an illusory and hallucinogenic condition:

15

Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts... A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights, receding [Gibson 1984, 51]

Of course, data *is* material. For example, although new computational developments such as cloud computing harken to an ephemeral and illusory cyberspace, they still make use of large-scale materiality through server farms. As Jean-François Blanchette notes, “computing systems are suffused through and through with the constraints of their materiality” [Blanchette 2011]. In addition, Ian Bogost and Nick Montfort demonstrate that it is crucial we take the materiality of computation seriously in our examinations of digital culture [Bogost and Montfort 2009]. Asserting data's materiality, however, is not enough. A more comprehensive framework is needed, and the digital humanities has yet to provide a widely adopted ontological basis for research projects that deviate from orthodox textual scholarship. Due to the nature of digital culture, it is important that this ontological framework earnestly examine not only the fluctuation of

16

data but also the changing relationship between different forms of data. As Joris van Zundert notes, “transformed by digital technology, text and digital editions – digital humanities data in general, as a matter of fact– become fluid, ‘living’, reaching a state wherein they are perpetually in a digital information lifecycle” [Van Zundert 2012]. In order to provide a more comprehensive framework for digital projects and to highlight the relationship between curation and conventional scholarship, I examine Manuel DeLanda’s expansion on Gilles Deleuze’s notion of “assemblages” in his *New Philosophy of Society* as a partial solution.

The concept of assemblages may be unfamiliar to many — although it is gaining some academic interest — and a key reason for this is that the literature on it is nebulous and scattered throughout Deleuze’s work. DeLanda reformulates Deleuzian assemblages into a comprehensive framework and seeks to broadly examine social institutions through a realist perspective. However, instead of viewing his framework as only relevant to social institutions, I adapt it as a broader way to view digital humanities data and scholarship. Specifically, through the notion of assemblages, I seek to conceptualize curatorial work and digital humanities praxis as resulting in “data assemblages.” While curation has made a case for the reexamination of data management techniques, it has often continued the rhetoric that poses a fundamental dichotomy between research and curatorial activities. Curation, in this sense, becomes a service that operates as the back-end of digital humanities work but is not an equal part of it. Data assemblages, on the other hand, provide one way to help curators and digital humanities scholars move forward while acknowledging their reliance on one another. By reframing the notion of projects, collections, and research as data assemblages, we can see the dynamic relationship between various groups and productions in the digital humanities.

17

Assemblages in their most basic definition are material component parts that form wholes. These wholes are notable in that they resist totalities and are made up of assemblages themselves. DeLanda asks us to remove the Hegelian focus on relations of interiority because the autonomy of component parts is questionable in a framework that defines their very characteristics based on their relative relationships. Instead, assemblages focus on relations of exteriority. These relations of exteriority are not exhaustive of the ontological properties of a component part. For instance, if an arm becomes removed from a body in a relationship of interiority it fails to be an arm. In contrast, in relationships of exteriority, the arm has the potential to maintain some aspects of itself as it becomes part of a new system of relations [DeLanda 2006, 9–11]. As a result, component parts become fluctuating in assemblages, or as DeLanda states, “a component part of an assemblage may be detached from it and plugged into a different assemblage in which its interactions are different” [DeLanda 2006, 10]. Component parts, notably, also remain in flux and are made up of different assemblages themselves.

18

In a data assemblage, the same focus on materiality allows us to reorganize our thinking of bits and their role in research. As noted earlier, data are material, and their material existence is not contingent on their relationships in curatorial practice. This moves away from earlier notions such as “collections” and “contextual mass” where the relationship between components is so strong that their characteristics within a new framework are problematic. Notably, while data curation allows us to gather bits in an assemblage that can then be used for further research, it perpetuates the perception that curation is done for research purposes rather than something analogous. In data assemblages, this is altered, as the relationships of interiority are no longer characteristic of the material but rather by relationships of exteriority. This allows us to maintain a focus on contextual relationships while still resisting totalities, and also allows us as researchers to focus on some aspects of data, even those assembled by data curators. Scholars can reconfigure this data for their own personal research, and the component’s relation of exteriority changes from one data assemblage (explicitly that of the libraries) to the data assemblage of the researcher (namely the project). Because the results of these data assemblages are not dichotomous but mutually reinforcing, they allow us to more clearly see the link between curatorial and text-heavy work.

