Scholarly Infrastructure as Critical Argument: Nine principles in a preliminary survey of the bibliographic and critical values expressed by scholarly web-portals for visualizing data

Jo Guldi <jguldi_at_mail_dot_smu_dot_edu>, Southern Methodist University

Revision Note

This article has been revised since its original publication. The language describing Zotero has been updated. The updated portion is here. The previous version of the article will remain available.

Abstract

What values does infrastructure-building represent? This article begins by situating scholarly practices around infrastructure within a broader transformation of twenty-first century life and indeed scholarship and learning by infrastructures, and distinguishing "scholarly" infrastructure from other kinds of infrastructure designed to share information that nevertheless lack scholarly engagement with analysis. This article compares the role of scholar-builders in crystallizing a set of theoretical concerns, data, and analyses to that of the architects of opera houses during the golden age of European opera, who structured, illuminated, and constrained possible future creations of art. The article next attempts to excavate a set of implicit values, while making room for the possibility that the list of values put forward here is only incomplete, and that the list of values itself is the subject of potential debate, critique, or dissent, some of which may take the form of building infrastructures differently than the patterns laid out here. A first section outlines a set of *bibliographic* values; while the article's second half turns to the power dynamics of infrastructure and a set of *critical* values encompassed by particular projects, before turning to the issue of why understanding these values is so essential to understanding infrastructure projects as a form of scholarly production that merit support and recognition by the community at large.

We live in a world connected not merely by technology but also by infrastructure, which seamlessly delivers data to our doors. What's more, in today's university, many a professor has become a designer of infrastructure as well as a consumer. While portals of this kind are developed by a minority of practicing scholars, they form one of the major digital complements that scholars tend to create for their traditional work. The reasons for this are seldom explicitly articulated in the existing literature that documents their use and purpose. The creation of public tools for visualizing data is increasingly important to many scholars, and it represents a non-negligible commitment in terms of time and grant money. [1]

This article argues that that choice and commitment are best understood as a kind of work deeply informed by an engagement with certain *values*, including scholarly transparency and political participation, which often align with scholars' moral commitments, intellectual heritage, and disciplinary traditions. Indeed, that alignment has been so powerful for many builders of scholarly data that they have chosen to pursue the building of infrastructure projects even when that choice was otherwise out of alignment with recognizable promotion incentives in their professional communities. [2]

In many cases, the infrastructure project advances an argument: but if web portals existed merely to convey arguments to a broader audience, they could just as easily take the form of an open-access article or monography. Why, then, [3]

would a scholar or a scholarly team choose to invest their energy in designing and constructing a web project?

The pages that follow then attempt to answer the question: what values does infrastructure-building represent? This article begins by situating scholarly practices around infrastructure within a broader transformation of twenty-first century life and indeed scholarship and learning by infrastructures, and distinguishing "scholarly" infrastructure from other kinds of infrastructure designed to share information that nevertheless lack scholarly engagement with analysis. The article next attempts to excavate a set of implicit values, while making room for the possibility that the list of values put forward here is only incomplete, and that the list of values itself is the subject of potential debate, critique, or dissent, some of which may take the form of building infrastructures differently than the patterns laid out here. A first section outlines a set of "bibliographic" values; while the article's second half turns to the power dynamics of infrastructure and a set of "critical " values encompassed by particular projects, before turning to the issue of why understanding these values is so essential to understanding infrastructure projects as a form of scholarly production that merit support and recognition by the community at large.

The principles and case-studies here suggest that infrastructure represents a conscientious choice about the format for disseminating that argument that reflects critical thinking about the value of scholarship. In designing infrastructure, many scholars today are amplifying the values embedded in their scholarship – whether values of accessibility, of replicability, or even of the political critique of institutions – and that for this reason, infrastructure needs to be looked on as a particular kind of gesture, one that not merely critiques the world but attempts to actively remake it. The author's personal history as a theorist/historian of infrastructure and sometime builder of infrastructure are offered by way of laying out the importance of the scholarly community engaging with the critical issues represented by a choice to build – and the hazards to professional careers that can occur when activities of this kind are undertheorized, underengaged, or otherwise underappreciated. Infrastructure thereby deserves to be elevated within a humanistic discourse that privileges action as well as contemplation.

Of Infrastructure

Web portals that visualize data may strike many scholars as a strange place to begin a discussion of infrastructure. Visualizing data, of course, is but part from the whole, even as a physical mailbox is invisibly connected to a network of roads and trains, law and labor, standards and trust — so familiar that it too disappears into the background of landscape taken everywhere for granted — which allows letters posted in one continent to be delivered promptly to another quarter of the world within days. The colossus of infrastructure that has developed in university libraries and digital humanities centers to make possible the scraping, cleaning, and serving of data is the subject of a great deal of documentation, much of which takes the role of making visible otherwise invisible commitments of expense, labor, maintenance, and ideation.

As understood in its broadest form, infrastructure encompasses any use of technology or information which connects communities. Science and technology scholars such as Janet Vertesi have defined infrastructure in terms of a technology or management structure distinguished by its deliberate design, its scale, its capacity to deliver information, and, frequently, its invisibility [Vertesi and Ribers 2019, 263–4]. In the digital age, new infrastructures have rapidly transformed how ordinary citizens as well as scholars access information. Individuals read e-books, consult maps, and digest newspapers whose delivery depends upon a vast infrastructure system ranging from transoceanic cables to smart phones [Starosielski 2015] [Sawyer et al. 2019].

Meanwhile, scholarship too has grown to depend increasingly on a world of invisible technologies. Already in 2010, Geoffrey Rockwell described an "infrastructure turn" that was shaping how a majority of scholars accessed data [Rockwell 2010], which entailed new worlds of labor, storage, and expense for universities, much of which required new skills at managing digital data. Tara McPherson, founder and editor of a journal that provided one infrastructure for delivering digital projects to readers, described the creation of a "broader culture of experimentation and change" among faculty, particularly those with a specialty in visual analysis, as requiring new developments in infrastructure that would service, validate, and share faculty experiments – from accessible to journals tenure and review processes [McPherson 2010].

Indeed, a broad range of institutional players came into being at most universities and within many disciplines, including digital humanities centers, digital humanities majors, institutional repositories, and new guidelines for dissertation-writing and for the tenure and promotion of faculty [Gold 2018] [Clement et al. 2017] [Flanders and Hamlin 2013]. Libraries began to engage data storage, standards that define the data, the data itself, metadata, and applications that retrieve and serve the data, as well as hardware, software, and structures for the management of labor [Mattern 2014].

Researchers can turn to an emerging field of similarly interoperable tools that put quantitative analysis of texts within the grasp of textual scholars from any tradition in the humanities or social sciences. Among those tools are *PhiloLogic*, UChicago's browser for topic-modeling text, analyzing word co-occurrence and re-use; *InPho Topic Explorer*, which allows users to apply topic modeling to their own body of text; *Voyant*, a simple web-interface which allows the user to pipe a selection of text through a variety of visualizations; *Palladio*, Stanford's application for visualizing correspondence such as letters within a social network; *CorTexT*, Jean-Philippe Cointet's extremely flexible and strong tool for language-based analysis on a body of text with temporal metadata; and *Archives Unleashed*, Ian Milligan's toolbox for digital humanities and social sciences analysis of archives.

There are also, of course, individual repositories of texts that offer particular tools for a unique set of documents, especially JSTOR's data for research portal, the online instance of the *Old Bailey*, as well as an expanding range of online portals for individual moments in time; those covered in this article include *Slave Voyages*, the *Legacies of the British Slave-Ownership* project, and *Torn Apart/Separados*, a project that documents the contemporary incarceration of immigrants and their families in the U.S. Effort will be made in this discussion to foreground the author's personal experience of the successes and failures of the *Paper Machines* experiment, and the strengths of the infrastructure that supplanted it, towards a genealogy of the design principles appropriate to infrastructure in the humanities and social sciences.

Even more recently, Roopika Risam has argued for the existence of a postcolonial DH, marked by an "ethos of building," where scholars create infrastructure to reconfigure access to suppressed voices that document the experience of colonization. She points to several examples of projects where scholars used infrastructure to reinterpret and challenge colonial world-making projects implicit in archives, by assembling alternative archives of information that supported research on the experience of colonized persons. Risam gives numerous examples that document the variety and importance of this work, including The Early Caribbean Digital Archive and the Online Tagore Variorum, which reassemble scattered texts documenting the experience of colonized persons, as well as Chicana por mi Raza, an archive of oral histories, ephemera, out-of-print publications that provide resources for Latina/o experience, and The Indigenous Digital Archive in New Mexico, which made available the records of Hope and Diné experience with boarding schools. Infrastructure in this case is at work reassembling the scattered voices of colonized persons, constructing alternative archives that challenge and actively provide alternatives to libraries whose holdings predominantly reflect the information matrix of the colonizer. In Risam's eloquent phrase, "The archive uses technology to push back against dominant cultural representations of indigenous communities" [Risam 2019, 36].

Scholarly creation of infrastructure has become a primary realm of critical encounter with public narratives about culture, identity, and the past. Scholars who have engaged this space have forged an entirely new modality of scholarly publishing, one with tremendous power to shape public discourse. It is not clear, however, that all forms of infrastructure constitute a critical intervention in scholarship. The Internet Archive is a stalwart repository for preserving the historical artifacts, the creation of which surely constituted a political act in its time. Many of its sub-collections – for instance, troves of anthropological films collected by universities over decades — represent some work of critical scholarship. Nevertheless, it would be hard to argue that the Internet Archive's little-curated collection of historical objects represents a scholarly intervention in any particular field. While not a scholarly piece of infrastructure by design, the Internet Archive's voluminous collections of scanned books, archival videos, and recordings nevertheless consists of a "potential "site of future critical interventions such as OCR, metadata, preservation, and collection analysis. Thus the distinction between scholarly and non-scholarly infrastructure is not hard and fast.

In a critical review of existing digital humanities and social sciences projects, I have attempted to derive an overview of the values that motivated those projects via their expression in public-facing websites. The idea that a piece of

infrastructure may encapsulate scholarly values is premised, in the first place, upon the implicit understanding that the designers of infrastructure have already made critical argumentation, embedded within the larger culture of critical thinking about technology and culture, and inspired by its insights. In other words, this article advances the proposition that the design of infrastructure, when it engages critically with existing thought about power, represents a form of scholarly argumentation.

