Toward a Diversity Stack: Digital Humanities and Diversity as Technical Problem

ALAN LIU

ALAN LIU, distinguished professor in the English department at the University of California, Santa Barbara, has worked in the areas of DH, Romantic literature, and literary and cultural theory. His most recent book is *Friending the Past: The Sense of History in the Digital Age* (U of Chicago P, 2018). He leads the 4Humanities project WhatEvery1Says, a DH big-data initiative funded by the Andrew W. Mellon Foundation that studies public discourse on the humanities.

HOW CAN THE SOCIAL AND ETHICAL COMMITMENT OF THE DIGITAL humanities to diversity, which it shares with much of the contemporary humanities, prompt the creation of an adequate platform, at once ideological and technical, for advancing scholarship related to diversity—a platform that DH might help develop as its unique contribution to the broader goal of communicating a humanities vision of diversity in a larger society where data science and machine learning are increasingly the currency of shared knowledge? Can ideas about diversity that are now core elements of what the humanities know contribute to what data science and machine learning know? In return, can the humanities' sense of diversity be inflected with new knowledge from those fields?

As in the humanities generally, *diversity* in DH means the inclusion of different cultures, identities, and people as research themes and as members of the scholarly community itself, with an emphasis on underrepresented races, ethnicities, genders, national origins, languages, abilities, and socioeconomic classes (including institutional and professional classes, which in DH means non-R1 universities and "alt-ac" scholar-practitioners). Somewhat at odds with the term as it is used in the broader humanities, *diversity* in DH also denotes an ideal of inclusive methodology—for example, coding *and* interpreting, curating large collections *and* making one-off exhibits, text encoding *and* text analysis, and researching *and* teaching. At base, though, methodological inclusion is also sociocultural inclusion because its aim is respect for the variety of DH practices worldwide.[1]

The overarching metaphor for all this diversity in DH is the "big tent." Connoting inclusiveness in all the senses mentioned above, *big tent* is the aspirational meme for the ideological side of the DH diversity platform. The meme originated at the Alliance of Digital Humani-

© 2020 ALAN LIU
PMLA 135.1 (2020), published by the Modern Language Association of America

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

ties Organizations (ADHO) conference in 2011 (DH2011), which emphasized methodological diversity.[2] As Matthew Jockers and Glen Worthey, the conference organizers, explained:

> The theme we've chosen for DH2011 is "Big Tent Digital Humanities," and this was meant, in part, to convey our own desire to include . . . the many different varieties of DH practice that surround us right here at Stanford. At the same time, the "big tent" seemed to us an appropriate metaphorical response to some of the debates that have flourished in the digital humanities worldwide, especially in the past few years, about the meanings and limits of the Digital Humanities designation. Although DH2011 belongs to everyone who participates in it, by choosing this theme we meant to announce publicly our own opinion that a broad and diverse and vibrant DH field, one in which a thousand flowers might bloom, is the sort of DH we at Stanford believe in and hope to promote. (vi)[3]

The next year's ADHO conference, in Hamburg, then foregrounded the sociocultural side of diversity by highlighting multiculturalism and multilingualism in its theme, "Digital Diversity: Cultures, Languages and Methods" ("Call for Papers").

Subsequently, DH blossomed, if not with the thousand flowers Jockers and Worthey called for (a common, tenfold misquotation of Mao Zedong's "let a hundred flowers blossom, let a hundred schools of thought contend"),[4] then with the meme of the big tent itself. Mentions of *big tent* echoed throughout DH in mixed tones of celebration, caution, qualification, criticism, and irony.[5] The topic to which this special issue of *PMLA* is dedicated, Varieties of Digital Humanities, is itself partly inspired by the meme. According to the program description for the panel of the same title at the 2018 MLA convention, DH is "expansive, movable, but precarious, a tent still not big enough in terms of diversity and access" ("347").

The figure of the big tent, of course, comes down to us from old-timey showcases of mass experience such as nineteenth-century religious camp-meeting revivals and big-top circuses.[6] Those were just two of the earlier tented architectures, institutions, and spectacle-forms, complete with their discursive conventions and belief systems ("Do you believe the incredible?," they both asked), that—to use Michel Foucault's term—were *dispositifs* ("apparatuses") of variety. Foucault defined a *dispositif* as itself a various form—a kind of metavariety:

> a thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions—in short, the said as much as the unsaid. Such are the elements of the apparatus. The apparatus itself is the system of relations that can be established between these elements. (194)

In particular, the open yet enclosed architectural, institutional, and discursive structures of tent revivals and big-top circuses were protomodern apparatuses of mass variety. They stage-managed the long, modernizing encounter—cultic, populist, democratic, or fascist—between an older, affinity-based sense of the *Volk* and an evolving awareness of ever greater social, gender, racial, class, geopolitical, and (in the case of circuses) even "special," in the sense of constituting a species, national or world variety. This was true of camp meetings and tent revivals associated with the Second Great Awakening in the United States, different versions of which drew the mass participation of black people, women, and geographically diverse populations in ways that had lasting sociocultural and political impact. It was true in another way of circuses. The largest big-top circuses, for example, were venues for witnessing a postimperial variety of

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

cross-species and racial and ethnic exoticism from around the world. As advertised in a late-nineteenth-century Barnum and Bailey poster depicting the world's human races and nationalities alongside animal species, circuses were "a glance at the great ethnological congress" and at a menagerie of "curious . . . animals" (*Barnum and Bailey* [fig. 1]).

We can easily add earlier and later apparatuses to the repertoire of modernizing open-yet-enclosed, inclusive-yet-exclusive showcases of mass variety—as witnessed, for instance, in the arena of the French revolutionary Champs de Mars, whose remaking for the 1790 Fête of Federation famously convened Parisians both low and high;[7] the searchlight-ringed Zeppelin field of the Nuremberg Nazi rallies parading the regional, age, and gender variety of the party's base;[8] the Maoist hundred-flowers campaign of 1956–57 (alluded to by Jockers and Worthey [vi]) with its "Democratic Wall" of "big character posters" at Beijing University (Meisner 179); and today's conceptual architecture of "open source" programming (not the "cathedral," Eric S. Raymond memorably said, but the "bazaar"). Even the most seemingly open paragons of the modern mass *communitas* among such examples appear from the viewpoint of some other ideal of openness to be closed police states. Richard Stallman, for example, asserts that open-source software is closed by contrast with the free-software movement.[9] The Nuremberg rallies and Mao's hundred-flowers campaign are extreme examples of how open variety can become its antithesis, closed purity—the "anti-rightist" campaign that followed Mao's hundred flowers, for instance, sent to the re-education labor camps the very intellectuals whose diverse ideas Mao had encouraged.[10]

One response in the DH community to the big tent ideal has thus been that the tent is still too closed. "Who's in and who's out?" has been a question in the field since the start of its big-tent ethos, when controversy arose in a DH panel at the 2011 MLA convention about whether one had to be a coder to be "in" the field.[11] However widely DH wanted to throw open its gates, it seemed to some that specialist, disciplinary, or institutional norms were gatekeepers. Especially concerned were scholars at the start of their careers whose focus on diversity and social justice made them wonder if DH was too closed for them. Members of the #transformDH collective movingly reflected:

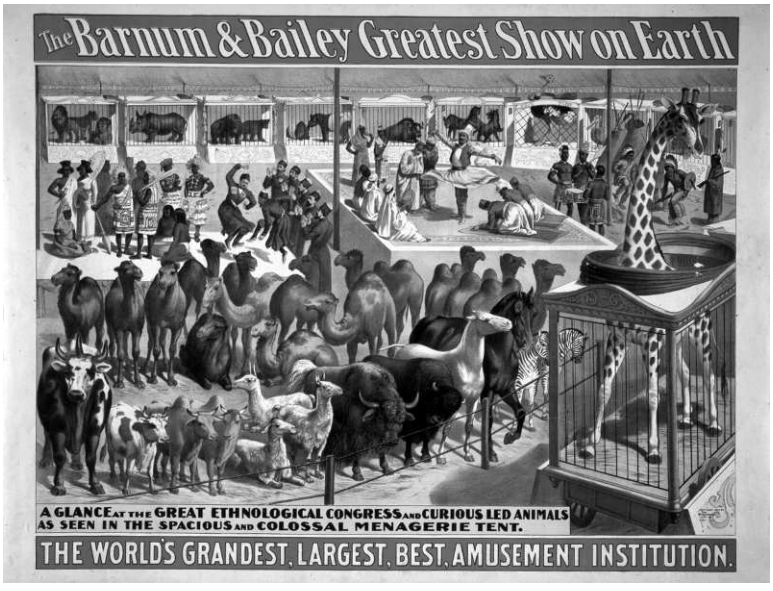


FIG. 1
Poster for the Barnum and Bailey Circus, circa 1895.