19

Assemblage theory is not only notable for stressing the material characteristics of data, but also because it provides an additional framework for the examination of these material parts. Through a framework of three separate dimensions that correspond with three axes, assemblages become more than just a focus on the need to assess fluctuation in ontology. Each axis allows us to investigate the link between digital materials and their corresponding data assemblages. I go through each axis individually using the example of a literature review. While some may contend that literature reviews are only used in conjunction with larger scholarly pieces, this is not always true. Perhaps, most

20

prominently, scholars often publish literature reviews in journals as articles overviewing the state of disciplinary fields, and graduate students sometimes submit them separately before embarking on the dissertation and thesis. Nonetheless, the general framework should be applicable to other forms of print scholarship. While such an exercise may seem tedious, I believe it will more lucidly describe the link between analogue and digital modes of scholarship. In addition, I also emphasize what each axis means in the spectrum of data curation and digital humanities praxis.

The first dimension of data assemblages focuses on the material and expressive roles of components [DeLanda 2006, 12]. The material role of assemblage parts correspond to their physical existence within an ecosystem, and in the case of the literature review, these are the actual books and articles. Often, scholars creating literature reviews will make copies of relevant pages and cut them up into different sections. They then place the different pieces thematically to create sections of an outline. The materiality of conventional research becomes more apparent through this process. In data assemblages, the material roles of component parts exist on one end of an axis with the expressive role on the opposite end. These are the actual characteristics of the books and articles, such as their color and form. These expressive roles should not be understood as inherent characteristics of a component part but rather as also contingent and in flux.

21

The most obvious analogy of material and expressive component parts in current digital media scholarship is within the framework of Matthew Kirschenbaum's data forensics. For Kirschenbaum, like DeLanda, focusing solely on the material structure is problematic. He posits that the material can be further broken down into forensic and formal materiality [Kirschenbaum 2012]. Forensic materiality focuses on individualization and the reduction of physical matter [Kirschenbaum 2012, 12]. Formal materiality, on the other hand, are the "imposition of multiple-relational computational states on a data set or digital object" [Kirschenbaum 2012, 12]. For Kirschenbaum, formal materiality can be understood as the "relative or just-in-time dimension of materiality, one where any material particulars are arbitrary and independent of the underlying computational environment and are instead solely the function of the imposition of a specific formal regimen on a given set of data and the resulting contrast to any other available alternative regimens" [Kirschenbaum 2012, 13]. It provides the illusion of immateriality in the material.

22

Because Kirschenbaum seeks to maintain the ontological distinctiveness of material objects through their reductive individualized traces, he is often ambiguous on the distinction between formal materiality and forensic materiality in things such as firmware. He implores us to not draw an analogy between the formal and forensic and hardware and software, since this hardline distinction is itself a result of industrial practices by companies such as IBM [Kirschenbaum 2012, 14]. Data assemblages remove this ambiguity by not only making the fluidity of the forensic a critical aspect of the material but also by leaving open the possibility of the heterogeneous nature of expressive and material forms. In data assemblages, the material role is exemplified by the electrical impulses in data. The expressive role represents the forms that these electrical impulses take. These roles and materials are constantly shifting allowing for a more formal structure to understand heterogeneous entities such as firmware. As DeLanda notes "these roles are material and may occur in mixtures, that is, a given component may play a mixture of material and expressive roles by conveying different sets of capacities" [DeLanda 2006, 12].