In the following sections, this article will take the first step towards articulating some principles that define the culture of infrastructure. These principles are divided in two parts, reflecting the distance between the concepts of "scholarship" and "criticality" that I have teased apart in the introduction. A set of "bibliographic" principles of infrastructure describe the principles that bring data into alignment with traditional bibliophilic concerns about verifiability and fact in the liberal arts, including web tools that link directly back to particular manuscripts in particular archives. A set of "critical" principles of infrastructure describe the kind of interventions that render the invisible visible, and which otherwise mirror the kind of critical thinking called upon by Arendt and others as a liberatory intervention in the public sphere.

The principles advanced in this essay are merely a starting point – a review of possible sources, both from the traditional humanities and social sciences and from critical politics, which begins with an autobiographical narrative of the author's own participation in the design and building of a piece of infrastructure. It is hoped that this starter list of principles will be amended and argued with, towards the creation of a more critical discourse about what humanistic infrastructure looks like and where critical thinking takes place, with the expected outcome of generating a scholarly discourse about the choices made in designing infrastructures for the public. An essay of this kind affords an opportunity to reflect on the public purposes and intellectual orientation intended in the building of infrastructure, and the extent to which humanities and social science critiques of the flow of information and power in society have resulted in the creation of alternative or transformative flows.

Infrastructure as Proto-Argumentation

In enabling the researcher to efficiently identify the patterns that characterize a vast scale of documents, infrastructure can seed dozens or hundreds of scholarly arguments. The infrastructure thus represents a *proto-argument*, if not an argument or analysis itself.

What do I mean by *proto-argument*? An infrastructure contains constraints that structure and direct the format of the argument eventually produced over them, even as the shape of opera houses in London and Turin framed civic ambitions, structuring how later compositions, libretti, and performances would be received; those opera houses hosted hundreds or thousands of later productions.

In relationships of this kind, the architect's work was often to envision and crystallize a set of intellectual values and ideological ambitions latent in opera. As historians of the opera have reasoned, the work of the architect was to offer a framework of imagination that included "civic ambition," "unabashed commercialism," and "international cosmopolitanism," which ran in parallel to textual forms of argument-making about history, civic identity, and teleology [Aspden 2019, 6]. Indeed, the opera house structured the reception of operas within a "penumbra of exclusivity" that mirrored the social structuring of other forms of enlightenment knowledge [Aspden 2019].

Hence the "proto" in *proto-argument*. The architect's work codified enlightenment values into a structure that would both enhance later performances, extend the potential of the. The creation of the vessel represents the moment in time when an ideology is crystallized into a form where it can be replicated. Much of the work of conceptualizing what *Aida* — as an epic about the tragedy of gender, race, and empire — might be was thus the work of a generation of architects and opera-designers who worked previous to Verdi's masterpiece. Similarly, the infrastructure allows the application of one *theory* of text or temporality to a million data sets.

Architecture and infrastructure both function simultaneously as creation, vessel, and constraint. They are proto-arguments in the sense that they structure the natural course of use, exchange, and argumentation that follows from an original design. One cannot stage a Brechtian play in an opera house. In a state of disaffection, one might *imagine* dissolving the fourth wall from the opera house at Turin, but one would have to build a totally different kind of theater in

order to realize it.

Similarly, scholarly infrastructures that curate particular choices of word analysis – grammar, topic, bag-of-words; transparent or not, open to interpretation or not – package an entire system of understanding about what a text is, what components of grammatical and semiotic structure or historical context matter, and who is allowed to participate in the interpretation of text. Scholarly infrastructure can be designed in such a way as to elucidate the critical function of scholarship or to diminish it; and to expand public access to polysemous interpretations or battles over the historical record or to channel them. All of those choices have enormous downstream implications for what scholarship can be executed, published, or received.

Builder-scholars, over the last decade, have made legion critical decisions that open up corpora to possible kinds of engagement. Google Ngrams does not allow the scholar to inquire about word co-location as PhiloLogic does; Cortext allows the user to highlight discontinuities in time in a way that few other resources do; each innovation in analysis represents an opening to a critical question, informed by concerns in linguistics or history, bringing the theories of disciplines into alignment with algorithms and corpus in such a way as to *scale* theory into an analysis with new documents.

Builder-scholars are also, as I shall argue below, responsible for making choices that govern who has the opportunity to inspect, to use, or to interpret data, thus structuring entire communities of knowledge. The builder's decisions – to model gender or not, to open the structure, to an enormous degree, whether and how fluidly later scholars who use the infrastructure will be able to compare the use of gender, race, or class in the text; whether hours, days, years, or decades of change over time are susceptible to analysis; and who is allowed to make that analysis – whether a scholarly elite or anyone at all.

To embrace this perspective made atop a web portal is to highlight the hybridity of labor in discoveries, where both the design of infrastructure and the scholarly analysis built upon those discoveries are crucial. When an award-winning paper is published that relies on the metrics of gender difference encoded in historical data made available by a public-facing portal, is the award properly given to the paper's author, or to the website's designer, who identified gender difference as a key component of that data, fit the data with the correct algorithm for analysis, and made the results available? Arguably, the award would properly be given to both: not only because the paper would be impossible without the labor of the data, but also because a large portion of the critical perspective implicit in the paper was given in advance by the infrastructure.

Above all, this line of reasoning underscores that the design and building of portals is itself an act of imagination, and possible critique. In the following section, I will propose some general lines of reasoning about what kinds of imagination and critique the present generations of portals embody.

The Bibliographic Principles of Scholarly Infrastructure

The Principle of Transparency with Respect to a Document Base.

One challenge of interoperability is making sure that the origin and nature of artifacts is preserved, even as information is exchanged. A variety of tools – including *Omeka* and *Fedora Commons* — make possible the tracking of metadata about archival objects [Gourley and Battino Viterbo]. A user in the humanities needs always to know on what documents and particular words a visualization is based. Only in this way can the user defend herself against the critical challenges of readers who will want to know if a homonym has been used in different sense, if a single word example counted by a computer is based on an OCR error, or if the statement counted as an earnest example of nationalism was made in jest.

For this reason, it is crucial that humanities infrastructure be designed so that every visualization is transparent. Each visualization based on data in an archive should be clickable; each word or point on a chart should link to another page that cites the documents and passages represented by that analysis. Ideally, each list of citations should link back to the snippet of original text that was counted to create the analysis. Where copyright allows, that snippet of original text would link back to the original scan of the document itself.

The University of Chicago's *PhiloLogic* platform is a paragon of this value: visualizations of keyword prevalence over time can be linked to original passages in the text, which link back to chapter and edition, with the result that the scholar is always clear about upon which original texts a visualization is based, and how those individual parts of the text relate to a larger text and cycle of publication.

Paper Machines singularly failed with regard to this principle of clickability, but other infrastructures are more satisfying. In *PhiloLogic*, each visualization of word counts over time links to a list of documents and passages that links to the original text in its context. In the *Old Bailey*, keyword searches and crime searches link back to passages that link to the original scans of the document. The user is never deluded with respect to the evidence and where it comes from. Adherence to the Principle of Transparency with Respect to a Document Base ensures that users, both scholarly and vernacular, can link abstract word counts to original text, whether by looking up the citation given or through a link to a verbatim copy of an original text.

The Principle of Interoperability with Existing Documents.

A basic orientation of Paper Machines that the prosthesis must be interoperable with the material that scholars already had on their computers. Despite the range of national and temporal orientations, every research scholar I knew worked with pdfs, which had recently become the standard digitized format for journal articles and scanned books. A single application would accept pdf's and churn out a text version of the article (presumably with the noise of headers, footers, and page numbers intact). Despite the noise, the information encoded in the text would be sufficient for us to design the application so that it could generalize about collections of pdf's, for instance creating word clouds that suggested the transition of pdf's in a single field from one decade to the next.

Paper Machines was thus designed to operate within existing infrastructure – depending on the folders of pdf's that users already had on their machines. I originally imagined it as a kind of box, into which a user would drag folders of documents, with the result of producing visualizations about the contents of those folders. Eventually, my co-designer, Christopher Johnson-Roberson, proposed a tool that would scrape text and generate visualizations from inside another existing piece of humanities infrastructure: *Zotero*, the application that many humanists use to organize their citations for books, articles, and archival materials. Inside *Zotero*, *Paper Machines* operated as a right-click menu that offered a set of commands by which a researcher could extract text from a file, or order that text to be generalized into topic models, word clouds, or networks of phrases.

Today, a variety of humanities applications, including *Palladio* and *Cortext* allow users to upload their own data, typically accepting files in the form of text or Excel document. Tools like *Voyant* and *InPho Topic Explorer* allow scholars to cut-and-paste raw text from any source. PhiloLogic works with any TEI-encoded text sent directly to its editors for uploading. Portals of this kind offer an invaluable resource for scholars with little technical background who wish to take advantage of existing digital tools to analyze text in a given form.

The Principle of Neutrality with Respect to Technical Literacy.

Consider sociologist Matthew Desmond's website, Eviction Lab, which currently serves data on local eviction rates to members of the public, directly feeding local housing campaigns, but also providing a solid stream of data to journalists and legislators via the site's API. Desmond's web portal is "friendly", we might say, in the sense of serving data within a sparsely-designed interface, set to a map of the United States; users can zoom into any locality, and explore rates of eviction and demographics for that site, as well as comparing one site to others. The portal delivers information; the cleaner the interface, the easier to use, and the more attention the portal attracts.

"Friendliness" is thus a function of obscurity, in a sense: more accessible than a monograph or scholarly journal, the portal obscures expertise; there are no footnotes or citations and few allusions to other organizations. One dimension of this friendliness is the obscuring of other forms of infrastructure – the many organizations whose work contributed to the gathering of the data; the many organizations at work, some of which initially complained at a lack of acknowledgement from Eviction Lab [Aiello et al. 2018]. Indeed, as we have seen, one function of infrastructure is frequently its invisibility: the cables that carry data across the ocean are submerged; users of libraries rarely consider the labor and investments

that created those institutions. Yet obscurity is not the only function of friendly design; rather, a clean and legible design also serves the function of inviting participation.

Friendly design tempts users to investigate a dataset or tool according to their own peculiar motivations, which have only been anticipated by the interface's designer in a general way. Indeed, one aspect of the accessibility of web portals is to obscure many of the acts of accreditation and comparison that are typical of journal footnotes and published monographs, explaining the conclusions that a scholar came to on the basis of data. Instead, in Eviction Lab, the user is invited to make their own comparisons between any two geographical entities, examining rates of eviction, poverty and race between a given city and the rest of the state where it is located.

Another dimension of the seeming accessibility of the data is thus that the uses presented to the public are potentially open-ended: the scholar who shares data about rent, city-by-city, cannot know the political ends to which arguments about that data will be put. Data about rents is interpreted in one way by advocates of the free market and another way by advocates of welfare services; but without facts to argue about, the discussion is won merely by the fact of identification with an ideology, and thus by authority, rather than discourse, debate, and reason.