> It was 2011, the year that "Big Tent DH" surfaced as a term to describe digital humanities as inclusive and welcoming of different disciplines. But for those of us whose academic homes were in gender and queer studies, race and ethnic studies, and disability studies, and whose personal and political work embraced the digital, it appeared as if the "big tent" was not big enough. Our social justice concerns seemed to enter so rarely into conversations and research, even in the "big tent" of the field. (Bailey et al. 72)[12]

Named after its Twitter hashtag, the collective was, like DH in general, an early scholarly adopter

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

of social media, blogs, and online platforms, whether those platforms were provided by Silicon Valley firms or by the academy itself (beginning with platforms like the Humanist Discussion Group [humanist@lists.digitalhumanities.org] and HASTAC Scholars forum [www.hastac.org/initiatives/hastac-scholars] and going on to the newer professional association "commons" spaces such as MLA Commons [mla.hcommons.org/], descended from the CUNY Academic Commons [commons.gc.cuny.edu/]). Social media, blogs, and digital commons seemed to be a tent even more open than the "big" one.

At this point, however, we should pause to consider what else the robust adoption of online media in DH was signaling: the glaring mismatch between the metaphor of the big tent and the actual technical platforms enabling the field's work. *Big tent* named the DH diversity platform as ideology. But it contributed no guidance at all for it as a technical platform. Putting this more carefully: what is at stake is not a binary difference between ideology and technology but instead a not-yet-understood need in the field for a fused techno-ideological apparatus—a platform in all senses—that can do urgently needed work. As the gradual shift in metaphor from big tent to commons might indicate, that work is to engage the humanities in responding to how Silicon Valley firms with closed platforms have appropriated "variety" without attending to diversity: as clustered, filtered, and profiled "aggregation."

So what would be a capable "platform" for DH work on diversity—one that is both a metaphor for ideology and an operational technology?

Below, I suggest part of an answer. It is not the whole answer because it focuses on technical advances needed for researching and thinking about diversity but not on how to increase the diversity of people working in DH, the humanities, the academy, and knowledge-work institutions generally.[13] Consideration of the institutional dimension of inclusion as it affects, or is affected by, DH would require drawing on additional approaches such as neoinstitutionalism, organizational-technology studies, and critical infrastructure studies—approaches that I take up in other contexts.[14]

My answer here just tries to deploy for the digital humanities the full sense of "platforms" today as inextricably both ideological and technical. In particular, I draw on Benjamin H. Bratton's powerful theoretical exegesis—at once technological, sociopolitical, and philosophical—of platforms in his magisterial *The Stack: On Software and Sovereignty* (2015). Bratton makes a persuasive case that the most important platform concept today—underlying such other contenders as the network—is the stack, or "The Stack," as he refers to it throughout to emphasize its status as what is ultimately a world system. Stacks are our time's fundamental ideological-cum-technical platform paradigm. They are at once what Bratton calls a form of "sovereignty" (political and otherwise) and a set of standards and protocols installed in apparatuses all the way down to the environmental level that Bratton calls, in one of his most impressive chapters, the "Earth" layer (75–107).

There are many instances of stack platforms today, ranging in extent from fundamental software structures to Bratton's world-systemic "The Stack." The example that has become paradigmatic is the "Internet protocol stack," which I depict in simplified form in figure 2.[15] This diagram shows how information flows laterally on the Internet from "host" computer to "host" computer (and from hosts acting as "servers" to end users) through networks of routers. "Information" in this context means well-formed communications that are sequenced, structured, validated, and configured to be usable (e.g., as a "document" readable as an e-mail message or Web page). However, the actual data flow

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

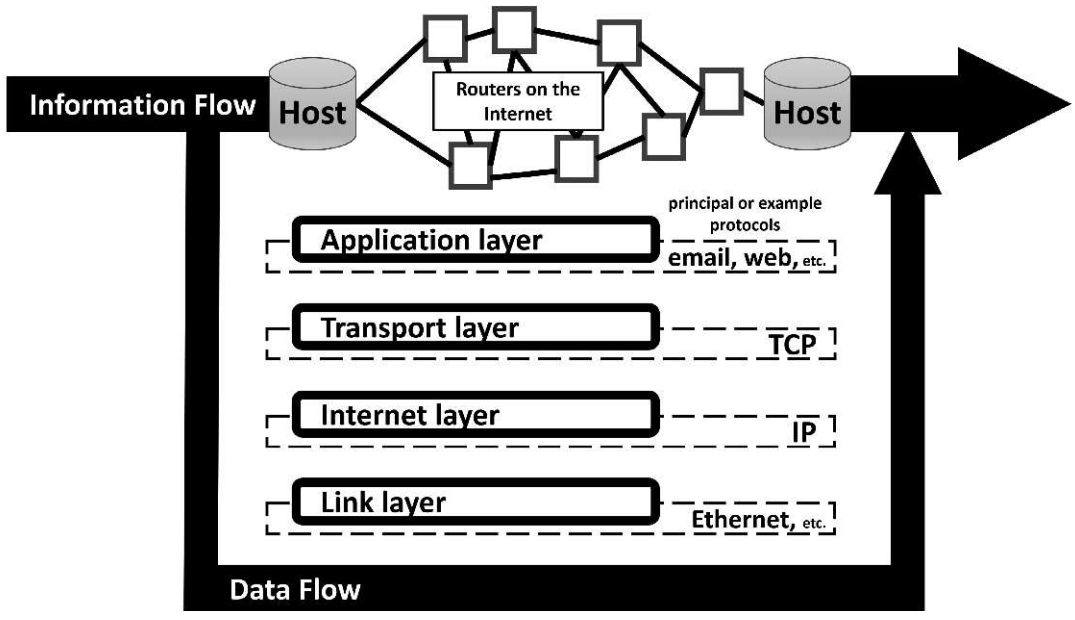


FIG. 2
Internet stack. Diagram by Alan Liu synthesizing and adapting many other available visualizations.

that conducts the information flow descends and reascends a vertical "stack" of protocols for formatting and handling data. As Bratton puts it, "In simple terms, a message is sent by a user down the stack, layer by layer, until it is transmitted laterally across the physical layer to the receiving node across the street or across the ocean. Upon being received, the message travels back up the stack, from the physical layer to the application layer, and is read by the next user" (61–62).

Each protocol layer implements standard ways of handling data at its level through what can be called data abstraction and modularity. *Abstraction* means that the lowest "link layer" transmitting data in a specific local area network (such as the Ethernet in one's department) is already an abstraction of the physical and electrical signaling media underlying everything. Then, unheeding of local systems, the "Internet layer" is a higher level of abstraction whose Internet Protocol (IP) transforms data into standard "data packets" adapted for transmission through a wider network. The "transport layer" is a yet higher level of abstraction whose Transmission Control Protocol (TCP) represents data packets as a data "stream" by ensuring proper packet sequencing, congestion avoidance, reliability, and so on. And at the top of the stack, the "application layer" (with transmission protocols such as *smtp*, for e-mail, and *http*, for the Web) is the level of abstraction supporting the complex software applications and interfaces that finally allow users to work with data as "information."

Modularity refers to the way each lower layer in the stack focuses on limited goals as simply and efficiently as possible, handing off more complex functions to the layer above it; while, reciprocally, each higher layer can remain oblivious to the inner workings of the layer below it. The clearest example is the modular separation of the "Internet layer" and "transport layer" according to the so-called end-to-end system-design principle.[16] "End-to-end" in this context means that the IP of the Internet layer just concentrates on pushing data packets through the routers of the Internet without guaranteeing their integrity or even arrival, while the TCP in the transport layer then operates on the more complex host computers at the "ends" of the transmission to support such higher-level tasks as establish-

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

ing the connection, checking the integrity of the packets, and arranging for their proper sequencing when they are received.