23

The second axis deals with the territorial and deterritorial roles of assemblages [DeLanda 2006, 12–13]. DeLanda cautions us against viewing assemblages as totalities but instead as continuously fluctuating collections of material. Territorialization is the process of component parts of an assemblage coming together. These can be "accidental" and evolutionary in nature or purposeful curatorial acts. Deterritorialization, on the other hand, represents the opposite process of breaking apart assemblages. Again, these parts do not lose their autonomy because their characteristics are not contingent on their relationships of interiority. In the example of the literature review, the reviewer exemplifies the territorializing aspect of the assemblage by placing corresponding works in distinctive sections of the literature review. He or she can also remove parts from different sections, and this exemplifies deterritorialization. Normally, things that we have traditionally seen as having "agency," such as sentient beings, deterritorialize assemblages, but non-sentient beings can also engage in the process. In this sense, data assemblages take seriously the role of quasi-objects and the need to reexamine what Latour refers to as the modernist constitution, which views nature and society as unattached [Latour 1993, 13–48]. For instance, a fire can erupt in the location where the reviewer is conducting his research and cause the work to become ash. In either case, the deterritorialization of the assemblages requires reexamining the

24

relationship between various component parts.

In data assemblages, the territorialization and deterritorialization of component parts are a result of the labor of digital humanities scholars and curators. For instance, as curators organize and collect these assemblages, they create new data assemblages made up of material components. Databases, in this sense, are digital materials that create larger digital infrastructure. Similar to how a scholar creates a literature review, digital humanities projects reconfigure electrical material into assemblages. Projects as diverse as digital historical maps and text mining initiatives are all, in this sense, analogous.

25

The third and final dimension of assemblages corresponds to a coding/decoding axis [DeLanda 2006, 14–16]. In many ways, this dimension should not be considered separate from the territorialization and deterritorialization dimension but as intertwined. Linguistic and genetic phenomena can be seen as part of this dimension. These are special cases that often provide the illusion of the immaterial and work at multiple scales of assemblages. Still, they contain a material functionality. Case in point, DNA represents a physical chromosome structure while language represents neurological phenomenon. These help in territorializing assemblages by giving them distinct linguistic structures. According to DeLanda, language often causes us to lose site of relationships because of the “linguisticity of experience, that is, the idea that an otherwise undifferentiated phenomenological field is cut up into discrete entities by the meanings of the general term” [DeLanda 2006, 45]. In a literature review, such a form of coding exists in the linguistic assertion of the arrangement itself, and the act of a scholar declaring the literature review as complete and ready for submission. A coding function also manifests itself in regards to the transitions and thesis of the literature review itself and highlights the way that coding can affect assemblages in diverse scalarity.

26

In data assemblages, the coding/decoding dimension is exemplified by the assertion of collections as singular totalities and metadata. Research projects and curatorial collections of data are not ontologically separate from one another but become distinct through this linguistic maneuvering. Due to the algorithmic coding of computer software, a territorialization of bits exists that move through the material structure of the computer processing unit. For librarianship, a more pertinent example is metadata that provides the ontological significance of material along with an examination of the connections between these materials.

27

Together, the three dimensions of data assemblages provide an additional framework for the discussion surrounding data curation and digital humanities. It also permits us to examine the various aspects of digital materiality and see their relationship between one another without confining them to relationships of interiority. I believe that such a framework makes it easier to create collaborative opportunities between different groups in the digital humanities because it posits the emergence of data assemblages as the resulting outcome of *both* research and curatorial labor.

28

## Case Study: The Digital Public Library of America

In order to provide an overview of how this framework works with an actual project, I use the Digital Public Library of America (DPLA) as a case study for data assemblages. The organization began in 2010 with forty leaders and sought to make “an open, distributed network of comprehensive online resources that would draw on the nation’s living heritage from libraries, universities, archives, and museums in order to educate, inform, and empower everyone in current and future generations” [About DPLA]. Through the support of the Alfred P. Sloan Foundation and the Berkman Center for Internet and Society at Harvard University, the DPLA launched after nearly two years of planning in April 2013. Although it calls itself a library, it does not circulate any non-digital materials and at this time, does not create new conventional content itself. Instead, it curates the digital metadata of other libraries and cultural heritage organizations and puts them in one place for easy access. Through this aggregating approach, the DPLA has been able to create partnerships with the HathiTrust, the Smithsonian Institute, the Internet Archive, and the Library of Congress. Still, despite its success, measuring the academic impact of the DPLA as a scholarly project remains difficult because the link between curatorship and scholarship is unclear. By reassessing the DPLA as a data assemblage, I seek to demonstrate that the DPLA’s efforts are ontologically analogous to other scholarly activities, such as a literature review.