Faith in the *process* of such a discourse to produce understanding, particularly when all players have access to original texts, is part of the legacy of the enlightenment. Indeed, encouraging contemporary people to engage the shared reality of the past – to interpret it and come to understanding on their own, to begin to understand the present in relationship to shared historical experience, however naively at first – is surely part of the core duty of scholars and universities. Builders can limit the extent to which users of topic models or other abstractions can away from historical truth by linking visualizations back to the original text – a strategy we discussed in the Principle of Transparency – such that a debate on social media might occur that took the form of a series of refutations grounded in precise arguments about what to make of particular passages of texts.

In the humanities, it is possible that certain scholars experience an aversion to the principle of sharing learned documents beyond the sacrosanct borders of those initiated into the elite habits of hermeneutics and historiography. An anecdote shared by one scholar/builder suggests that a grant to release a public topic model of Jefferson's correspondence was crushed, at one point, by a reviewer who feared that political forces in the public would demonize Jefferson on account of topics that referred to slaves.

Sharing the tools of abstracting and contextualizing arguments in their original context is surely part of the process of helping the public to perceive the past as accessible and a source of reason or rational debate to begin with: it is tools that can be argued about, linked to original passages of text, that can prove to naïve readers that the past is something about which rational arguments can be had, where arguments about the past have to be backed up by reasonable citations of an authoritative text stream. Engaging the tools doesn't relinquish the battle over interpretation; rather, it moves the battle of interpretation out of the arcane world of footnotes, into a world of (preferably transparent) graphs and passages of text that any literate person can begin to understand. I shall return to this question of how tools help contemporary societies to establish a shared understanding of their past and present in the discussion of a later section, the Principle of Creating Knowledge About the Social World.

Today, portals such as *Philologic*, *Topic Explorer*, *Voyant* and *Palladio* make readily available a graphic user interface for documents brought in by the user, wrought out of precious seasons of user-testing the redesign, with the purpose of delivering a tool easy to use to those with no training. Single-archive portals such as the *Old Bailey* offer multiple transformations so that a user can explore the history of court cases by gender, crime, name, and other archive-supplied variables. All of these tools aim on bringing the facility and nimbleness of code to a base of users whose priority is research, rather than mastery of the architecture of coding.

In 2010-2, I worked on a similar piece of infrastructure – *Paper Machines* – which could be embedded within the existing Zotero portal for storing documents, allowing users to generate word clouds and other visualizations for the pdf's organized on their computer. Our starting assumptions with *Paper Machines* departed from a moment when most scholars in the humanities lacked the technical expertise to convert files from pdf to clean text. We anticipated that scholars would want to harness the tools of aggregating information across thousands of the pdf's that each of us

already had on our desktops, and that Google Books had just made available in enormous numbers. In order to accomplish this, scholars would need access to highly interoperable tools. By the nature of the community in which I was embedded, any tool I worked on necessarily needed to bring insights to a plurality of scholars who worked on a range of topics, periods, languages and nations, who possessed a range of technical proficiencies from ample to extremely limited. Such an imagined user base underscored the importance of working with existing infrastructure, of transparency and usability as values. The idea is to present a low bar of technical literacy with the end of enabling all users – regardless of their background – to replicate a scholarly analysis, or to run their own.

In the case of our work, it was vital that *Paper Machines* serve the largest number of users with the smallest technological interventions to produce an easy transformation of the data into visual form, organized so as to be as convenient as possible to novices. It was intended to be facile and transparent, easily installed and easily opened.

<div style="float:right">43</div>

Extended beyond *Paper Machines*, this principle raises questions about the design of library and museum catalogues, national newspaper collections and archives of political debates, and indeed any collection of archives whose user base includes a general public which may not be literate. If the institution wishes to justify itself to a public taxpayer-base, to create transparency about national history or the nature of its political system, then builders of humanities infrastructure must follow the precepts of user interface design to enable the movement from question to analysis through a transparent, largely visual interface.

<div style="float:right">44</div>

Indeed, the barriers to technological literacy may be more vast than the infrastructure itself. As Miriam Posner has argued, barriers may be institutional and cultural rather than defined by software. Successful digital humanities centers typically require a great deal of independence from university administration — as well as support — in order to provision researchers with the tools of committed research [Posner 2013].

<div style="float:right">45</div>

The Principle of the Primacy of Pattern Recognition.

Today, the designers of algorithms can borrow from an immense assortment of possible clustering algorithms from Machine Learning and statistics. It is not clear, in all cases, how the transformation of language encoded in these processes are applicable to humanities concerns. Put simply, pattern recognition is the coin of the realm of the humanities and social sciences, structuring insights from linguistics, history, archaeology, sociology, and political science, since at least the age of Ernst Cassirer.

<div style="float:right">46</div>

In the design of Paper Machines, algorithms were chosen that would foreground pattern recognition and make available insights to which scholars could give their assent – on the basis of relatively easy-to-explain insights from computer science. My graduate school education coincided with the linguistic turn in British history, and my first point of orientation to the digital humanities was the possibility of measuring language as an index of political change. Another basic orientation was that we would foreground forms of analysis that already replicated, insofar as possible, the sorting and pattern-recognition functions of ordinary scholars of cultural, social, or political language.

<div style="float:right">47</div>

Establishing and foregrounding the patterns being measured in a text has a clear benefit: it underscores examples of human behavior that other observers can easily recognize, grounded in facts that can be shown to correlate with the analysis. For this reason, pattern recognition is the gold standard of analysis in the humanities and social sciences.

<div style="float:right">48</div>

For this reason, likewise, drawing attention to cultural patterns has transformative potential because it directly relates analysis to fact. Inspired by the overwhelming power of cultural criticism based in patterns, digital humanities students at Berkeley addressed the #blacklivesmatter movement by creating an Online Hate Index (OHI) that would recognize patterns in twitter and social media [Vacano et al. 2017]. Such a tool, in theory, has transformative political potential as much as any Frankfurt School analysis of culture, on the same principle: that enlightenment can be tagged to self-conscious reflexivity about cultural phenomena, and that awakening, therefore, can be harnessed to the recognition and analysis of patterns.

<div style="float:right">49</div>

Not every online archive necessarily helps readers to navigate by using visualizations that foreground repeated patterns in the text. Earlier schools of infrastructure design – for example, many hyperlinked scholarly editions, web-born exhibits

<div style="float:right">50</div>

of art history, and keyword-searchable web repositories too numerous to mention – simply delivered text or visualizations to the user, one list or item at a time.

The skills of rendering insights from repeated patterns are associated with the school of Cultural Analytics, especially publications in the *Journal of Cultural Analytics*, which has begun, in recent years, to harvest the most successful examples of pattern-recognition aided by advances in statistics and machine learning [Bamman et al. 2014] [Kraicer and Piper 2019] [Lansdall-Welfare et al. 2017] [So et al. 2019] [Underwood et al. 2018]. The natural next expression of the methods laid out in these journal articles – which count, for example, named subjects by gender and race, or identify words associated with gender and race, or cluster documents temporally – is for those toolsets to be made usable by the public through the design of portals that would harness the full capability of our present knowledge of pattern-recognition. | 51

To foreground the tools that draw attention to patterns based on repeated use of words – from basic keyword-counting to named entity extraction to topic modeling – reduces the technological intervention to something grounded in theories familiar to scholars of language such as Wittgenstein, Cassirer, Laswell, Austin, Skinner, Koselleck, and their heirs. A pattern-recognition toolkit is relatively transparent to users with respect to its possible uses and implications. At the same time, insisting on the primacy of pattern recognition must not come into conflict with the reality of the plural desires of humanists. | 52

For those insights to be useable, and to garner consent, it is not enough that the types of analysis used be composed of concepts familiar to humanists and social scientists; they must also be replicable. | 53

The Principle of Replicability.

A concept introduced in another article in this issue, the "pipeline" promise that the results of working with any given text base can be generalized. A term borrowed from computer science, the pipeline suggests a structure that enables the transformation of data through particular modules, each of which is completed before the next takes over. Conceptually, describing the way that a particular data visualization was produced through a pipeline underscores the need to transparently document the choices that comprise any particular approach to data. Sharing a pipeline, in the form of open data or open-source code, means being fully transparent about how a scholar arrived at a particular conclusion. Sharing pipelines also opens the gates to constructive scholarly critique, for instance, inviting other scholars to run the same code on the same data, but with a different cleaning algorithm or clustering equation, to see how different the results might be. | 54

Scholarly journals in the humanities have begun to make this kind of sharing possible by adopting relationships with existing infrastructure that supports the sharing of code and data alongside shared ideas and visualizations. For instance, the *Journal of Cultural Analytics* asks authors to contribute their data and code to the Harvard Dataverse, an existing open-source infrastructure that allows a user to share and label datasets and code so that they are transparent and findable. The *Hansard Topic Relevance Identifier*, *HaToRI*, discussed in another essay in the same issue of *Digital Humanities Quarterly* – aspires to the "pipeline" concept documented here, where every choice in the production of an analysis is documented, and any choice can be questioned, with the result of producing different material. | 55

Sharing pipelines encourages inter-disciplinary, inter-temporal, and inter-national replication of results, comparison, and extrapolation. A pipeline designed around the Hansard debates for Great Britain can be plugged into the EU debates, the debates of the American congress or NYC council, with the same transformations executed on the same infrastructure. | 56

Building infrastructure in this way makes good on the promise typically ascribed to *critical theory* in the humanities and social sciences since the era of Hegel and Marx, where comparisons of social phenomena from diverse places and times can be rendered understandable and compared in aggregate. Infrastructure designed to work only with a single group of texts does not do this, but infrastructure that is interoperable with a great range of material in theory allows vast comparisons and contrasts that will enable the next generation of social theorists, theorists of literature, and theorists of society to draw the generalizations that will structure our knowledge of the world. | 57

Eve Kraicer, Andrew Piper and other scholars are pursuing an intellectual agenda in the design of code pipelines that make possible the critical reading of gender, class, and race in large-scale bodies of texts [Kraicer and Piper 2019]. Their work differs from traditional endeavors, of course, in that it is replicable – that is, the body of code that Kraicer and Piper apply to American best-selling novels since the 1970s can, with the adjustment of perhaps five lines of code, be applied to Victorian novels or Renaissance broadsides, assuming clean data has already been prepared.