Abstraction and modularity in the IP stack led in principle to three consequences that significantly affected the combined technical and sociopolitical evolution of the Internet. One is layer or platform agnosticism, by which I mean that each layer in the stack does not care about, or even necessarily know, what is happening in the layer above or below it. This is the underpinning of so-called net neutrality, the tenet that lower data layers should not be throttled in speed or otherwise manipulated to determine, for instance, whose video-streaming service runs best on top of the stack. (The fact that layer agnosticism can in fact be compromised so that lower-level constraints hinder some services compared to others is why net neutrality is a contested issue.) The second consequence is freedom of innovation (in the idiom of Silicon Valley). New software and social uses can be slotted at will into the stack as long as their supporting layers continue working as expected. The third consequence is then a nuance of platform agnosticism and innovation that can be called layered emergence, a phenomenon with immense technical and ideological implications (including, as I will argue, in our context of diversity). What emerges stepwise from semiautonomous actions in the underlying layers of a stack platform is not deterministic or predictable and can be for better or worse. Some of the most important points in Bratton's seventeen-feature description of platforms relating to stacks are thus:

> 1. As opposed to other macrogovernance institutions, platforms do not work according to detailed premeditated master plans; rather they *set the stage for actions* to unfold through ordered emergence. (47)
>
> 7. Like centralizing systems, platforms consolidate heterogeneous actors and events into more orderly alliances, but they themselves are *not necessarily situated in a true central position* in relation to those alliances in the same way that, for example, a master planning committee or federal capitol building would be. (48)
>
> 9. Even as platforms guarantee identities to the *Users* of its systems, for better or worse, they do not provide these evenly or equally. (49)
>
> 17. Platform sovereignty may be planned or unplanned, universal or specific, generative or reactive, technologically determined or politically guaranteed. *Platform sovereignty is automatic under some circumstances and highly contingent under others, and it may function differently in relation to different components of the platform system.* (51)

The interaction of even the most standardized and seemingly well-behaved layers lower in stacks allows for the emergence higher up of divergent systems and behaviors, including what Bratton calls the "*User*" on the top layer of his world-system stack but what in our discussion of diversity we can call identity.[17]

To put into service the above schematic of platform stacks, my partial answer to the question of what would be a capable platform—at once metaphoric and operational—for diversity work in DH is not a big tent but a diversity stack. Of course, it would strain credulity to strive for exact correspondence between a "diversity stack" and specific information platforms like the Internet protocol stack. What is important in my suggestion is only the essential logic of stack platforms: modular layers of abstraction with their combined consequences. The diversity stack I outline is thus organized in layers of abstraction, each limited in its goals because it is about doing one thing well. Building capacity for diversity research is not the responsibility of any specific modular layer. It is the goal at the top where the stack as a whole might support fresh, applied ways of thinking about diverse identities.

Fair warning: the diversity stack I call for confronts DH with hard challenges for which there are no quick fixes. In all humility, DH

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

will need to collaborate with computer scientists, data scientists, *in silico* STEM scientists, social scientists, archivists, data journalists, and others outside the academy in Silicon Valley firms or nongovernmental public-service organizations like Datakind (www.datakind.org). The shared goal is to foster a virtuous circle in which research on diversity helps shape technical innovation and, in turn, technical innovation designs new ways to understand and act on diversity. And another caution: these are not even all the layers that need to be developed in a diversity stack, only some critical ones.

I start from below in my proposed diversity stack and climb upward. However, to compress my argument, even at lower levels I will indicate how a hand-off might be made to higher-level abstractions in technical approaches to diversity.

Multilingual DH

Since language is the source material of so much humanities work, a fundamental lower-level issue bearing on diversity is multilingualism.[18] DH should coordinate a multidisciplinary campaign to address the technical problem of digital work across languages, which the 2012 ADHO conference highlighted in its call for papers "fostering multilingualism and multiculturalism" and which the ADHO special-interest group Global Outlook::Digital Humanities (GO::DH [www.globaloutlookdh.org/]) has emphasized as a key diversity issue. As the Translation Toolkit working group of GO::DH puts it, "the complexity and sensitivity of the issue of multilingualism touches on the large scale problems of economic inequality, colonial history, and political issues that have shaped the world we live in and, by extension, the landscape in which we work" ("About: Why a Translation Toolkit?").

Today, the difficulty that DH has in working across languages impedes the comparatist research needed for diversity. To start with, as digital humanists such as Domenico Fiormonte and Isabel Galina Russell point out, DH has prioritized a geopolitically narrow range of languages (primarily English) as its object of study and its medium of professional exchange. It also draws its technical vocabulary (often baked into tools and encoding standards) mainly from English. This not only constrains the scope of research but hinders collaboration in the international DH scholarly community itself, though inventive social-engineering and other solutions have been devised for some multilingual scholarly activities. For example, GO::DH created a translation toolkit for use at academic conferences ("Translation Toolkit"), which the bilingual 2018 ADHO conference, in Mexico City, recommended in its call for papers ("Digital Humanities 2018").

Just as fundamentally, a core technical problem in DH is the inability to work with multiple languages simultaneously because DH tools are often functionally monolingual. For example, while text-analysis methods are language-agnostic, they usually only work meaningfully on a single language at a time, thus making it difficult to pursue comparatist approaches. DH should work as a multilingual collective to develop, systematize, automate, and scale up solutions like that for topic modeling English *and* Latin documents prototyped by James Jaehoon Lee and Paul Dilley in their project Mining the Renaissance ("Mining").[19] The goal would be to apply machine learning to extensive, mixed-language corpora (collections of texts) of the kind needed to investigate diverse cultures and many other comparatist issues—in the United States, for instance, corpora of English *and* Spanish news articles, novels, TV dialogue, and song lyrics. Research by computer scientists such as David Mimno and his collaborators on "polylingual topic models" indicates relevant new technical approaches (Mimno et al.). It is also conceivable that neural-network

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

translation methods like those deployed by Google in its translation platform since 2016 will someday allow DH to analyze text in multiple languages based on their intersecting "interlingua"—machine-generated, emergent, and transitional language forms that are a kind of pure comparatism.[20]

Such efforts could lead not just to operational advances in work on diversity but also—handing off to higher-level layers of abstraction—conceptual changes in such work. While multilingualism pursued through AI "interlingua" is not exactly "intersectionality" or "contact zones" (influential concepts related to diversity originally developed by scholars such as Kimberlé Crenshaw, Patricia Hill Collins, and Mary Louise Pratt and adapted for DH by Roopika Risam and Élika Ortega), it suggests interesting variations on those themes. For instance, what could the humanities add to their understanding of intersectionality and contact zones if they analogized them to the interaction of "artificial neuron" layers in AI neural networks? "Convolutional" neural networks for image recognition that work by sequentially detecting and pooling specific "features" in images may be a particularly apt example (and also provide a convenient segue to the multimedia problems I next take up). Critical determinants in such neural networks include the number of neuronal layers (each with many processing nodes) and the interactions in, and between, layers sensitized for the discrimination of different kinds of features. Could the humanistic concept of diversity be modeled analogously on how separated layers of identity nodes—in Crenshaw's own memorable analogy, people "stacked" in a "basement" one atop another at different levels of disadvantage (151–52)—interact socioculturally when discriminating (and discriminating against) the features of others?[21] (After all, we know that neural networks are already operating algorithmically in this way, if poorly so, through biased AI face recognition of races and ethnicities.) In other words, what patterns of intersectionality in addition to those already understood in intersectionality theory might the humanities discern through borrowing from new technical paradigms? (We note that the vocabulary of intersectionality theory—for example, "matrices" and "vectors" of domination—is already pseudo-mathematical in ways that seem ready to enter into conversation with technical models.[22]) And, reciprocally, how might humanistic insights, once conversant with new technical paradigms, then be added back to those paradigms to inflect them in turn?

Multimedia DH

Another lower-level issue bearing on diversity concerns digital images and audio. DH has been very text-centric. The reasons are many, including the preponderance of texts in the humanities, more flexibility in how texts can be used under copyright law, and the greater technical tractability of texts (which can be quantified and analyzed as discrete "tokens"). In addition, text materials can be managed computationally with the assistance of extensive, flexible, fine-grained, and open taxonomies (the basis of text encoding, for instance) that are generally either more robust or more widely used than equivalents for images or audio.[23]

Nevertheless, digital humanists have begun "distant reading" the audiovisual record of human experience. Projects include HiPSTAS (digital analysis of spoken-text audio files [blogs.ischool.utexas.edu/hipstas/2012/11/14/welcome-to-hipstas]), Lev Manovich's "direct visualizations" ("image graphs" of large numbers of scalable images arranged by their metadata [20–23]), and Frederic Brodbeck's "cinemetrics" (information about films "such as the editing structure, color, speech or motion . . . extracted, analyzed and transformed into graphic representations so that movies can be seen as a whole and easily interpreted or compared side by side").[24] However, there is much new, rapidly

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

advancing technical research in this area that could be explored by digital humanists. Again, neural-network AI, such as AI for image recognition, indicates new horizons. We might imagine applying an image-recognition neural network to large collections of visual materials important to the humanities and arts. It seems obvious, for instance, that Heinrich Wölfflin's influential early-twentieth-century *Principles of Art History*, which compares "linear" with "painterly," "planar" with "recessional," and "closed" with "open" styles based on formal "motifs," awaits a machine-learning reinvention that could produce equivalents of such observations from Wölfflin as the following: "Nothing is more illuminating than to compare the similar curve of the arm in the two pictures [by Botticelli and Lorenzo di Credi]. The sharp elbow, the spirited line of the forearm, the radiant spread of the fingers on the breast, the energy which charges every line—that is Botticelli" (2).