29

The DPLA is an important case study for the scholarly use of data assemblages for three important reasons. First, created by a largely academic organization housed at Harvard University, it requires constant funding and infrastructural

30

requirements from scholarly resources, and these often cut into and compete from funding for more orthodox academic research. By examining the DPLA's relationship to routine scholarship, we can better make sense of the pitfalls and benefits of sharing and competing for the same resources. Second, created by both academics and other leaders, the DPLA makes apparent the collaborative and similar nature of traditional scholarship and curatorship. This convergence of roles is important not only for collaboration but for future tenure reviews and job evaluations as individuals move more fluidly between academia and infrastructure customarily seen as supporting it. For instance, Dan Cohen, the current Executive Director of the DPLA, gave up his professorship at George Mason University to head the organization. For Cohen, his movement to the DPLA advanced the work he already did in academia at the Roy Rosenzweig Center for History and New Media [Cohen 2013]. Yet, in the current academic climate his work at the DPLA, would not count towards a tenure review. Finally, the DPLA is an important analysis of data assemblages because it outlines many of the same elements in its philosophy, "three main elements," and mission.

The collective nature of the components that the DPLA aggregates exemplifies the material and expressive components of the first axis of data assemblages. The DPLA's "Philosophy" page illustrates the importance of materiality for the organization: "To help organize and structure metadata, the DPLA falls back on some pretty philosophical concepts. Perhaps the most important of these is the idea that things in the physical world can be represented at a number of different levels" [Philosophy-DPLA]. While the DPLA turns each item into metadata that it aggregates, this accumulation also results in the material bits of these items to become flattened ontologically as they move through the DPLA's servers. The expressive component of the first axis deals with the forms of these bits, and by manipulating the arrangement of electricity, the DPLA is able to create an aggregation of metadata that it allows others to handle.

The territorialization and deterritorialization of material bits is, in many ways, the core of the DPLA's mission. On its "About" page, the organization lists "three main elements" that outline the nature of the DPLA. The first of these focuses on how the DPLA seeks to be "a *portal* [emphasis in original] that delivers students, teachers, scholars, and the public to incredible resources, wherever they may be in America" [About DPLA]. In the parlance of data assemblages, this portal is a result of the territorialization of material bits in different libraries and cultural heritage organizations. The second main element of the DPLA demonstrates its role in deterritorializing these material bits. According to the organization, the DPLA seeks to be "A *platform* [emphasis in original] that enables new and transformative uses of our digitized cultural heritage" [About DPLA]. The organization hopes to accomplish this using Application Programming Interfaces (APIs) that permit others to manipulate the its own data. For instance, *Serendip-o-matic*, an application built in a single week for the One Week | One Tool project at the Roy Rosenzweig Center for History and New Media, allows scholars to search for relevant sources in the DPLA using any text they may have [Serendip-o-matic Team 2013]. By doing so, the application lets them deterritorialize material bits on its servers and manipulate them to make curatorship easier and more integral to scholarly research. The importance of examining relations of exteriority also becomes more evident. As Catherine Adams and Terrie Lynn Thompson highlight, "Within a sociomaterial reading, data is not really a thing but rather a relational effect: it is what it is in a particular moment because of the temporal and spatial networks of relations in which it is ensnared" [Adams and Thompson 2014]. As these bits move through various networks, coding and decoding perpetuate the perception of these bits as frozen totalities.

The third axis of coding/decoding corresponds to the process of declaring discrete material entities as unified wholes. The DPLA's main website exemplifies this by fashioning bits in curatorial form as virtual timelines, maps, and exhibits. As mentioned earlier, we should not view the coding and decoding aspects as necessarily different from the territorializing and deterritorializing aspects as both highlight our ability to see things as distinct or compiled. Furthermore, along with manipulating the coding and decoding of its own resources, the DPLA also makes a concentrated effort to understand the linguistic and rhetorical moves that work for and against it as a library. Case in point, in its final "main element," the DPLA seeks to be "an advocate for a strong *public option* [emphasis in original] in the twenty-first century" [About DPLA]. Specifically, it hopes to create a way for people to have access to different sources of knowledge, and it presents itself in a longer line of public libraries in America. This rhetorical move flattens the dissimilarity between digital libraries from more traditional ones.