Potential replicability is underscored where scholars share actual code – or create a web portal that allows a body of code to be applied texts supplied by the scholar. A scholarly body of code, such as historian Lincoln Mullen's "textreuse", a toolset for comparing large bodies of text, opens doors for further scholarship, thus matching in significance works of critical theory, potentially rendering Mullen a contemporary Bourdieu. In *InPho Topic Explorer*, philosopher Colin Allen provides an interface where scholars can submit any body of text to topic modeling, thus allowing humanists and social scientists access to a preferred tool for the machine-enabled clustering of documents.

Drawing attention to the wealth of postcolonial projects in the space of the digital humanities, Roopika Risam has underscored the way that digital technology allows scholars not merely to assert, but to replicate a critique of empire through the assemblage of alternative archives. Reviewing numerous examples, she shows how scholars have created and share alternative archives that run in parallel to the maps, laws, censuses, and other documentation of empire. Only through building — not merely through scholarship — she reasons, can scholars truly challenge the dominance of the colonized voice [Risam 2019]. Unlike a one-time publication, which might critique the official archive without enabling other scholars the tools of directly accessing more voices, infrastructure projects of this sort actively provide the work for a hundred or a thousand new research projects and student papers.

This potential replicability of a theory across data-sets represents the fulfillment of the universalizing principles of the humanities since Vico, that is, the desire to make sense of multiple times and places. At the same time, the critical theory impulses of Piper and his cohort also have crafted forms of code (such as vector analysis of gender-index words) that enable scholars to lift nuanced semantic differences from very diverse bodies of text. In other words, the skilled reader, addressing this material, should be able, with the help of critical infrastructure, to preserve and describe cultural nuances that differentiate one moment from another, peering with ever finer attention into the characteristics and tensions within a corpus.

I am not saying, of course, that different cultures will not require different infrastructures or different kinds of code. Scholars of Chinese history, for example, reaffirm that many forms of digital humanities pipelines developed for western texts – which, for instance, depend upon word order for grammatical understanding – would be misapplied to their field, even while forms of analysis that jettison word order – including semantic network analysis and topic modeling – appear to work fine. The promise of infrastructure to extend and replicate scholarly inquiry is clear; the limits of that promise need still to be worked out through many case studies and much experimentation.

The Principle of Creating Knowledge About the Social World.

An implied extension of a transparent document base is that documents are a reflection of a shared historical context that has a reality and permanence, even within opposing interpretations: Jefferson's contemporaries might have argued about the tolerability of human slavery, but none would have contested that slavery was a contemporary reality of deep and persistent concern, structuring their economic and social world to a deep degree. A piece of infrastructure can be – in certain circumstances – a tool for generating assent. Through encouraging consensus around some insight into social or economic reality, pattern recognition can offer the basis for collective action. More than simply advising a critical point of view, infrastructure can provision citizens and movements with data that actually then becomes part of their work.

It is clear that the humanities and social sciences have a great deal to offer, on an interpretive level, to the public, beyond merely being purveyors of data. Philosophers and feminists from Hannah Arendt to Seyla Benhabib have long valorized the role of the humanities and social sciences in the public sphere, in particular citing its capacity to enliven critical thinking about contemporary discourses of politics and identity. Working in this tradition, many scholars are used to understanding "critique" and "complication" as aligned with a political critique of empire, politics, and exclusion. More recently, Rita Felski has argued that critique reaches its limit when the hermeneutics of suspicion is turned everywhere

[Felski 2015], and has urged scholars to remember their value as curators of memory, critics of politics, and composers of stories.

Inviting the public to inspect data for itself is an action that implies political consequences. Consider the political context that motivated historians David Eltis and Catherine Hall – both historians of slavery — have turned into builders of infrastructure, in the midst of British debates about remunerations for slavery. Such projects as these represent broadly the fact that scholars are using infrastructures for serving cultural and political artifacts to the public as a way of engaging, instructing, and facilitating further cycles of engagement by the public.

Their projects have granted the public access not only to a curated debates about the history of slavery, but also directly to massive troves of quantitative, spatial and economic data about the British slave trade (Slave Voyages, The Legacies of British Slavery). Of the two projects, it is Hall's – the "Legacies of British Slavery" – that has perhaps most directly sought to create an epistemic bridge to contemporary discourse, insofar as they profile the continuing profits that Europeans have reaped as a result of legacies of the past.

Both portals did three things that critical essays cannot do: (1) They reduced the arguments over Britain's involvement in the slave trade down to a single visualization. (2) Indirectly, they invited the public to view the data upon which the visualization was based for themselves, offering it in the agnostic guise of "raw data" rather than as a series of footnotes compiled by an expert in the service of a project. We might, of course, take issue with "raw data" in principle, agreeing with Johanna Drucker that all data is already cooked: but the public fascination with raw data remains. And both scholars implicitly took advantage of the public fascination with raw data to demonstrate, in their portals, the extent of data upon which their visualizations rest. (3) They made the data interoperable, allowing the public to test conclusions for themselves.

At the same time, other scholars reviewing those projects might raise questions about whether all examples are equally "critical" in the sense of how profound is the new perspective it offers on the past. A scholarly edition represents a critical intervention in Literary Studies and is celebrated when evidence is mustered to persuade readers of a dramatically new interpretation of a canonical text. Digital editions, like *Frankenbook*, the most recent digital edition of *Frankenstein* – an exemplars of the genre, built on several generations of digital edition work — would have to pass this test of "critical" interventions in order to attract the notice of some departments. Clearer examples of critical interventions where humanities and social sciences knowledge is used to create a radical break with present-day narratives of politics and identity include *Torn Apart/Separados*, which documents child separations at detainment centers along the national border, providing users with data, visualizations, and a critical narrative. At the same time, *Torn Apart* ejects a critical humanistic perspective into materials in the public sphere.

In sociologist Matthew Desmond's ethnography, *Evicted* , Desmond provides accounts of families in Milwaukee who faced eviction, and the resulting stress to savings, education, children, health, and jobs that resulted from their experience. In Desmond's companion website, *Eviction Lab*, he provides users across the United States with information about the rate of homelessness and eviction in their neighborhoods. The infrastructure offers insight into the local realities of neighborhoods outside the Rust Belt. Another companion website, *JustShelter*, offers corridors to action by connecting users with organizations that advocate for housing rights.

Like Desmond's *Eviction Lab*, *Torn Apart* offers the public an open-source databank that renders visible previously invisible patterns of expenditure and violence. *Torn Apart* thus creates a scholarly intervention in a realm that was previously little addressed by scholars, or if addressed, in such a partial way as to make less of an intervention than *Torn Apart's* open-source databank. Rendering visible the invisible, and making accessible the previously inaccessible data of knowledge, such infrastructures as these attempt interventions that differ in content, scale, and power from other scholarly productions, and they are therefore deserving of special attention as models of criticality and engagement with the public sphere.

The social power of documents to create an informed consensus becomes evident in digital humanities portals that illuminate politically contested episodes of national history. In portals such as *Slave Voyages* and the *Legacies of British Slavery* site, historians of Britain have published, for the benefit of a wide public, the primary records and quantitative

distillation of the history of persons enslaved by the modern west. The portals, by their design, offer members of the public, as well as other scholars, access to primary-source artifacts, original numbers, visualizations and maps, as well as critical interpretations that, for instance in the case of the Legacies project, link the historical record to present-day companies and families whose wealth was bound up with the trade.

In such a circumstance, where the material governed by a site is potentially political, the designers of infrastructure have an opportunity to thwart the logic of naysayerdom in the era of fake news by firmly circulating primary-source evidence to users. By providing direct links to the primary source material on which the analysis is based, the *Legacies* site all the more firmly asserts the possibility of a new consensus around the history of British slavery and its repercussions.

72

Perhaps a more important aspect of scholarly infrastructure as argument functions in the realm of a critical understanding of the potential audiences and points of access to scholarly data. In the age of peer-to-peer networking with humanities and social science data, new collaborations are possible that never could have been accomplished in the age of individual research in the archive, and many journalistic theories have sought to characterize the generative potential of "hive mind" or many-to-many learning (Shirky). More recent research, however, has emphasized that many-to-many discursive spaces online create echo chambers that can be deleterious to learning, feedback, or critical thinking (Nguyen).

73

Information flows of any kind may generate echo chambers. Social theory foregrounds the potential of any flow of information to either intensify power or undercut it. It suggests that the choices made in the construction of institutions and infrastructure determine whether the experience of information will create a cult-like experience of information – where outside information is met with suspicion – or whether new information, data, and expertise will be experienced as the essence of liberation.

74

It might appear, of course, that a principle that demands consensus about reality is one of top-down power exerted over the user base by the designers. Indeed, the principle of creating knowledge about the social world operates in tension with another principle: that of a plurality of humanistic desires.

75

The Principle of a Plurality of Humanistic Desires.

Yet another orientation was to imagine our potential users as a plural community with a diversity of desires and perspectives with regard to electronic archives. In designing *Paper Machines* for an imagined diverse public of my fellow researchers, I had no ambition of creating a specific tool that would "understand" documents; rather I would leave it to users to organize their documents into relevant categories – for example, comparing decades, authors, subject-matter, or genres of writing. Instead, the tool would only implement algorithms that would prosthetically "extend" the ordinary capabilities of scholars: generalizing about large bodies of text, enabling the researcher to develop new insights into the material. By designing an infrastructure with fairly humble goals, we would offer the services of these algorithms to researchers who had little time away from the archive to code.

76

This principle holds more instruction for institutions such as libraries, museums, and national archives that have a particular responsibility to a non-scholarly user base. If the infrastructure is supposed to serve the general public, it must be accessible and transparent to a general public, whose desires from the archive may be genealogical or political in a narrow sense, rather than a reflection of current scholarship. The tools at hand should enable the finding of more documents according to a plurality of desires.

77

Indeed, Joris van Zundert has lately underscored the limits of humanities infrastructure projects tailored to a single archive – and the need of tools to help scholars who deal with heterogenous data [Van Zundert 2012]. The desire principle, in theory, would facilitate not only collective research programs but also individual trajectories of research within those archives – and across multiple archives. Zotero allows a user to collect documents from multiple databases, for instance gathering trials from the Old Bailey and books from Hathi Trust alongside pamphlets help on JSTOR. By working within Zotero, *Paper Machines* made room for such erratic document-gathering habits, and attempted to offer tools that would allow the researcher to generalize about the collection as a whole.

78

Once a scholar who used Zotero had collected her own archive, *Paper Machines* was designed to facilitate individual exploration and experimentation with as broad an array of tools as possible. By trying on different categories, researchers could summon algorithms to quantitatively interpret texts, in much the same way as researchers' own imagination typically sorts texts into different categories by decade, genre, and theme. The researcher would thus be able to *test* her subjective assessment of different categories using the quantitative power of a tool. Quantitative testing, we assumed, might reveal the user's subjective categories to be grounded in fact. Alternatively, it might destabilize the user's categories and suggest other patterns, established by regular word order, that might complement or challenge the user's assessments.