A specific research direction for DH related to diversity might be to train machine learning on image collections representing different nations, cultures, and epochs. This would allow for comparatist data analysis of depicted objects (e.g., What objects are in pictures? How many of each are there?), formal structures, color patterns, perspective systems, and material media. The goal would be to seek unexpected relations among images not accommodated as "movements" or "styles" in standard cultural histories segregated by Western, Eastern, African, indigenous, and other art. Comparatist work of this kind could be further enhanced by linking image metadata discovered through machine learning not just to image taxonomies like Iconclass but to named and geotagged entities in the historical textual discourse surrounding images. Or, again, DH might contribute to research on image machine learning itself by applying to recognition, classification, and annotation methods the long tradition of humanistic thought on *ut pictura poesis*, ekphrasis, and related verbal-visual problems, especially by updating the tradition to take account of diversity. For instance, digital humanists could provide a historically longer, more multicultural baseline for addressing such questions as, What defines an adequate textual description of an image?[25] Reciprocally, what is an adequate visualization of a text?

However, without further progress in scalable, granular, and flexible ways of analyzing multimedia, DH will not be able to explore such avenues embracing the full spectrum of graphical-cum-textual visual knowledge that Johanna Drucker calls "visual epistemology,"[26] not to mention aural, haptic, choreographic, and other epistemologies (8). The unabridged amplitude of culture will thus remain tantalizingly out of reach, including the multimedia corpora of diverse peoples whose past or present cultures flower most brilliantly in domains outside text—for example, in oral performance, art, decoration, music, and dance. So, too (handing off here to higher layers of abstraction), the potential for new ideas about the nature of diversity—ideas latent in today's digital laboratory of multimedia with all its algorithms for sampling, filtering, transforming, synthesizing, and other operations—will remain untapped. For instance, how might we understand diversity differently if we explore digital methods of mixing, equalization, or Gaussian blurring that suggest alternate paradigms of inclusion and exclusion? So much of today's multicultural experience, after all, comes by way of multimedia music, video, and other mixed performances at once channeled and altered through digital methods, which thus need to be studied for the way they are not just "effects" but conceptually and socially important ideas of "remix" diversity.

Representative DH Corpora

Above the layers in the diversity stack focused on textual and audiovisual media is the layer

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

of abstraction devoted to collection work. Just as the transport layer in the Internet protocol stack is more abstract than the underlying Internet layer because it makes data "streams" out of "packets," collection is higher in the diversity stack because it makes "corpora" and "archives" (and also, in the idiom of Commonwealth and other nations practicing DH, "heritage") out of media. The crucial questions bearing on diversity at this level concern not just the provenance and well-formedness of corpora but their representativeness. How do the facets of representativeness in a collection (social, political, economic, cultural, linguistic, religious, racial, ethnic, gender, and otherwise) intersect?[27] How can such facets be annotated and measured in computationally tractable ways? Further, in the case of missing or silenced archival materials, how can computational methods help us infer the fuller human pattern? Finally, how do digital methods of aggregation as they impinge on notions of statistical "normalization" (making varying scales or distributions of data comparable) affect our very "norms" of social representation and diversity today? These are specifically DH versions of some of the key questions about diversity and inclusion.

Though at first glance it may seem reductive and controversial, I suggest that a necessary step for digital humanists is to collaborate across nations and groups; consult with archivists, social scientists, and ethnographers; and comb the demographic-classification systems of the International Organization for Standardization (ISO), governments, and nongovernmental organizations (NGOs) to create a digitally tractable, extensible taxonomy of diversity—one that can be expressed in, or linked to, more general ontology protocols for provenance, prosopography, and social-event description requiring the categorization of actors and groups. Examples of such general protocols are the Text-Encoding Initiative (TEI) P5 guidelines (especially the section "Personal Characteristics" in chapter 13, "Names, Dates, People, and Places" ["Names"]); the Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF) schema (eac.staatsbibliothek-berlin.de/); the W3C PROV data model ("PROV-Overview"); and the Conflict and Mediation Event Observations (CAMEO) framework. For instance, CAMEO—an "event"-description standard arising from the international studies research field—includes a "codebook" (*CAMEO*) of actor and action designations designed to assist in computationally mining and analyzing textual materials related to contemporary political and military conflicts (with their linked actors, actions, and localities). The codebook's "Ethnic Coding Scheme," or CAMEOECS, taxonomizes with three-letter alphabetic codes almost six hundred ethnic groups from around the world in association with their "primary countries of settlement" (*CAMEO* 111, 113).[28] Approached in the right spirit (with consideration for the self-identification and multiple identifications of groups), an effort to adapt for the humanities an extensible, machine-readable diversity taxonomy of this sort—one that for computational tractability demands explicit, agreed-on declarations—could foster much-needed discussion and data-driven study of diversity otherwise blocked by incommensurable understandings of what we are talking about when we talk about diversity.

Why risk the seeming reductiveness and controversy of a diversity taxonomy?[29] Because, as indicated by the "conflict and mediation" framework abbreviated in the acronym "CAMEO," naming groups who live in tense, intersecting relations with each other is a first step in knowing how to gather data about the whole pattern of the human ensemble and use it for social-justice or any other kind of research, especially at big-data scales requiring computation. Because, as in the case of census forms or ethnicity checkboxes on college applications, people in diverse societies

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

in fact negotiate diversity taxonomies in real life, which should be better understood in their complex national and world interrelations. And also because in the DH field itself, there is too much possible international misunderstanding about diversity. Witness James Smithies's blog post of 2013 entitled "Speaking Back to America, Localizing DH Postcolonialism." Writing as a New Zealander giving pride of place to indigenous cultures in postcolonial discussions about diversity, Smithies responded to an initiative called Postcolonial Digital Humanities (#dhpoco [dhpoco.org/]), which had started up in the United States, by reflecting, "It didn't represent anything I understood postcolonialism to be, or speak to the postcolonial world(s) I know. Critical Theory, yes. A powerful melange of identity theory and feminisms, queer theories and activist politics, but not 'my' postcolonialism." "Some localization is needed," he argued, "because we need to ensure that indigenous peoples in our part of the world get to engage with DH as colonised people and have the full benefit of our local postcolonial discourse(s) when they do."

What Smithies terms "localization" is one goal of an extensible DH diversity taxonomy—a kind of "diversity XML" in which elements and attributes are defined for particular kinds or regions of diversity. The reciprocal goal is generalization: the ability to abstract notions of diversity systematically across the levels of individuals, groups, nations, and larger formations (all the way up, as it were, to the most common "namespace" in the schema, humanity).[30] Creating a shared, transparent DH vocabulary of diversity—one that facilitates navigation across identity elements and attributes—would better position DH, and the humanities generally, to intervene in the diversity taxonomies now used algorithmically by institutions, corporations, and governments for their own purposes: for instance, to determine who can immigrate, how voting districts are drawn, which consumers to target, and, in general, how resources and identity of all kinds are allocated. While there can never be one right taxonomy (because in the last analysis there is no uncontestable justification for any specific political-economic allocation of resources or identity), there can be different taxonomies—or, better, taxonomies of difference.