Still, despite its success, the DPLA continues to face numerous challenges. While the academic community is supportive of the organization's mission, it continues to view the activities as secondary to scholarly research. The lack

31

32

33

34



of a scholarly thesis or traditional format also means that the labor of those in the DPLA may seem less valuable in future tenure reviews or job applications. As a curatorial activity in the library community, the DPLA has a better record in gaining support, although criticisms were prominent before launch [Enis 2013]. As one group of curators notes, “Digital aggregations can provide essential metastructures for unifying distributed content.... however, the act of bringing together and providing access to a large number of collections does not guarantee that the resulting aggregation will be a useful resource for researchers” [Palmer, Zavalina, and Fenlon 2010]. By reexamining the materiality of both scholarly and curatorial activities as data assemblages, the work of both becomes more fluid and the link between the two becomes more apparent and continuous.

## Implications for Future Research and Collaborations

According to Sayeed Choudhury, Mike Furlough, and Joyce Ray:

35

We have on the one hand, a community, or a subset of several communities, that has been working on the “back end” of digital production from the generation of raw data to the construction of an organized product that can be accessed, and, on the other hand, another community — publishers — who work on the “front end” of scholarly communications, from manuscripts to publication [Choudhury, Furlough, and Ray 2012]

Data assemblages provide a structure for future digital humanities research that collapses this dichotomous vision of back-end and front-end digital production. While scholars of the humanities have increasingly pushed for interdisciplinary research, the resulting methodologies have continued to stress endeavors traditionally considered “academic.” This has forced curators and librarians to continuously reassess the work they do with a focus on orthodox researchers. This same attention on research needs, however, has resulted in a glancing over of the importance of curatorial activities as something valuable in and of themselves. Data assemblages, as a result, have broad implications for curators, digital humanists, and the broader digital age beyond the mere theoretical.

36

For curators, the notion of data assemblages allows us a way to view curation as an activity in and of itself. Tenure, in traditional academic settings, has relied on the dissemination of scholarly information, usually in the form of published articles. While such scholarly publications are important, they perpetuate, at least in the humanities, an ideal of the lone scholar working outside of any broader non-academic infrastructure. Curators in this context fail to be given due credit for their work despite being in the academic ecosystem and often must rely on the occasional acknowledgement within a scholarly monograph for appreciation of their labor. In many ways, this problematizes not only the way that conventional scholars view research but also the ways curators see their work. As Trevor Munoz and Julia Flanders write, “data curation is...both a scholarly research area and a praxis and, increasingly, data curation too uses a distinct methodology to advance itself as both a discipline and a profession” [Flanders and Muñoz 2011]. These curatorial activities will be critical to researchers, especially since groups like the data curation community take the link between their work and other researchers to be central to their praxis. The data assemblage model pushes this further by asking us to question the distinction between research and curatorial activities altogether. Because all data assemblages exist on an ontologically flat surface, these activities should not be differentiated from — or at least not seen as antithetical to — scholarly publication.

37

For digital humanists, the data assemblage model, along with causing us to question the nature of scholarly publishing, allow us to create new forms of scholarly activities that are not oriented around traditional narrative structures: maps, games, databases, etc. In other words, the larger digital ecosystem causes us to reexamine our scholarly publishing and the role of diverse media formats in the academy. By viewing scholarly activity as something that must be in the form of print, we perpetuate the notion that narrative structures are the only meaningful forms of communication. Unfortunately, the coding ability of narrative is hard to get away from, and digital scholars should be wary of viewing their activity as separate from data curation and vice versa due to these narrative characteristics. Of course, the value of scholarship is not strictly in the format it takes but also in its ability to stay in conversation with other scholars. Curatorial activities, however, also do this when they seek collaboration with other curators. While these activities are not stated

38

explicitly as scholarly dialogue, we should be wary to discount them as existing in a structural vacuum.