Through these operations, researchers could iteratively model sub-portions of a corpus. They could view a body of texts as if through different lenses, reorganizing the corpus into different subcategories, and comparing the perspective raised by these subcategories, for instance rapidly glancing at the perspective of the corpus organized by decades and that of the corpus organized by genres or themes. Similar design principles governed the University of Virginia's *Collex* , designed to facilitate the collection and exhibition of a suite of documents gathered by a researcher according to his or her own research program; they continue to inform the design principles behind *Omeka*, the go-to resource for narrating a public story grounded in well-annotated archival artifacts. Both tools allowed scholars to extrapolate their own language for analyzing cultural documents, providing basic annotations like maps or timelines to enhance the user's capabilities of storytelling.

Paper Machines and other analogues like *Voyant* and *Cortext* enhanced individual story-telling with the possibility of data analysis. They provide researchers with a to test different quantitative perspectives on the corpus, rapidly switching between the perspective enabled by two different mathematical models, for example word count and topic model.

Yet *Paper Machines* added one facet that the others lacked: because it was built in Zotero – a platform that allowed scholars to share collections of documents with each other – *Paper Machines* facilitated research on shared archives, where a group of scholars might agree about a subset of documents of interest, and then embark on different algorithmic approaches to understanding those documents. This zone of inquiry would take place between individual research on mass shared collections and individual research on individual collections, occupying what Paul Edwards might dub the "intermediary scale" of research infrastructure. It would make room for shared, sociable research desires to emerge – not merely on an individual or collective scale – but also within the space of micro-communities.

The Critical Principles of Scholarly Infrastructure

Thus far, the principles have focused on the relationship between a non-technical user base and a technical process of transformation, although embracing the principles has, as has been shown, implications for the power of the digital humanities to redound onto critical theory and seed interventions across the disciplines. Another set of principles govern the relationship between users and the erudite nature of archives themselves; these principles are best understood within a review of writings about power dynamics and infrastructure.

The implicit power dynamics involved with supporting digital infrastructure within the university itself have been subject to inspection and critique for at least two decades. After all, as critics of modern technologies of communication have long understood, power is at work at each scale in an infrastructure system. As many scholars in what is now called "critical infrastructure studies" have shown, at each scale and within each part of the pipeline, choices in values can change what is implemented [CI Collective]). On a national scale, highway systems, water systems, and other infrastructure are designed with the desire with potential interest groups in mind: for example, water policy in early twentieth-century California shored up the advantages of large landholders by giving them privileged access to irrigation (Worster), while the design of eighteenth-century road networks in Britain and nineteenth-century land redistribution programs in America was designed to give economic access to small farmers and workers (Worster, Guldi, Foner). A choice of values – which either reinforce existing power dynamics or counteract them — is thus at work in the design of infrastructure at every level of scale.

Digital infrastructures play into the same dynamics: they redound power dynamics of rich and poor nations and rich and poor communities, experts and non-experts, by reinforcing existing differentials of access. On a global scale, internet

infrastructure is effected by choices about internet providers, such as networks of physical cables that connect rich countries to the exclusion of the poor translate realpolitik into technology [Starosielski 2015]. On the geographical scale of each community, smart phones and GPS usage require some levels of literacy and other forms of expertise in order to be useful [Sawyer et al. 2019]. What's more, the choice to maintain or not to maintain structures the accessibility of infrastructure and its benefits [Mattern 2018] [Graham and Thrift 2007] [Henke 1999]: as when local maintenance of water facilities, delayed in Michigan, results in working-class populations being poisoned, or when differential maintenance across Parisian neighborhoods results in a seamless experience for an elite minority and a degraded experience for the rest [Denis and Pontille 2014].

In creating digital infrastructures, the builders of web portals have opportunities to counteract these power differentials. Builders who favor the strategies of "minimal computing" put their websites within reach of citizens in the global south or poor neighborhoods who frequently must access the internet through second-rate connections [GO::DH 2018].

Other builders of infrastructure have primarily conceived of themselves as designing the tools for the coming of another community beyond the imagination of the designers. For example, the activists who designed toolkits for participatory mapping in the 1970s and 1980s imagined that the toolsets they wove together would be used by communities for purposes beyond what the original activists could imagine: and indeed, put into action, the participatory maps came to be used for the purposes of monitoring pollution, redistributing land, and protecting indigenous property rights (Guldi). Just so, the historical function of Britain's nineteenth-century road network cannot be entirely understood through the work of engineers, but only through the use of the network made by the road's users – including Methodists, journeymen, and radicals – who used the roads to create communities sometimes antithetical to the capitalist and nation-building purposes imagined by the network's designers [Guldi 2012]. The power of relinquishing control over infrastructure often makes the infrastructure of comparatively long use and greater effectiveness overall.

Some scholars will resist a form of scholarly practice marked by the relinquishing of control. They may argue that the value of traditional scholarship reflects the scholar's command of language, citation, and transparency of sources to elucidate what facts have been used and what argument is being made. Indeed, many scholarly infrastructure project support direct control over the transmission of bibliographic information. While a web-born scholarly edition of Mary Shelley's *Frankenstein*, for example, may offer a useful resource to the teachers of Victorian literature grounded in printed and manuscript archives, the web-born format almost entirely replicates the form of traditional scholarly editions – something entirely useful for the purposes of tenure and promotion. Similarly, carefully-annotated digital collections of historical documents such as *Vistas: Visual Culture in Spanish America, 1520–1820* (https://vistas.ace.fordham.edu/) represent the equivalent labor of an art history exhibition or heavily-researched book. In format, these tools differ very little from traditional scholarly publications in books, excepting the use of hyperlinks and the scale of their accessibility. Scholarly infrastructure with a traditional form may nevertheless be radical in terms of its content; both the edition of *Frankenstein* and *Vistas* open up important questions about gender and race. Not all authors of scholarly infrastructure make the relinquishing of control over their sources or the sources' interpretation a feature of their design.

Projects that emphasize the "critical" use of infrastructure make at least some gesture towards the relinquishing of control, pointing towards the wider uses of the community. As such, they overlap with the final instance of the humanistic principles, the "principle of a multiplicity of humanistic desires." Infrastructure of this kind resonates with participatory traditions in urban planning, where planners consult the local community about the best course of development in a particular neighborhood [Friedmann 1973] [Guldi 2017]. There are, moreover, traditions of literature and philosophy where the relinquishing of control is the hallmark of utility. For Roland Barthes, one of the hallmarks "writerly literature" is that it can inspire multiple readings and uses in cultures far from its point of origin, and so "make the reader no longer a consumer but a producer of the text" [Barthes 1974, 4]. The creation of a meaningful infrastructure project is marked by the project's ability to support the user's active production of knowledge.

The principles outlined below investigate how scale of exposure is deliberately used by certain websites to place texts of data in the public realm, often marked by the relinquishing of control over the ends to which the later products are put. Frequently, infrastructure projects of this sort work towards the multiplication of projects well beyond those imagined or intended by the designers. This article's conclusion will strive to understand what this multiplication implies for the work

of scholarly infrastructure.

The Principle that Democratization of Access to Information Breeds Better Democracy.

The experiments with infrastructure that press the principle of neutrality with respect to technical literacy to its furthest extent are those that underscore mass participation, and they come to us typically from science and technology studies – a discipline familiar with theories of how technologies define the limits of their own participation or exclusion. In Science and Technology Studies, researchers have experimented with radical infrastructures that extend the principle of technical accessibility past the archive and web application, into a physical infrastructure accessible even to individuals whose computer and internet access is limited. In his dissertation on formaldehyde in FEMA trailers, anthropologist Nicholas Shapiro worked with chemists to commission post cards that would test the formaldehyde content in the trailers' walls, if occupants mailed the postcard to a testing facility [Zhang et al. 2019]. The paper infrastructure of chemical-sensitive postcards thus made chemical testing of pollution in the environment radically available to ordinary citizens, allowing occupants of the trailers to test their environment for noxious chemicals by themselves. Infrastructure projects such as these have pressed the bar of what it means to offer information directly to the public.

Early in the scholarly infrastructure turn, many designers of infrastructure believed that delivering data and stories via the web offered an opportunity to democratize access to both scholarship and primary sources. In 2008, Harvard historian Robert Darnton was already making the case for open access to the digitized books scanned by Google, Howard Rheingold's article, "Participative Pedagogy: For a Literacy of Literacies", for example, argued for digital practices that broadened access to "literacy", broadly defined, with the end of teaching "the ways people use knowledge and technology to create wealth, secure freedom, resist tyranny". For Rheingold, this "literacy of literacy" was more important than teaching the how-to of particular kinds of code or technology; in essence, Rheingold was arguing that the age of the internet made more necessary the proliferation of engaged and critical readings of culture [Rheingold 2008].

It is hard to imagine a humanities infrastructure that mirrors' Shapiro's postcards in bringing the knowledge of the university radically to bear for communities in need, but some archivists have begun to theorize the shape of such a project. At the Columbia University Library, Alex Gil's *Nimble Tents* creates digitization kits to help communities rebuilding after a natural disaster to map their resources. Gil has also created "The Translation Toolkit", another kit with which communities in impoverished neighborhoods and the developing world can record their own stories and send a copy to Columbia for preservation and translation – with the idea that recordings belong ultimately to the community, and that Columbia can aid communities in their analysis and documentation [Gil 2018].

The elaborate series of scholarly articles introducing students to the Old Bailey archive (http://www.oldbailey.org) expose students to themes of poverty, violence, gender, and race in Britain's past. The website obviously holds the same merit as (at least one) traditional textbook, if not many more, by virtue of its carefully prepared archive and the many visualizations thereof: the main use of this part of the website is to expand the scale of potential readers who can potentially access these essays.

Consider *EvictionLab*, which shares social science data about the demographic effects of eviction policies around the nation. *EvictionaLab* can be seen as complementary response to the social and economic critique of the harm served by eviction against families of color offered by Matt Desmond's traditional scholarship, ethnographies published in the form of scholarly articles. The scale of harm that Desmond uncovered in eviction was a form of harm *at scale*. A critique – even a powerful one – barely approximates the scale of the harm that he discovered.