Assisted by an extensible diversity taxonomy, DH could go on to create computational ways to assess the balance and configurations of diversity in any corpus of materials (or other object of study)[31]; to represent the absent presence of undercounted, censored, or erased voices (e.g., through digital inference methods); and to triangulate the whole idea of a representative corpus with such older paradigms of representativeness as "archives," "editions," "canons," and (in a related but different sense of *corpus*) "corpus linguistics." A metastep—requiring a handoff to higher layers of software applications in the diversity stack—will then be to create tools that facilitate the assessment and enhancement of diversity by addressing not corpora but their collection platforms: digital repository systems (such as Samvera [samvera.org/]), "digital asset management" (DAM) systems, and "content management systems" (CMS). For example, imagine that improved versions of the kinds of metadata bearing on diversity that now appear in CAMEOECS or the TEI P5 prosopography (e.g., "faith," "language knowledge," "nationality," "sex," "age," "socioeconomic status," "occupation," and "education")[32] could as standard practice be entered in institutional repositories and database-backed DAM and CMS systems, thus allowing a whole secondary ecosystem of metadata-analysis tools and plug-ins to be created to calculate measures of diversity in collections according to optional filters (i.e., points of view about diversity) that are themselves diverse. Assessment tools of this kind could then be followed by linked-data or API methods for adding missing parts of the hu-

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

man picture—for example, by automatically amplifying a corpus or Web site with materials drawn from the Digital Public Library of America (dp.la/), Europeana Collections (www.europeana.eu/), or Humanities Networked Infrastructure (HuNI [huni.net.au]).

Of course, it may not ultimately be possible to access all materials required for a representative corpus, reach consensus on what norms of representativeness or even basic schemes of classification and organization should apply, or agree on the facets of a corpus to weigh or infer. But simply innovating better ways to declare, track, measure, standardize, and then work with collections based on metadata about diversity would go a long way toward achieving the ideal of what might be called (adapting Katherine Bode's notion of a "scholarly edition of a literary system" [97–102]) scholarly editions of diversity systems. Properly edited, a corpus does not just present a collection of materials but makes visible the editorial practices by which it was created, selected, balanced, corrected, and transformed to become a representation of sociocultural diversity.

The DH Chronotope

Seen at an even higher level of abstraction, corpora represent in their coverage of materials what are de facto temporal and spatial "worlds." The next layer up in the diversity stack I call for is thus (to use M. M. Bakhtin's term) an improved "chronotope" of those collected worlds. Especially needed is a digitally traversable and manipulable chronotope able to slice time and space differently to project diversity as an alternate distribution of human experience—for example, the experience of seasonal migrant laborers—over time intervals and geopolitical domains.[33] We might say by analogy that just as a "topic" in topic modeling is a "probability distribution over terms" in a document set (Blei), so diversity is a probability distribution of the "human" over the chronotope of historical, demographic, legal, economic, governmental, journalistic, literary, and other datasets, where that distribution is the outcome of past unjust allocations of space-time (e.g., territory or visa durations). As in a well-done topic model, modeling the theme of human diversity as a chronotope should not just confirm stereotypes but discover surprises about how people and their representations inhabit space-time.

Of course, much work in DH and related fields is already focused on spatial and temporal analysis. For example, robust branches of spatial research in digital history, literary studies, geography, and archaeology use such methods as mapping, geographical information systems (GIS), lidar site modeling, and, indeed, topic modeling.[34] Such spatial research addresses either (or both) what may be called the external and internal chronotopes. *External* means the real-world geographic domain of the actors, events, or works being studied, while *internal* refers to mentions or representations of place inside works (including in fiction) amounting to what the *Literary Atlas of Europe* calls the "geography of fiction" ("Mapping") and Elizabeth F. Evans and Matthew Wilkens term "the literary-geographic imagination." In regard to temporal study, a similarly broad interdisciplinary effort in DH has experimented with altering what might be called the normative "time window" of humanistic study. Traditionally, days, months, years, or decades are the intervals for studying cultural phenomena. But in a return to what Fernand Braudel called the *longue durée*, computational distant reading opens the time window wider. Representative literary-historical studies in this regard include Andrew Goldstone and Ted Underwood's "The Quiet Transformations of Literary Studies" and Underwood and Jordan Sellers's "The *Longue Durée* of Literary Prestige." In the history discipline, similarly, Jo Guldi has argued for using DH

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

methods to reemphasize the *longue durée* (Guldi; Guldi and Armitage). Meanwhile, the media-archaeology and electronic-literature cognates of DH shrink the time window to the scale of what Wolfgang Ernst calls the computational "microtemporal level" (97). For example, Jessica Pressman, Mark C. Marino, and Jeremy Douglass's *Reading Project: A Collaborative Analysis of William Poundstone's Project for Tachistoscope {Bottomless Pit}* provides a dazzling reading of a work of electronic literature that delves into code events at the millisecond level (Pressman et al., e.g., 39–40, 128).

But such spatiotemporal work in DH notwithstanding, the various disciplinary branches of the field have not yet collaborated to provide shared recommendations for how researchers who are not expert in spatiotemporal methods should describe, encode, and analyze such notions as world (and world system), nation, locality, period, era, trend, movement, generation, life, or event. Many DH projects thus mold space and time en bloc in such formations as "nineteenth-" or "twentieth-century" works in "Britain" or "America." Or they just accept as their chronotopical donnée the boundaries of such corpora as Early English Books Online (EEBO), Eighteenth Century Collections Online (ECCO), or the HathiTrust Digital Library—at most adjusting the space-time parameters of such corpora through relatively light methods of spatial analysis (e.g., based on place of publication or named-entity-recognition geotagging of mentioned locations) or temporal segmenting (e.g., breaking a corpus into decades). Much of the more nuanced spatiotemporal research mentioned above is thus left on the shelf, along with even more leading-edge research today in such areas as event modeling (e.g., Karl Grossner's essays "Topotime" and "Event Centrality"), longitudinal topic modeling (e.g., Blei and Lafferty; Wang and McCallum; Wang et al.), and advanced timeline visualization (e.g., Kräutli).

This matters for work on diversity because en bloc formations of space and time taken "as is" are precisely the ones that have been historically closed to diversity. What is needed for any kind of DH work, but especially research related to diversity, is a kind of congress of DH and other chronotopical specialists to produce a set of protocols, data-file templates, scripts, and examples for working with space-time that is informed by advanced spatiotemporal research but designed for use by nonspecialist DH scholars wishing to open up the fissures and alternate formations in the great en bloc chronotopes so as to welcome diverse cultures into our space-time commonplaces. We need a standards recommendation, in other words, that essentially declares, Here are at least four things you should do to make your project chronotopically meaningful as well as open to diverse understandings of how people live in space and time. Again, the idea of extensible standards applies, since different scales and kinds of space-time description or analysis are appropriate for different research materials. But the main idea is that DH would gain from having a shared understanding of how to expose the space-time information in a corpus; regularize that information using gazetteers, archival authority files, and biographical sources; infer additional information using machine-learning tools; and model and transform the resulting chronotope.

This is how to reimagine the chronotope as a plenum of spaces and times that from the first, and not as an afterthought or at the margins, can be a home for diversity. A 2018 essay by James Jaehoon Lee, Blaine Greteman, Jason Lee, and David Eichmann is exemplary in this regard both for its innovative combination of DH methods and its focus on the chronotope of diversity. Titled "Linked Reading: Digital Historicism and Early Modern Discourses of Race around Shakespeare's *Othello*," the work reports on projects that embed "topic models and vector space models

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

within historical book networks . . . to explore both the language associated with race in texts that mention Moors and the extensive networks of booksellers, printers, and publishers who produced these texts." In particular, the work investigates early modern racial diversity as a geographic imaginary of difference to suggest "alternative definitions of early modern race in which geographical and spatial variables were at least as important as more readily recognizable factors such as skin color and anatomy in determining conceptions of race in Shakespeare's time."

DH Theory of Identity

At the top of my suggested diversity stack I place what is likely the highest abstraction layer related to thought about diversity: not the "*User*" in Bratton's world stack but what the contemporary humanities have deeply meditated on under the name *identity*. Drawing on multilingual and multimedia collections and the chronotopes of humanity they model, DH now has a chance to update for our data-science age the kind of knowledge about diverse identity that W. E. B. Du Bois and his team at Atlanta University presented in the American Negro Exhibit at the Paris Exposition Universelle in 1900 by means of sixty-three data visualizations reporting on the condition of black people (recently reproduced in the gorgeously illustrated *W. E. B. Du Bois's Data Portraits: Visualizing Black America—The Color Line at the Turn of the Twentieth Century*, edited by Whitney Battle-Baptiste and Britt Rusert [Du Bois, *W. E. B. Du Bois's Data Portraits*]). With new datasets, analyses, and visualizations, DH can now follow up on such work both by discovering additional answers about diversity and, equally important, by framing new concepts of it. Speculatively, DH might even contribute a DH theory of diverse identity.