More broadly, the notion of data assemblages allows us to reconsider our place in the digital ecosystem. According to Matthew Kirshenbaum, we must confront the .txtual condition. As he shows, “the preservation of digital objects is *logically inseparable* from the act of their creation — the lag between creation and preservation collapses completely, since a digital object may only ever be said to be preserved *if* it is accessible, and each individual access creates the object anew” [Kirschenbaum 2013]. In other words, a deep examination of digital material shows that access and duplication are inherent to the digital world, and there is no separation of access from creation. We can never really touch the same file because the postmodern nature of computers means files are continuously located in new memory, which are fluctuating electrical pulses. Such electrical impulses are themselves part of a slow evolution into the computer that begins with parts assembled throughout the world. Kirshenbaum argues that the “fingerprint” of the files that we touch in the .txtual condition causes them to change. Of course, the same fingerprints apply to the real world, as criminal forensics has made clear. We see that the assemblage of objects such as books changes along with the data itself, as it is stored in and out of the correlating neurological memory of the observer. In data assemblages, no claims to ontological totalities are made. As a result, the .txtual condition vanishes, or at least becomes temporarily manageable.

39

Furthermore, data assemblages expand on our understanding of ourselves in the digital environment and as researchers. New media scholar Lesley Gourlay argues, “The academy has become saturated by technologies to the point that there can be no meaningful distinction made between digital and analogue, embodied and virtual, ‘face-to-face’ and ‘online’” [Gourlay 2011]. In their examination of digital materialities and posthumanism, Catherine Adams and Terrie Lynn Thompson discuss the need for four *posthuman fluencies* [Adams and Thompson 2014]. The first and second of these fluencies deals with our ability to understand our shared agency with digital technology and the way that these digital technologies lead to a de-skilling and up-skilling of research activity. Data assemblages make this shared agency more apparent since the ability to territorialize and deterritorialize not only deals with sentient agency, such as humans, but also, as in Latourian actor-network-theory, the agency of non-sentient actors such as computers, cameras, and servers. The third of these posthuman agencies deal with the ability for data to become frozen in time for research and the importance of flux for data assemblages. “When data is viewed as frozen *but* [emphasis in original] lively and mobile,” write Adams and Thompson, “new enactments and understandings of data are possible” [Adams and Thompson 2014]. The final of these agencies concerns the importance of understanding the tensions and affordances that come with increased collaboration and fragmentation of research practices. By viewing curatorial and digital humanities scholarship as data assemblages, these posthuman fluencies become more apparent and allow us as scholars to grapple with new questions about our research praxis in the digital age.

40

As digital humanities increases as a field, the need to examine our methodological assumptions will be more apparent. The field has defined itself as a revolutionary force in the academy, and this same revolutionary impulse provides scholars a leg up in collaborative techniques that bridge curatorship with more conventional digital humanities scholarship. As Kathleen Fitzpatrick notes, “the key problems that we face again and again are social rather than technological in nature: problems of encouraging participation in collaborative and collective projects, of developing sound preservation and sustainability practices, of inciting institutional change, of promoting new ways of thinking about how academic work might be done in the coming years” [Fitzpatrick 2010]. This piece sought to outline one way to increase collaboration and acknowledgement by recognizing the material ecosystem surrounding digital scholarship. Still, it is not perfect, and I hope that others will carry out further research and to make the link between the curation and conventional scholarship even more clear.

41

## Works Cited

**About DPLA** Digital Public Library of America, About DPLA. Available at: <http://dp.la/info/> [Accessed September 15, 2014a].

**Adams and Thompson 2014** Adams, C. & Thompson, T.L., 2014. Interviewing the Digital Materialities of Posthuman Inquiry: Decoding the encoding of research practices. In *Proceedings of the 9th International Conference on Networked Learning*. Ninth International Conference on Networked learning. Edinburgh. Available at: <http://www.networkedlearningconference.org.uk/abstracts/pdf/adams.pdf>.