Becoming a builder of infrastructure allowed Desmond to deploy his research *at scale*. By creating a public portal that could answer to a multiplicity of local demands for information about the local face of eviction, Desmond was able to offer a level of detail that would be hard to rival in a traditional book or article. In a sense, the act of building allowed Desmond to enact a remedy of an appropriate scale for facing the trouble that he had elsewhere critiqued. Thus while critique illuminates our knowledge of harm, the limited effectiveness of critique hampers it: infrastructure-building, by contrast, in the form of data portal's like Desmond's, is a form of scholarly activity that activates a critique at scale, demonstrating the relevance of an economic critique to geographical locales.

The era of scholarly infrastructure has thus begun to forge a fundamentally new idiom of scholarly communication, one where the intellectual labor of critique can be married to the possibility of broad application at scale. The proposed equivalence is ethical in nature: a deep harm, which scales to the level of billions of families, requires a remedy that can travel further than a single well-cited journal article. Deep harm requires a remedy at scale: potentially, a remedy at scale also therefore requires not critique alone but also data infrastructure.

In many cases, by using scale and relinquishing control over the uses of data, infrastructure invites broadcast participation in arguments over social facts and public documents. In so doing, infrastructure does the work that many scholarly essays, for reasons of their small audience, often simply cannot do: elevating a political critique into the public sphere.

The Principle of Community Ownership.

In an article elsewhere, I have argued for the relevance to social science and the humanities of citizen science and other movements that collect and analyze data. For scholars pursuing the history of rent and eviction, for instance, the labor of an individual scholar is dwarfed by the scale of research imaginable if members of the public took up a scholarly charge to begin collecting data about the history of rent prices, rent laws, and eviction rates in every city around the world. To be sure, scholarly acceptance of such a crowdsourced database would depend upon adhering to the foundational principle of transparency, as well as existing standards for the documentation of the *provenance* of a document [Cheney et al. 2012]. The payoff of building avenues towards crowdsourced knowledge, however, would be an undertaking of research that would dwarf contemporary undertakings in its scale and potential for insight. A truly global program of insight would result.

Critical theory has long recognized the existence of principles of democratic participation against which initiatives and organizations can be measured. In 1968, Sheri Arnstein's "ladder of citizen participation" arranged different forms of information sharing in the public sphere on a hierarchy of democracy, ranging from "manipulation" and "therapy" on one extreme to the devolution of power at the most democratic. Arstein's theory underscored the power dynamics in information flows from government experts to the people, and raised the possibility of new kinds of citizen-governed institutions, a move that reflected current concerns of African-American populations in American cities, but mirrored radical ideas about governance going back at least to the era of the Paris Commune [Guldi 2017]. This early theory of power and democracy in political information offers a pattern for understanding the radical potential of scholarly information on the internet.

Today, examples of citizen science abound, where users contribute observations, insights or desires, which are exchanged over a community-run infrastructure. Science and technology theorists have often critiqued the first generation of citizen science for the lack of discursive engagement with a program of observation laid out by experts: citizens were mobilized to count birds, to cite one famous example, but not to raise critical questions about how or why existing climate science had failed to be taken up in American public discourse.

There is enormous variation in whom different portals allow to play and how. The *Legacies of British Slavery* project makes data available to users, but stops short of allowing users to shape their own inquiries, holding tight to the overall interpretation of the moment in what may be informed by a concern for preserving the sacrosanct authority of interpretation at a time of rising ethnic nationalism. *Eviction Lab* gives away the data, allowing users to make what they will of demographics and instances of families ejected from their homes. The *BookNLP* software package allows users to isolate variations in the handling of gender, but lacks a public portal (as of now) that would allow a tech-illiterate user to engage such an analysis. Each of these decisions results from a complex of ideas about the scholar's role, the expertise of their intended audience, whether scholarly production serves purposes outside the academy, and what the scholar's relationship to that public is or ought to be. Those decisions are, by nature, philosophical, political, and professional; they are also subject to critical debate.

Outside the academy, communities are transforming science into the practice of citizen science by realizing Arnstein's principle of participation and community ownership. For example, the organization Public Lab hosts "Research Notes" where community members like social scientist Nicholas Shapiro post requests for technical expertise – in Shapiro's

case, requesting help with the technical end of formaldehyde testing. In another online community, a community of small farmers called *FarmHack* shares blueprints for appropriate technology to make small-scale farming more economically viable. But experiments of this kind are less common in the digital humanities.

Humanists who share the values of democracy would be well-advised to consider whether the practice of citizen science has its place in the interpretation of text: for example, whether ordinary citizens require the tools of inspecting, via distant reading, the congressional or parliamentary debates of their own government.

104

The Labor of Infrastructure Building

In the foregoing sections, I have surveyed a range of possible values expressed by the builders of scholarly infrastructure. A proposition of this kind is important because the intellectual engagement of infrastructure-builders with critical thought has not always been self-evident to outsiders.

105

Understanding value as a factor in the kind of work embodied by scholarly infrastructure projects – including the bibliographic, critical, political values represented by engaging in project-building as opposed to traditional scholarly documentation in the form of books and articles – is crucial to having those projects accepted and interpreted as interventions in the realm of scholarship, where critical commentary routinely engages the kind of intellectual values represented by a project and the way that data, visualizations, and other forms of labor support those values. Understanding the range of possible values underlying those commitments is therefore crucial, as communities and scholars come to terms with the stake of committing energy to infrastructure-building rather than to traditional scholarship products such as grants and articles.

106

The metaphors that I have offered above – comparing scholarly infrastructure to an opera house – are also productive in a number of directions that cannot be exhausted in the course of this article, especially involving the question of where labor or ideological contributions occur. Panning out from the opera and opera-house itself, one can see any given production of *Aida* is itself the product of architect, composer and librettist, conductor, and cast. A reader-centric interpretation even pushes this list to include the audience in the theater on any given night, and their manifold interpretations of the opera [Locke 2006].

107

Just so, the work of digital production and interpretation includes all of those forms of invisible infrastructure that include scanning and cleaning texts, storing and serving them, before the design of the infrastructure; the infrastructure itself, any scholarly articles whose interpretive approach depends on the infrastructure in question, and any individual users of the infrastructure, whether or not they later become authors of analyses published in traditional academic venues.

108

The multiple locations of labor and interpretation in such a process constellate a broader transition to a hermeneutic regime of information at scale. Elsewhere, I have argued that shifts of this kind are marked by a deferment of "critical thinking", where insight is produced at a single point of interpretation and argumentation with the text, into a processes of "critical search", where critical engagement is required over the entire length of an information pipeline [Guldi 2018a].

109

Appreciation for the values implicit in scholarly infrastructure has not always been the tone by which digital scholarship is received by larger communities in the university. In one inflammatory editorial in the *Chronicle of Higher Education*, a scholar characterized the digital humanities as being the tool of the neoliberal takeover of the university and an ally in the adjunctification of the professoriate [Brennan 2017]. In at least one tenure process with which the author is familiar, a department threatened to stage a kind of inquisitorial panel to monitor a digital humanist already published in traditional books and articles, because engagement with digital humanities and social sciences question – including building infrastructure – was understood as somehow sullying the life of the mind with the dirty matter of service.[1]

110

The remedy for this failure is debate and conversation that theorizes what it is that people do when they build infrastructure, and at what stages scholarly discourse, humanistic discourse, technological acumen, dissent, and political thought enter the process. Scholars who think critically about infrastructure must advertise their critical engagement in the form of a scholarly essay that reflects on the format of the infrastructure adopted and its implications.

111

In teasing apart the range of values represented by modern infrastructure, I have argued that infrastructure building implies, at minimum, a commitment on the part of the scholar to *scale* and a relinquishing of *control*, which imply a choice of values, including a set of "traditional" values, such as transparency with regards to document base and the knowledge of society, which are aligned with long-standing commitments in the humanities and social sciences, and "critical" values, such as participation and the critique of power, which are aligned with radical traditions in the academy and beyond.

Recognizing the choice values at stake in building infrastructure is essential, I believe, to understanding much of the work in the digital humanities and social sciences as well as other parts of the university today: there is something on offer that has not been illuminated in the voluminous literature on infrastructure, digital humanities, or the crisis of the humanities, which deserves to be recognized, praised, critiqued, and pressed, because it presents a challenge to contemporary practices of scholarship, and scholarly communities must decide whether that challenge is worth embracing.

What is this offering, and why has it been so submerged?

On the surface, infrastructure represents a different species of animal from the scholarly article, whether single-authored or authored by many individuals. In some cases, like Andrew Goldstone's *Topic Viewer*, they are the creation of a single individual, but the individual's responsibilities include using code to who cobble together existing packages for data cleaning, analysis, and visualization, as well as scholarly insight about the kinds of cleaning, analysis, and visualization likely to produce insight from texts of a particular type. These designs thus resemble scholarly articles or narratives in that they provide users with an orientation to the data, based on wide experience; but they also resemble cartography or traditional kinds of data analysis in that they entail working with vast amounts of evidence to produce visualizations useful to different parties for different reasons. As a monument of scholarly labor and insight, the building of infrastructure typically represents a major scholarly intervention – entailing more labor than that included in the typical journal article in the humanities or social sciences. It bears the weight of a book.

In many cases, the committees that assign credit to academics for publishing books and articles resist crediting the development of software with the same power on the basis of the *plurality* of labor represented by the creation of an infrastructural project. Infrastructure often represents the product of collaborations, where "labs" of scholars and students, or hybrids of scholars working with university staff and/or private contractors, produce the design for the project and its execution. It can, after all, be difficult to isolate the contribution of a single party working on a project [Edmond 2015]. Typically, in these collaborative labs, coders and scholars collaborate to design a visual and interpretive strategy that governs how users interact with the data. The end product reflects ongoing collaborations, much as an edited book may be conceived of and rallied by a single scholar but may represent the work of many hands.

In some of the most highly-developed web portals that serve humanists, the infrastructure requires ongoing updates managed by graduate students, salaried coders, full-time staff, and administrators, whose tasks may range from insuring interoperability with users' machines to managing lawsuits (for instance in the famous case in which Thomson-Reuters sued Zotero. In any case, the range of labor provided – from coding and design to data collection, cleaning, analysis, and more – is more diverse than the traditional scholarly processes of archival research, analysis, and writing, even where facaulty act mainly in an administrative or visionary role rather than producing code themselves. Overseeing the development of infrastructurerequires the scholars involved to conceive of their arguments in visual terms, and to produce multiple visualizations that align with different perspectives on the data in question.

Yet the plurality of labor necessary to realize infrastructure is only part of the story of why the university community is sometimes hostile towards infrastructure projects. A turn towards personal autobiography will be useful here. As a scholar of the history of infrastructure who researched eighteenth-century road building and wayfinding through traditional archival means, I came to the digital humanities primarily as a researcher interested in how keyword search was changing my own techniques of research. Initially, I resisted the idea of building infrastructure, preferring to think of myself exclusively as an author of books and articles and a consumer of tools developed by others.