By "discovering additional answers," I mean that the lower layers of multilingual, multimedia, collection, and chronotopic work that DH pursues will produce collections and data that variously confirm, adjust, extend, or correct what scholars understand about the past and present of multicultural identities.[35] Those layers (improvising on Du Bois's "color-line")[36] will, as it were, color in with actual data all the "mosaics," "kaleidoscopes," and "patchworks" allegorizing multiculturalism on the covers of so many literature anthologies in the 1990s.[37]

And by "framing new concepts," I mean what may turn out to be the most unique contribution DH can make to humanities scholarship on identity and diversity: serving as a conduit for data science and machine learning ideas that help us rethink not just "color lines" but the very notion of lines, bounds, or positions of identities.

Consider, for example, that from a data science and machine learning point of view the diversity stack I outline relies at every level on algorithms operating in high-dimension mathematical spaces. These are conceptual spaces in which—to take the case of text analysis—document sets consist of hundreds of thousands or more so-called dimensions (one for each unique term) measuring term frequencies or other attributes. In their interrelation these dimensions allow for the statistical recognition of patterns—for instance, which words or documents tend to co-occur, which are more "like" others or cluster together, and so on. I previously mentioned neural-network AI machine learning. An application of such AI for studying high-dimension language space that has generated great interest in DH is word embedding, which uses so-called shallow neural networks to model semantic relations between words as "vectors" positioned and oriented relative to other words in mathematical space so that those behaving in semantically similar ways are proximate to each other.[38] James Jaehoon Lee and his coauthors, for instance, adopt word embedding ("vector space models")

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

as one of their methods. A graph of words in vector space in their essay shows a word-embedding model of EEBO from 1500 to 1623 suggesting how the term *complexion* defined "a phenotype modulated by 'physicke,' 'constitution,' 'melancholy,' 'tempers,' and the 'temperature' of the local environment, far more than an immutable genotype" (fig. 3).[39] "Indeed," the authors add, "'blacke_moore' pairs with 'behavior,' rather than words associated with the body or skin color."

What this instance shows is that word-embedding models have the uncanny ability to reveal the cultural logic behind how societies speak, complete with biases. The standard example used to explain word embedding is the computationally generated answer to the question, King is to Queen as Man is to what? Based on normative English documents, the answer the computer suggests is an area in the high-dimension semantic space where the nearest word is *woman*. This is an obvious case that we do not need computation to solve, of course. But there are many non-obvious or controversial semantic logics in social discourse that read like a psychoanalysis of the biases that make diversity such a charged issue. For DH to shed its old-timey "big tent" paradigm in favor of a contemporary approach to diversity, it will need to explain—and, yes, also create metaphors for—high-dimension mathematical space. Such space, with all the domains of big data and the Internet it spans, is the new *big tent*.

Potentially new concepts of identity and diversity inhere in such DH space because the data lines (vectors) that become visible there do not behave like physical lines in analog space or color lines in society that either connect or divide, bridge or wall off. Instead, data science and machine learning analyze the relation of anything to any other thing in high-dimension spaces through myriad alternate methods of differencing and, reciprocally, similarity finding—for example, term frequency–inverse document frequency (TF-IDF), cosine similarity, principal component analysis (PCA), hierarchical and partitioning clustering,[40] topic modeling, word embedding, and social-network analysis (with the latter's metrics of centrality, degree, distance, betweenness, and so on). While high-dimension methods are necessarily also "dimensionality reduction" methods for making human sense of it all (e.g., as a two- or three-dimensional graph), such reduction is informed by permutational possibilities in the background suggesting novel or uncertain ways of thinking about identity.

In sum, there is a whole mathesis of identity applicable to intersectionality, contact zones, and other ideas that can inspire us to re-ask such questions related to diversity as, Are you like me? Are these million documents produced by or about your group like those bearing on my group? Or even, Are there different and like ways in which we can be different and alike? If identity, as the highest layer of the diversity stack, is the outcome of what I earlier called "layered emergence," then we need to understand at least in concept the mathesis of difference and similarity buoying it up from the lower levels of the stack so that, as I also put it, it is

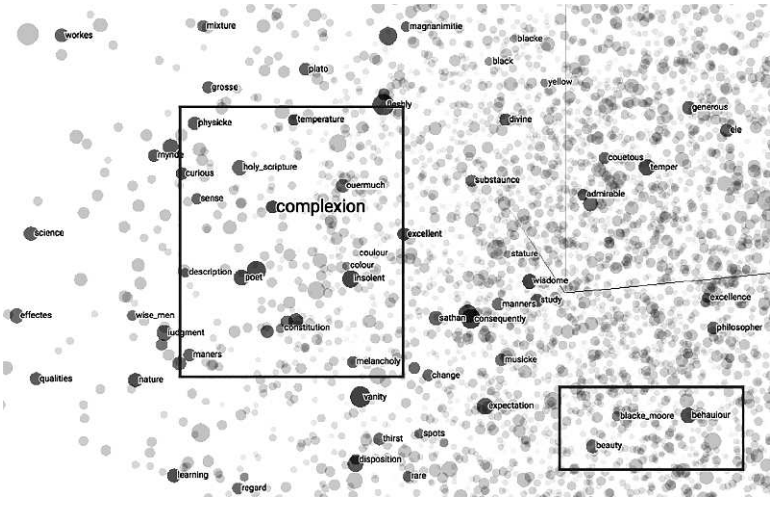


Fig. 3
Word embedding graph from Lee et al. CC Attribution 4.0 International license, https://creativecommons.org/licenses/by/4.0.

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

"not deterministic or predictable and can be for better or worse." DH's technical goal is to mine the mathesis of difference and similarity to explore unexpected formations, trends, and linkages influencing what we think identities have been and can be. And its critical goal is to assist the humanities in arguing for the better, not the worse, among possibilities.

After all, major corporate aggregators of online media now use big-data algorithms to interrogate similarity and difference (what their users "like," for instance). They do so because they want to own people's identities, aggregated in their "profiles." Meanwhile, there are governments that also interrogate data about similarity and difference (e.g., at borders) to construct their versions of profiled identities, not to own but to *dis*own them. DH needs to theorize and harness the profound power of contemporary differencing and similarity finding as part of its platform for researching—which is to say, not just interrogating but understanding and imagining—a more richly human, diverse future beyond the big tent.

Notes

This article is a significantly expanded and revised version of a talk I presented at the panel Varieties of Digital Humanities at the 2018 MLA convention (Varieties). The original talk was posted on my blog (Liu, "Digital Humanities Diversity").

1. See Earhart ("Digital Humanities") on localized DH practices around the world. My own visits to DH centers and events in other geopolitical regions have impressed on me the sense that different nations invest in different balances of the kinds of DH work. Countries that place a greater value on digitizing national or indigenous heritage, for example, concentrate more DH expertise and resources in their galleries, libraries, archives, and museums. While no large-scale, systematic comparisons of DH practices and priorities across nations have to my knowledge yet been made, some studies have been conducted on specific nations and DH sectors (see, e.g., Kizhner et al.; Lewis et al.; and Cha). My thanks go to @FrueheNeuzeit (Stefan Heßbrüggen), @JamesLeeUC (James Jaehoon Lee), and @brettbobley (Brett Bobley) for suggesting these sources in response to a query I sent out via *Twitter* on 16 March 2019 (@alanyliu, "Dear DH twitter" and "Place"). See Fiormonte and Galina Russell for astute critiques of what Fiormonte calls "geopolitical and . . . cultural-linguistic" imbalances in the DH field (440).

2. The conference of the Alliance of Digital Humanities Organizations is the field's premier annual international meeting. It is held each year in a different region of the world.

3. The conference's theme, "Big Tent Digital Humanities," was originally announced in its call for papers ("Call for Papers Announcement").

4. On this slogan, see Kraus.

5. Examples of works with *big tent* in their titles include Terras, "Peering inside the Big Tent"; Svensson, "Beyond the Big Tent" (and the section "Getting Rid of the Big Tent" in chapter 2 of his *Big Digital Humanities*); Earhart, "We Aren't the Big Tent"; Maria de la Cruz et al., "Mapping Cultures in the Big Tent"; Watrall, "Archaeology, the Digital Humanities, and the 'Big Tent'"; and Urberg, "Pasts and Futures of Digital Humanities in Musicology: Moving towards a 'Bigger Tent.'"

6. Terras attributes the trope more to religious than to circus history ("Peering"). I read the religious (especially evangelical) and mass-amusement sides of "big tent" tradition in a single, more general frame. Kyriakodis discusses a series of tent-revival events that took place in the United States during the so-called Third Great Awakening in 1858. See also my comments below on the Second Great Awakening. Circuses originated in the later eighteenth century, but their heyday was the nineteenth century, especially in the United States. The big tent (or "big top") circus was invented in 1825 (Davis).