- Anderson 2008** Anderson, C., 2008. The End of Theory. *Wired Magazine*, 16. Available at: <http://www.simulex.it/Materiali/Articoli/Chris%20Anderson-The%20end%20of%20theory.doc> [Accessed July 2, 2013].
- Anderson 2012** Anderson, S. & Blanke, T., 2012. Taking the long view: from e-science humanities to humanities digital ecosystems. *Historical Social Research/Historische Sozialforschung*, pp.147–164.
- Blanchette 2011** Blanchette, J.-F., 2011. A Material History of Bits. *Journal of the American Society for Information Science and Technology*, 62(6), pp.1042–1057.
- Bogost and Montfort 2009** Bogost, I. & Montfort, N., 2009. Platform Studies: Frequently Questioned Answers. In *Digital Arts and Culture 2009*. UC Irvine. Available at: <http://escholarship.org/uc/item/01r0k9br> [Accessed July 12, 2013].
- Borgman 2009** Borgman, C.L., 2009. The Digital Future is Now: A call to Action for the Humanities. *Digital humanities quarterly*, 3(4). Available at: <http://works.bepress.com/borgman/233/> [Accessed July 2, 2013].
- Burgess 2011** Burgess, H.J. & Hamming, J., 2011. New Media in the Academy: Labor and the Production of Knowledge in Scholarly Multimedia. , 5(3). Available at: <http://www.digitalhumanities.org/dhq/vol/5/3/000102/000102.html> [Accessed August 18, 2014].
- Choudhury, Furlough, and Ray 2012** Choudhury, S., Furlough, M. & Ray, J., 2012. Digital Curation and E-Publishing: Libraries Make the Connection. In *Against the Grain Press*, pp. 476–483. Available at: <http://docs.lib.purdue.edu/charleston/2009/OutofBox/2/> [Accessed July 11, 2013].
- Cohen 2013** Cohen, D., 2013. The Digital Public Library of America, Me, and You. *Dan Cohen*. Available at: <http://www.dancohen.org/2013/03/05/the-digital-public-library-of-america-me-and-you/> [Accessed September 15, 2014].
- Cohen and Rosenzweig 2006** Cohen, D.J. & Rosenzweig, R., 2006. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*, Philadelphia: University of Pennsylvania Press.
- DeLanda 2006** DeLanda, M., 2006. *A New Philosophy of Society: Assemblage Theory and Social Complexity*, London; New York: Continuum.
- Denbo2014** Denbo, S., 2014. Committee on Professional Evaluation of Digital Scholarship by Historians. *American Historical Association*. Available at: <http://blog.historians.org/2014/01/committee-professional-evaluation-digital-scholarship-historians/> [Accessed September 15, 2014].
- Duff and Johnson 2002** Duff, W.M. & Johnson, C.A., 2002. Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives. *The Library Quarterly*, 72(4), pp.472–496.
- Enis 2013** Enis, M., 2013. DPLA Launches, Librarians Respond. *The Digital Shift*. Available at: <http://www.thedigitalshift.com/2013/04/digital-libraries/librarians-respond-to-dpla-launch/> [Accessed September 15, 2014].
- Fitzpatrick 2010** Fitzpatrick, K., 2010. Reporting from the Digital Humanities 2010 Conference. *The Chronicle of Higher Education Blogs: ProfHacker*. Available at: <http://chronicle.com/blogs/profhacker/reporting-from-the-digital-humanities-2010-conference/25473> [Accessed September 15, 2014].
- Fitzpatrick 2011** Fitzpatrick, K., 2011. *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*, New York: New York University Press.
- Flanders and Muñoz 2011** Flanders, J. & Muñoz, T., 2011. An Introduction to Humanities Data Curation. Available at: <http://guide.dhcurator.org/intro/> [Accessed July 2, 2013].
- Gibson 1984** Gibson, W., 1984. *Neuromancer*, New York: Ace Books.
- Gourlay 2011** Gourlay, L., 2011. Cyborg Literacies and the Posthuman Text. In *Workshop on New Media, New Literacies, and New Forms of Learning, December. London, UK*. <http://blogs.ubc.ca/newliteracies/files/2011/12/Gourlay.pdf>.
- Gray et al. 2002** Gray, J. et al., 2002. Online Scientific Data Curation, Publication, and Archiving. In *Astronomical Telescopes and Instrumentation*. pp. 103–107. Available at: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=874929> [Accessed July 12, 2013].
- History-DPLA** Digital Public Library of America, History-DPLA. Available at: <http://dp.la/info/about/history/> [Accessed September 15, 2014b].
- Kirschenbaum 2012** Kirschenbaum, M.G., 2012. *Mechanisms: New Media and the Forensic Imagination*, Cambridge, Mass.; London: MIT Press.