The flowering of new tools in the digital humanities – the "infrastructure turn" identified by Geoffrey Rockwell (2010) —

persuaded me that scholars with active the gap between existing tools for annotation, curation, and story-telling – such as *Zotero* and *Omeka* – and the array of text-mining and visualization practices coming into being from computer science. Some of those projects were created by affiliates of research universities such as George Mason, which had a firm commitment to building infrastructure, and whose graduates were distinguished by their knowledge of servers and code. Others were the work of senior faculty, who had already built a career composed of multiple books, who turned to building infrastructure at an advanced stage in their career when they could command a multi-year research project to build an infrastructure around an infrastructure of particular interest to themselves and their students. Admiration for those accomplishments made me question archival research as the sole expression of scholarly research and argumentation, and opened the door for my curiosity about infrastructure-building.

In 2010, I began writing grants and talking to colleagues about the possibility of building an infrastructure for the digital humanities and social sciences. I imagined that building a tool that would distill scholarly knowledge into frequently-repeated patterns, extending the scholar's power of recognizing patterns over tens of thousands or millions of documents, and thus acting as a kind of scholarly prosthesis. Working in research-focused departments nevertheless forced me to reconcile my admiration for scholarly infrastructure with my own needs, as well as the needs of my students and colleagues – all of us, of course, being researchers at first-tier research programs, who were incentivized above all to discover new outlooks on the past, whether from archivally-new materials or from a new perspective on events already known but insufficiently understood. It was hard to imagine a single tool that would serve this purpose. The graduate students who I worked with at Chicago had extremely diverse research backgrounds and interests – ranging from early-modern Germany, to modern Latin America and nineteenth-century Chicago. The tool that would serve my research-focused students would necessarily need to work with documents of concern to them all, using existing language-processing algorithms and visualizations to offer insight into documents of different kinds. Thus it was important to build a tool that would allow researchers invested in text from a range of backgrounds to profit from the tools of word-count analysis and topic modeling then being used by a handful of scholars in the digital humanities.

Many examples mark how senior scholars skeptical of digital methods have held up the career of junior scholars — even those who have otherwise fulfilled traditional requirements for tenure — and this certainly seems to be the case for builders of infrastructure. Such affairs undermine the progress of the field, as one anecdote will illustrate. Consider the case of a scholar, known to me, whose interest in digital methods had led her — after finishing a traditional monograph in her field — to build a piece of digital infrastructure also publish a treatise investigating the future of her field with respect to the digital humanities When she did so, she was punished by her department.

The grounds for tenure laid out at her hiring were reversed; she was ordered to write a third book on a traditional, archival topic that was ordained by her committee, and it was conveyed to her that the department was outright hostile to her continued use of digital methods. Far from recognizing the value of a piece of infrastructure developed by a junior scholar, her department counted infrastructure as an albatross for her career.

In the vita that she had submitted to her committee for tenure, this scholar could count several accomplishments unusual for historians, including the aforementioned infrastructure, as well as a significant grant. Following the advice of colleagues in other departments, this scholar circulated a draft in which the infrastructure and grant-related projects were classified under a heading of "book weight projects". Her committee recommended that she remove the infrastructure project to a section on "miscellaneous projects". The message to her was clear: infrastructure would not be valued alongside books and articles as a marker of scholarly labor, intellectual value, or theory-driven critique. The double-binds that this individual faced – resistance to the digital humanities, resistance to infrastructure-building, and few incentives to document the critical thinking behind existing infrastructure experimentation — amounted to an incentive for delay. She delayed writing further about infrastructure, in part, because the culture of resistance she faced meant that even this kind of scholarly commentary on experimentation would itself be rendered suspect by tenure review committees by the same hostility that the article needed to challenge.

Things might have gone otherwise for the individual and infrastructure in question, of course, had there been intellectual apparatus that considered the development of infrastructure as a form of scholarship, infused with the values of critique: her investments of time might have been viewed as a viable form of scholarship and a valuable investment in future

forms of scholarly community, rather than a dangerous action inappropriate for junior faculty. The problem was essentially that investments in scholarly infrastructure were obscure from the point-of-view of existing scholarly institutions; they could be seen neither as a form of scholarly argumentation, nor as a critical reflection on contemporary scholarly practice. In order for that resistance to give way, scholars would need to document the history, debate, and values out of which new experiments with infrastructure were emerging.

Institutional resistance to a culture of experimentation, however, is a wicked problem, in that no one solution is sufficient to overcome it: a single major publication is rarely sufficient to shift a culture. In my own career, I began to develop such a justification for scholarly infrastructure projects in my second book, *The History Manifesto*, which set forward the promise of tools like topic modeling for opening up longue-durée analysis of problems to historical analysis. `124`

Failure to recognize or support infrastructure as a scholarly achievement has consequences. Today, the infrastructure in the story is no longer maintained — a fact that highlights the differentials of commitment between senior faculty and junior faculty. As with the train system in the Parisian suburbs, junior faculty projects unprotected by senior administrators are liable to be refused maintenance and so degrade. I delayed writing explicitly about infrastructure as scholarly production and intellectual critique. Very few articles reflect on the process of building or state a case for the role of argumentation and thought in infrastructure design; there was very little, in 2013 and 2014, to point to by way of a firmly-established "infrastructure turn" that would be acceptable to senior historians. Indeed, at 2019, this essay offers a much-delayed assessment of the role of critical thinking in the design of *Paper Machines* in 2009-13. `125`

The double-binds that I faced in 2014 — resistance to the digital humanities, resistance to infrastructure-building, and few incentives to document the critical thinking behind existing infrastructure experimentation — amounted to an incentive for delay. After initial hostility to the broad-brushstrokes with which we praised the potential of the digital humanities in the *Manifesto*, I delayed writing further about infrastructure, in part, because the culture of resistance I faced meant that even this kind of scholarly commentary on experimentation would itself be rendered suspect in tenure review committees by the same hostility that the article needed to challenge. `126`

We might refer to a culture that resists innovation, because the terms that inspire innovation are obscure and hostility is rife, as a "culture of delay". `127`

My own partial and much-delayed engagement with these problems may be symptomatic of larger failures of acknowledgement for the builders of scholarly infrastructure, as well as our own hesitance to explain our values. Scholars who have come to make major investments of time and energy in building infrastructure rarely justify those projects in scholarly or public venues, perhaps because the intellectual labor behind the work or the time dedicated to infrastructure projects is likely to be dismissed, and the labor judged with hostility. The failure of institutions to recognize or reward scholarly infrastructure compounds almost certainly compounds scholars' slowness to highlight the intellectual values or merit expressed by their infrastructure building projects. `128`

I therefore propose that the remedy for culture of delay is, ironically, more theory: theory in the service of illuminating the alignments that compel some scholars to experiment with building and sharing infrastructure at scale. The builders of infrastructure need to theorize openly about the choices they have made and the values that those choices represent. The foregoing principles have been offered for the sake of waymarking out some broad and familiar territory, but future theorists of scholarly infrastructure will necessarily have to take up the challenge of setting up their own waymarkers, or arguing with and thus refining the extremely preliminary terms in which I have framed values such as "community ownership" or "transparency". In that process of scholarly debate, a trail of references will be created that will familiarize critical thinking about infrastructure for mainstream communities of scholars. `129`

Conclusion: The Work of Scholarly Infrastructure Projects

What sort of work, then, is the labor that goes into a humanistic web site? This article has already suggested that the principles governing that academic labor reflect not only bibliographic and scholarly traditions, but also discourse from critical theory about the liberatory potential of knowledge and participatory research. A strong argument has thus just been made for infrastructure-building, in the humanities, as a site of intellectual argumentation, so long as the process `130`

of designing and building the infrastructure entails some moment of reflection upon the principles that governed its design and their inspiration (which need not, of course, replicate the principles outlined here).

Clear though the argument may be for critical reflections on infrastructure as a form of scholarly argument, some account must be given of why so little critical discourse exists on the building of digital humanities projects of this kind.

<div style="text-align:right">131</div>

This article has proposed nine criteria of critical thinking about infrastructure demonstrated in mainstream projects which demonstrate an engagement by the designers with ideas about the humanities. Two of those criteria – those I have labeled "critical principles" – demonstrate that scholarly infrastructure embodies political and argumentative work prized by some scholars, for example, those concerned with the defense of democracy, or the radicalization of community access to knowledge. Projects that demonstrate these criteria show that the designers have actively prepared the infrastructure so as to engage with the concerns of their discipline, and building, under these concerns, represents an act of engaged criticism and argument. Many of what I have called the "bibliographic principles" also demonstrate an engagement with critical thinking, and also with the political insights of social theory, in the form of an engagement with infrastructure as a tool of power. Infrastructure projects that demonstrate these both criteria suggest that practitioners have not only engaged with the concerns of their discipline, but with the critical theory of the last half-century as well, particularly as translated through the social/historical infrastructure turn and its critique of the infrastructure of knowledge-making.

<div style="text-align:right">132</div>

In other words, the realization of the critical principles in acts of scholarly development demonstrate that a wide variety of infrastructure-builders are actively engaging a wide-ranging critique of the institutions of knowledge and their political power. Here, scholarly engagement and argumentation are taken to an extreme largely unknown in the traditional disciplines. A double infrastructure turn – combining social insight and technological building – is at work, which merits the inspection of scholars across the academy. Infrastructure building is the present-day site of engagement with the insights about power and governmentality that defined the scholarly infrastructure turn.

<div style="text-align:right">133</div>

Acknowledgements

<div style="text-align:right">134</div>

Notes

[1] The scholar in question left the university in protest, prior to the activation of the inquisitor panel or the tenure review board. The review panel was proposed alongside a series of requirements that would have allowed the department to dictate the scholar's future research projects, demands that, in the context of the scholar's publication history, appeared designed explicitly to punish new research agendas.

Works Cited

Aiello et al. 2018 Aiello, Daniella, et al. "Eviction Lab Misses the Mark". *Shelterforce* (August 22, 2018), accessed Apr. 8, 2020. https://shelterforce.org/2018/08/22/eviction-lab-misses-the-mark/.

Arnstein 1969 Arnstein, Sherry R. "A Ladder Of Citizen Participation". *Journal of the American Institute of Planners* 35, no. 4 (1969): 216–24. https://doi.org/10.1080/01944366908977225.

Aspden 2019 Aspden, Suzanne. "Introduction". In Suzanne Aspden, ed., *Operatic Geographies: The Place of Opera and the Opera House*. University of Chicago Press, 2019.