7. Ozouf studied the ideology of openness staged by this arena and other French Revolution festivals. See, especially, her chapter "The Festival and Space" in *Festivals and the French Revolution* (126–57).

8. We ordinarily think of Hitler's Nuremberg rallies as showcases of racial purity rather than "variety." However, the impression of variety was certainly one of their mass and scale effects. Eyewitness accounts of the 1927 rally such as that in Rosenberg and Weiß emphasize the amassing of intra-German (and Austrian) ethnic and regional variety in marches on the field. In later rallies that included the Nazi Women's League and the girl's branch of the Hitler Youth, women were also part of the variety on show. See Bytwerk's account in the *German Propaganda Archive* of the women's contingent at the 1936 Nuremberg event.

9. Stallman writes, "Any activity that promotes the word 'open' tends to extend the curtain that hides the ideas of the free software movement."

10. On the antirightist campaign, see Kraus 254–57.

11. The lightning rod in this regard was Stephen Ramsay's talk at the 2011 MLA convention, "Who's In and Who's Out" (posted on his blog).

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

12. According to its Web site, #transformDH is "an academic guerrilla movement seeking to (re)define capital-letter Digital Humanities as a force for transformative scholarship by collecting, sharing, and highlighting projects that push at its boundaries and work for social justice, accessibility, and inclusion" ("About #transformDH").

13. For discussion and data-driven analysis of diversity in the DH scholarly community (especially as reflected in participation in the annual ADHO conferences), see Eichmann-Kalwara et al.

14. See, e.g., my "Drafts for *Against the Cultural Singularity*" and the site *Critical Infrastructure Studies.org*, which I started and maintain (cistudies.org).

15. In describing the Internet protocol stack below, I am aided by plentiful technical explanations available online and by such works intended for humanities and other scholars as Bratton's book and James Smithies's chapter "Towards a Systems Analysis of the Humanities" in his *The Digital Humanities and the Digital Modern* (113–51). My description simplifies the Internet protocol stack by overlooking some of its specific protocols and features. I also do not discuss similarities to or differences from the earlier OSI (Open Systems Interconnection) stack model.

16. Though preceded by earlier developments, the now classic statement of the end-to-end principle is that of Saltzer et al. in the early 1980s.

17. Bratton envisions a contemporary age of "The Stack" with ascending layers of Earth, Cloud, City, Address, Interface, and User. The latter "user" of this stack, which for Bratton includes nonhuman, such as machinic, subjects, forms atop "columns" (as he puts it) of emergent, contingent connections up and down the underlying layers (66–72).

18. "Lower" is a relative concept. I do not attempt to discuss even lower levels of technical tasks such as Unicode character encoding and OCR (optical character recognition) that are also important for research on diversity, e.g., in regard to multilingualism.

19. My thanks go to Lee for providing additional information about the project.

20. On "interlingua" in Google's neural-network translation service, see Coldewey. For an excellent long-form article on Google's development of its neural-network-based translation service, which replaced its older system in 2016, see Lewis-Kraus.

21. Another particularly suggestive class of neural networks in this context is that of "generative adversarial networks" (GANs), which are also used for image recognition. The "adversarial" relation of the two networks involved in GANs (one "generative," the other "discriminative") suggests additional ways of thinking about what an "intersection" or "contact zone" might be. (For an introduction to neural networks, see "Neural Network," "Convolutional Neural Network," and "Generative Adversarial Network." I am also grateful to Fabian Offert, whose in-progress dissertation has taught me how the "interpretability" of image-recognition neural networks is of interest from a humanities perspective.)

22. For the idiom of "matrix" in intersectionality theory, see, e.g., Collins 227–29, and for "vector" in intersectionality theory, see, e.g., Ritzer and Stepnisky 241–43.

23. Iconclass is an example of an art-historical image-classification system (www.iconclass.nl/home). For good thinking on image taxonomies, see Burford et al. Examples of taxonomic research for computerized classification of music genres or sound effects include Li and Ogihara; and Moffat et al.

24. Complementary to Brodbeck's approach to the data analysis of films is another kind of datacentric method in film studies: one focused on data about the film industry. See, e.g., the Kinomatics project (kinomatics.com/).

25. For an example of computer science research on using machine learning for the textual annotation of images, see Murthy et al.

26. See also my discussion of "graphical knowledge systems" and "graphical knowledge modeling systems" in *Friending the Past* (87–92).

27. The following are some of the facets of source-material "representativeness" in the corpus of journalistic and other discourse that the in-progress 4Humanities.org project WhatEvery1Says (we1s.ucsb.edu/), which I direct, is assessing in its study of how the public views the humanities: nation and region of publication, genre (e.g., newspaper or magazine), medium (e.g., newspaper, TV, or social media), circulation (audience size), declared political identity, and self-identified association if any with a gender or sexuality group, religion, cultural heritage, or institutional sector (e.g., universities or the higher education profession). (My thanks go to Abigail Droge and Lindsay Thomas, who led the creation of the project's metadata for the aforementioned facets.) See Tahmasebi et al. for a thoughtful, systemic approach to the trade-offs between representativeness and reduction implicit in the process of interpretation, especially in DH but also in "traditional" humanities. For additional research bearing on the problem of representativeness or absent materials in corpora and archives, see "Bibliography."

28. For context about CAMEO and earlier international frameworks for encoding actors and actions in conflict events, see Gerner et al. and Heap et al. The full CAMEOECS list of ethnic groups is in table 5.1 of *CAMEO* (114–30). By my count (the table rows are unnumbered), it includes 594 groups. The CAMEOECS taxonomy was created through a carefully thought-out process that started with consulting the ISO codes for languages and the Ethnic Power Relations (EPR) dataset (*CAMEO* 111–12). My thanks go to Scott Kleinman for first directing me to CAMEO.

29. I am aware of the risks of the computationally oriented taxonomic approach to diversity scholarship I suggest. Greater taxonomic awareness and agency in this regard can be socially or ethically positive. (For an exam-

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

ple, see Terras, "On Changing.") But if executed poorly, it could also prompt dissensus among regions, nations, and groups; reinforce historical power differentials between those who get to classify and those who are classified; and in other ways prove to be oppressive. (For admonitory instances, see Bowker and Star's chapter "The Case of Race Classification and Reclassification under Apartheid" in their classic study of classification, *Sorting Things Out* [195–225] and Os Keyes's discussion of "administrative violence" in applying data science to queer people.) I venture to make my suggestion, however, because I am persuaded—especially after studying public discourse at the scale of millions of media articles and social-media postings for the WhatEvery1Says project—that there is a case to be made for the humanities, assisted by DH, to contribute its taxonomic perspective to the myriad taxonomies of diversity otherwise flourishing in the wild in popular culture and in the controlled vocabularies of state or organizational bureaucracies. An ethical code of conduct and agreed-on protocols for deliberation would need to be put in place to guide the creation of such a DH taxonomy of diversity. But the gain for all the complication and risk would be to help reshape computationally managed and analyzed classificatory metadata about the world's peoples more humanely.

30. In the computer science sense, a "namespace" is an abstract logical area (though specified by exact uniform resource identifiers, or URIs) bounding a particular domain. In that domain, a schema or other protocol establishes the conventions for naming entities and attributes. Namespaces prevent collisions in conventions. Text-encoding protocols such as XML, HTML, and TEI thus all accommodate the declaration of namespaces in which specific naming rules can be applied. In my analogy here, diversity in the wake of identity politics is a kind of namespacing. But, as in intersectionality theory designed to highlight messy collisions and overlaps, it is a namespacing that sometimes encourages rather than avoids namespace collisions.

31. Argamon and his fellow authors, for example, provide a breakdown of the metadata (including "name(s), race, age, gender, nationality, ethnicity, occupation, sexual orientation") in the database *Black Drama—1850 to the Present* (search.alexanderstreet.com/bld2). As Risam observes, their study is effectively "intersectional" in approach, asking the question, "Can the database be intersectional?"

32. See section 13.3.2.1 ("Personal Characteristics") of the TEI P5 guidelines ("Names").

33. On "chronotope," see Bakhtin's "Forms of Time and of the Chronotope in the Novel" in his *Dialogic Imagination* (84–258). For simplicity, I leave out of consideration another fascinating aspect of the problem: the relation between the chronotopic "world" represented by corpora and the institutional or organizational worlds modeled by archives curated according to *respect des fonds*, or "the principle of grouping records by the administration, organization, individual, or creating body in which they originated" (J. Bailey). J. Bailey provides an excellent critical discussion of the way the digital era challenges *respect des fonds*.