- Kirschenbaum 2013** Kirschenbaum, M., 2013. The .txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary. *Digital Humanities Quarterly*, 7(1). Available at: <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html>.
- Latour 1993** Latour, B., 1993. *We Have Never Been Modern*, Cambridge, MA: Harvard University Press.
- Lee 2005** Lee, H.-L., 2005. The Concept of Collection from the User's Perspective. *The Library*, 75(1). Available at: <http://www.jstor.org/stable/10.1086/428693> [Accessed July 2, 2013].
- Lord 2004** Lord, P. et al., 2004. From Data Deluge to Data Curation. In *Proceedings of the UK E-science All Hands Meeting 2004*. pp. 371–375. Available at: <http://www.allhands.org.uk/2004/proceedings/papers/150.pdf> [Accessed July 2, 2013].
- Modern Language Association 2013** Modern Language Association, 2013. Guidelines for Evaluating Work in Digital Humanities and Digital Media. *Modern Language Association*. Available at: [http://www.mla.org/guidelines\\_evaluation\\_digital](http://www.mla.org/guidelines_evaluation_digital) [Accessed September 15, 2014].
- Munoz 2013** Munoz, T., 2013. Data Curation as Publishing for Digital Humanists. In CIC Center for Library Initiatives Annual Conference. Columbus, OH. Available at: <http://trevormunoz.com/notebook/2013/05/30/data-curation-as-publishing-for-dh.html>.
- Munoz and Renear 2013** Munoz, T. & Renear, A., 2013. Issues in Humanities Data Curation. Available at: <http://cirssweb.lis.illinois.edu/paloalto/whitepaper/premeeting/> [Accessed July 2, 2013].
- National Endowment for the Humanities** National Endowment for the Humanities, Digital Humanities Start-Up Grants. *National Endowment for the Humanities*. Available at: <http://www.neh.gov/grants/odh/digital-humanities-start-grants> [Accessed July 12, 2013].
- Palmer, Zavalina, and Fenlon 2010** Palmer, C.L., Zavalina, O.L. & Fenlon, K., 2010. Beyond Size and Search: Building Contextual Mass in Digital Aggregations for Scholarly Use. *Proceedings of the American Society for Information Science and Technology*, 47(1), pp.1–10.
- Philosophy-DPLA** Digital Public Library of America, Philosophy-DPLA. Available at: <http://dp.la/info/developers/codex/policies/philosophy/> [Accessed September 15, 2014c].
- Rosenzweig 2003** Rosenzweig, R., 2003. Scarcity or Abundance? Preserving the Past in a Digital Era. *The American Historical Review*, 108(3), pp.735–762.
- Salo 2011** Salo, D., 2011. Respect My Praxis. *#alt-academy: Alternative Academic Careers*. Available at: <http://mediacommons.futureofthebook.org/alt-ac/pieces/respect-my-praxis> [Accessed July 11, 2013].
- Serendip-o-matic Team 2013** Serendip-o-matic Team, 2013. Serendip-o-matic: Let Your Sources Surprise You. *Serendip-o-matic*. Available at: <http://serendipomatic.org/> [Accessed September 15, 2014].
- Svensson 2010** Svensson, P., 2010. The Landscape of Digital Humanities. *Digital Humanities Quarterly*, 004(1).
- Unsworth 2001** Unsworth, J., 2001. A Master's Degree in Digital Humanities: Part of the Media Studies Program At The University of Virginia. In Congress of the Social Sciences and Humanities 2001. Québec, Canada.
- Van Zundert 2012** Van Zundert, J., 2012. If you build it, will we come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities. *Historical Social Research/Historische Sozialforschung*, pp.165–186.
- Wachowski 1999** Wachowski, A. & Wachowski, L., 1999. *The Matrix*,
- Yakel 2007** Yakel, E., 2007. Digital Curation. *OCLC Systems & Services*, 23(4), pp.335–340.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.