Bamman et al. 2014 Bamman, David, Ted Underwood and Noah Smith. "A Bayesian Mixed Effects Model of Literary Character". In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014: 370-379.

Barthes 1974 Barthes, Roland. *S/Z: An Essay*. Trans. Richard Miller. New York: Hill and Wang, 1974.

Brennan 2017 Brennan, Timothy. "The Digital-Humanities Bust." *The Chronicle of Higher Education* (October 15, 2017). https://www.chronicle.com/article/the-digital-humanities-bust/.

CI Collective CI Collective. "CI Studies Bibliography". https://cistudies.org/critical-infrastructures-bibliography/, no date.

Cheney et al. 2012 Cheney, James, Anthony Finkelstein, Bertram Ludaescher, and Stijn Vansummeren. "Principles of Provenance (Dagstuhl Seminar 12091)". Edited by James Cheney, Anthony Finkelstein, Bertram Ludaescher, and Stijn Vansummeren. *Dagstuhl Reports* 2, no. 2 (2012): 84–113. https://doi.org/10.4230/DagRep.2.2.84.

Clement et al. 2017 Clement, Tanya E., et al. "Challenges for New Infrastructures and Paradigms in DH Curricular Program Development". *DH*, 2017.

Denis and Pontille 2014 Denis, Jérôme, and David Pontille. "Material Ordering and the Care of Things". *Science, Technology, & Human Values* 40, no. 3 (2014): 338–67. https://doi.org/10.1177/0162243914553129.

Edmond 2015 Edmond, Jennifer. "Collaboration and Infrastructure". In *A New Companion to Digital Humanities*, 54–65. John Wiley & Sons, Ltd, 2015. https://doi.org/10.1002/9781118680605.ch4.

Felski 2015 Felski, Rita. *The Limits of Critique*. University of Chicago Press, 2015.

Flanders and Hamlin 2013 Flanders, Julia, and Scott Hamlin. "TAPAS: Building a TEI Publishing and Repository Service". *Journal of the Text Encoding Initiative*, no. Issue 5, Issue 5, Text Encoding Initiative Consortium, Apr. 2013. *journals.openedition.org*, doi:10.4000/jtei.788.

Friedmann 1973 Friedmann, John. *Retracking America: A Theory of Transactive Planning*. Garden City, NY: Doubleday, 1973.

GO::DH 2018 GO::DH. "Minimal Computing (GO::DH Site)". *Minimal Computing*, 2018. http://go-dh.github.io/mincomp/about/.

GPP 2017 Grassroots Policy Project. "Building Social Movement Infrastructure". *Racial Equity Tools*, 2017. http://www.racialequitytools.org/resourcefiles/socialmovement_infrastructure.pdf.

Gil 2018 Gil, Alex. "Nimble Tents". College Park, Maryland: University of Maryland (Tuesday, November 6, 2018). https://mith.umd.edu/dialogues/dd-fall-2018-alex-gil/.

Gold 2018 Gold, Matthew K. "Response to Critical Infrastructure Studies Panel", 2018. https://hcommons.org/deposits/item/hc:17831/.

Gourley and Battino Viterbo Gourley, Donald and Paolo Battino Viterbo. "A Sustainable Repository Infrastructure for Digital Humanities: The DHO Experience." Digital Heritage, Euro-Mediterranean Conference. Springer, Berlin, Heidelberg, 2010, 473–81, https://doi.org/10.1007/978-3-642-16873-4_38.

Graham and Thrift 2007 Graham, Stephen, and Nigel Thrift. "Out of Order: Understanding Repair and Maintenance". *Theory, Culture, and Society* 24, no. 3 (2007): 10–11. https://doi.org/10.1177/0263276407075954.

Guldi 2012 Guldi, Jo. *Roads to Power: Britain Invents the Infrastructure State*. Cambridge, Mass.: Harvard University Press, 2012.

Guldi 2017 Guldi, Jo. "A History of the Participatory Map." *Public Culture* 29, no. 1 81 (January 1, 2017): 79–112. https://doi.org/10.1215/08992363-3644409.

Guldi 2018a Guldi, Jo. "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora". *Journal of Cultural Analytics* (2018), https://doi.org/10.22148/16.030.

Guldi 2018b Guldi, Jo. "Global Questions About Rent and the Longue Durée of Urban Power, 1848 to the Present". *New Global Studies* 12, no. 1 (2018): 37–63. https://doi.org/10.1515/ngs-2018-0012.

Henke 1999 Henke, Christopher R. "The Mechanics of Workplace Order: Toward a Sociology of Repair". *Berkeley Journal of Sociology* 44, no. 1999–2000 (1999): 55–81. http://www.jstor.org/stable/41035546.

Jyoti 1992 Jyoti, Hosagrahar. "City as Durbar: Theater and Power in Imperial Delhi". *Forms of Dominance: On the Architecture and Urbanism of the Colonial Enterprise*, 1992, 83–105.

Kraicer and Piper 2019 Kraicer, Eve and Andrew Piper. "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction." *Journal of Cultural Analytics* (2019), https://doi.org/10.22148/16.032.

Lansdall-Welfare et al. 2017 Lansdall-Welfare,Thomas, Saatviga Sudhahar, et al. "Content Analysis of 150 Years of British Periodicals". *PNAS (Proceedings of the National Academy of the Sciences)*, vol. 114, no. 4, Jan. 2017, pp. E457–E465, doi:10.1073/pnas.1606380114.

Liu 2012 Liu, Alan. "Where Is Cultural Criticism in the Digital Humanities?", in Matthew Gold and Lauren Klein, eds, *Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2012), accessed January 29, 2019,

http://dhdebates.gc.cuny.edu/debates/text/20.

Liu 2016 Liu, Alan. *Drafts for Against the Cultural Singularity* (book in progress). 2 May 2016. http://liu.english.ucsb.edu/drafts-for-against-the-cultural-singularity

Locke 2006 Locke, Ralph P. "*Aida* and Nine Readings of Empire". *Nineteenth-Century Music Review*, vol. 3, no. 1, June 2006, pp. 45–72. *Crossref*, doi:10.1017/S1479409800000343.

Lothian and Phillips 2013 Lothian, Alexis and Amanda Phillips, "Can Digital Humanities Mean Transformative Critique?" E-Media Studies 3.1 (2013): DOI: 10.1349/PS1.1938-6060.A.425.

Mattern 2014 Mattern, Shannon. "Library as Infrastructure". *Places Journal*, June 2014. *placesjournal.org*, doi:10.22269/140609.

Mattern 2018 Mattern, Shannon. "Maintenance and Care". *Places Journal*, 2018. https://placesjournal.org/article/maintenance-and-care/.

McPherson 2010 McPherson, Tara. "Scaling Vectors: Thoughts on the Future of Scholarly Communication". *Journal of Electronic Publishing*, vol. 13, no. 2, Fall 2010, doi:http://dx.doi.org/10.3998/3336451.0013.208.

Nimble Tents "What is the Nimble Tents Toolkit?"https://nimbletents.github.io/, no date.

Posner 2013 Posner, Miriam. "No Half Measures: Overcoming Common Challenges to Doing Digital Humanities in the Library". *Journal of Library Administration* 53, no. 1 (January 1, 2013): 43–52. https://doi.org/10.1080/01930826.2013.756694.

Rheingold 2008 Rheingold, Howard. "Participative Pedagogy: For a Literacy of Literacies". In *Freesouls*, ed. Joi Ito. Taiwan: Adams, 2008. http://freesouls.cc/essays/03-howard-rheingold-participative-pedagogy-for-a-literacy-of-literacies.html.

Risam 2019 Risam, Roopika. New Digital Worlds : Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy. Evanston, Illinois: Northwestern University Press, 2019.

Rockwell 2010 Rockwell, Geoffrey. "As Transparent as Infrastructure: On the research of cyberinfrastructure in the humanities". *Online Humanities Scholarship: The Shape of Things to Come*. Edited by Jerome McGann. Rice University Press, 2010. https://cnx.org/contents/PVdH0-lD@1.3:_USvuzFn@2/As-Transparent-as-Infrastructure-On-the-research-of-cyberinfrastructure-in-the-humanities.

Sawyer et al. 2019 Sawyer, S., Erickson, I., & Jarrahi, M. "Infrastructural competence." In J. Vertesi & D. Ribes (eds.), *Digital STS: A field guide*. Princeton, NJ: Princeton University Press, 2019, 235-252.

So et al. 2019 So, Richard, et al. "Race, Writing, and Computation: Racial Difference and the US Novel, 1880-2000". *Journal of Cultural Analytics*, 2019. *Crossref*, doi:10.22148/16.031.

Starosielski 2015 Starosielski, Nicole. *The Undersea Network*. Raleigh: Duke University Press, 2015.

Underwood et al. 2018 Underwood, Ted, et al. "The Transformation of Gender in English-Language Fiction". Cultural Analytics. February 12, 2018. https://culturalanalytics.org/article/11035.

Vacano et al. 2017 Vacano, Claudia von, Abigail T. De Kosnik, and Stephen Best. "Digital Humanities at Berkeley and the Digital Life Project". In *DH*, 2017.

Van Zundert 2012 Van Zundert, Joris."If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities". *Historical Social Research / Historische Sozialforschung* 37, no. 3 (141) (2012): 165–86.

Vertesi and Ribers 2019 Vertesi, Janet, and David Ribes. *DigitalSTS: A Field Guide for Science & Technology Studies*. Princeton University Press, 2019.

Vismann and Winthrop-Young 2008 Vismann, Cornelia, and Geoffrey Winthrop-Young. *Files: Law and Media Technology*. Stanford, Calif.: Stanford University Press, 2008.

Wagner et al. 2015 Wagner, Claudia, et al. "It's a Man's Wikipedia? Asssessing Gender Inequality in an Online Enyclopedia". The International AAAI Conference on Web and Social Media (ICWSM 2015), https://arxiv.org/abs/1501.06307v2.

Weil 2006 Weil, Benjamin. "The Rivers Come: Colonial Flood Control and Knowledge Systems in the Indus Basin, 1840s-1930s". *Environment and History*, 2006, 3–29.

Williams 2008 Williams, R. "Cultural Origins and Environmental Implications of Large Technological Systems". *Science in Context* 6, no. 02 (2008): 377–403.

Zhang et al. 2019 Zhang, Siyang, Nicholas Shapiro, Gretchen Gehrke, Jessica Castner, Zhenlei Liu, Beverly Guo, Romesh Prasad, et al. "Smartphone App for Residential Testing of Formaldehyde (SmART-Form)". *Building and Environment* 148 (January 15, 2019): 567–78. https://doi.org/10.1016/j.buildenv.2018.11.029.