34. One of the most novel uses of topic modeling in DH is Schmidt's application of the method not to words in document sets but to latitude-longitude coordinates in historical whaling-ship logs ("When"). The "topics" returned in topic models of such data are the geographic routes of ship voyages (though these topic models can also be deceptive, Schmidt points out).

35. Risam surveys some DH projects that "provide models for how to approach DH through an intersectional lens" of diversity.

36. In *The Souls of Black Folk* (1903), Du Bois writes, "[T]he problem of the Twentieth Century is the problem of the color-line" (3).

37. Literature "readers" of the era with images resembling mosaics, kaleidoscopes, or patchwork quilts on their covers include Rico and Mano; Walker; and Perkins and Perkins.

38. For an explanation of word embedding, see Schmidt, "Vector Space Models."

39. The total corpus that Lee and his coauthors used to create topic models and word embeddings consisted of "all documents in the 25,363 text Early English Books Online (EEBO) / Text Creation Partnership Phase I corpus, and specifically . . . the 14,708 paragraphs (from 3,147 texts) in the EEBO corpus mentioning all variants of the word 'Moor,' including Shakespeare's plays and editions of *Othello*." They then partitioned the corpus into different temporal ranges to be analyzed for particular purposes.

40. On these kinds of clustering, see the Lexomics Research Group's guide to "cluster analysis" ("Cluster Analysis").

Works Cited

"About #transformDH." *#transformDH*, transformdh.org/about-transformdh/.

"About: Why a Translation Toolkit?" *Global Outlook:: Digital Humanities*, go-dh.github.io/translation-toolkit/about.

@alanyliu. "Dear DH twitter: my travels to DH centers, programs, etc. in different parts of the world over the years makes me guess that the balance of _kinds_ of DH work (measured by funding, researchers & practitioners, # of projects, etc.) is regionally different—e.g., in nations that." *Twitter*, 16 Mar. 2019, 2:30 p.m., twitter.com/alanyliu/status/1107031314044084224.

@alanyliu. "place a greater relative emphasis on digitizing national and/or indigenous 'heritage.' Can anyone remind me if there is a survey or study that backs up (or refutes) that impression with empirical

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

evidence, whether through a systematic census or other means?" *Twitter*, 16 Mar. 2019, 2:30 p.m., twitter.com/alanyliu/status/1107031390636302336.

Argamon, Shlomo, et al. "Gender, Race, and Nationality in Black Drama, 1950–2006: Mining Differences in Language Use in Authors and Their Characters." *Digital Humanities Quarterly*, vol. 3, no. 2, 2009, www.digitalhumanities.org/dhqdev/vol/3/2/000043/000043.html.

Bailey, Jefferson. "Disrespect des Fonds: Rethinking Arrangement and Description in Born-Digital Archives." *Archive Journal*, June 2013, web.archive.org/web/20170919162159/http://www.archivejournal.net/essays/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives/.

Bailey, Moya, et al. "Reflections on a Movement: #transformDH, Growing Up." Gold and Klein, pp. 71–79.

Bakhtin, M. M. *The Dialogic Imagination: Four Essays*. Edited by Michael Holquist, translated by Cary Emerson and Holquist, U of Texas P, 1981.

The Barnum and Bailey Greatest Show on Earth—A Glance at the Great Ethnological Congress and Curious Led Animals. . . . Circa 1895. Strobridge Lithography, Library of Congress Prints and Photographs Online Catalog, www.loc.gov/pictures/item/98500053. Chromolithograph.

"Bibliography—Corpus Representativeness." *WhatEvery1Says*, we1s.ucsb.edu/research-resources/we1s-bibliography/bibliography-corpus-representativeness/.

Blei, David M. "Topic Modeling and Digital Humanities." *Journal of Digital Humanities*, vol. 2, no. 1, Winter 2012, journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei.

Blei, David M., and John D. Lafferty. "Dynamic Topic Models." *Proceedings of the Twenty-Third International Conference on Machine Learning*, ACM, 2006, pp. 113–20. *ACM Digital Library*, doi:10.1145/1143844.1143859.

Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly*, vol. 78, no. 1, Mar. 2017, pp. 77–106.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. MIT Press, 1999.

Bratton, Benjamin H. *The Stack: On Software and Sovereignty*. MIT Press, 2015.

Braudel, Fernand. "History and the Social Sciences: The *Longue Durée*." *On History*, translated by Sarah Matthews, U of Chicago P, 1980, pp. 25–54.

Brodbeck, Fredreric. "Cinemetrics." *Cinemetrics*, cinemetrics.fredericbrodbeck.de/.

Burford, Bryan, et al. "A Taxonomy of the Image: On the Classification of Content for Image Retrieval." *Visual Communication*, vol. 2, no. 2, June 2003, pp. 123–61, doi:10.1177/1470357203002002001.

Bytwerk, Randall. "The Rally of the German Women's League." *German Propaganda Archive*, 1998, research.calvin.edu/german-propaganda-archive/pt36frau.htm.

"Call for Papers." *DH2012*, www.dh2012.uni-hamburg.de/conference/archive-calls/cfps/index.html.

"Call for Papers Announcement." *DH2011*, dh2011.stanford.edu/?page_id=97.

CAMEO: Conflict and Mediation Event Observations Event and Actor Codebook. Directed by Philip A. Schrodt, version 1.1b3, Mar. 2012, *Computational Event Data System*, eventdata.parusanalytics.com/cameo.dir/CAMEO.Manual.1.1b3.pdf.

Cha, Javier. "Digital Korean Studies: Recent Advances and New Frontiers." *Digital Library Perspectives*, vol. 34, no. 3, Aug. 2018, pp. 227–44, doi:10.1108/DLP-04-2018-0013.

"Cluster Analysis." *In the Margins*, by Scott Kleinman et al., version 35, 19 Aug. 2016, scalar.usc.edu/works/lexos/cluster-analysis?path=topics.

Coldewey, Devin. "Google's AI Translation Tool Seems to Have Invented Its Own Secret Internal Language." *TechCrunch*, 22 Nov. 2016, techcrunch.com/2016/11/22/googles-ai-translation-tool-seems-to-have-invented-its-own-secret-internal-language.

Collins, Patricia Hill. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 2nd ed., Routledge, 2000.

"Convolutional Neural Network." *Wikipedia*, 21 Mar. 2019, 21:19 UTC, en.wikipedia.org/w/index.php?title=Convolutional_neural_network&oldid=888858225.

Crenshaw, Kimberlé. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago Legal Forum*, vol. 1989, no. 1, Dec. 2015, chicagounbound.uchicago.edu/uclf/vol1989/iss1/8.

Davis, Janet M. "America's Big Circus Spectacular Has a Long and Cherished History." *Smithsonian.com*, 22 Mar. 2017, www.smithsonianmag.com/history/americas-big-circus-spectacular-has-long-and-cherished-history-180962621.

"Digital Humanities 2018: 'Bridges/Puentes'—Call for Papers." *DH2018*, dh2018.adho.org/en/cfp/.

Drucker, Johanna. *Graphesis: Visual Forms of Knowledge Production*. Harvard UP, 2014. MetaLABprojects.

Du Bois, W. E. B. *The Souls of Black Folk*. Edited by Brent Hayes Edwards, Oxford UP, 2007.

———. *W. E. B. Du Bois's Data Portraits: Visualizing Black America—The Color Line at the Turn of the Twentieth Century*. Edited by Whitney Battle-Baptiste and Britt Rusert, W. E. B. Du Bois Center at the University of Massachusetts Amherst / Princeton Architectural Press, 2018.

Earhart, Amy E. "Digital Humanities within a Global Context: Creating Borderlands of Localized Expres-

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

sion." *Fudan Journal of the Humanities and Social Sciences*, vol. 11, no. 3, 2018, pp. 357–69, doi:10.1007/s40647-018-0224-0.

———. "We Aren't the Big Tent: Disciplinarity and Digital Humanities." "Day of DH: Digitalculturebooks Authors Reflect on Growth and Growing Pains," edited by Korey Jackson, *Digitalculturebooks*, 27 Mar. 2012, www.digitalculture.org/2012/03/27/growth_growing-pains/.

Eichmann-Kalwara, Nickoal, et al. "Representation at Digital Humanities Conferences (2000–2015)." *Bodies of Information: Intersectional Feminism and Digital Humanities*, edited by Elizabeth M. Losh and Jacqueline Wernimont, U of Minnesota P, 2018, dhdebates.gc.cuny.edu/read/untitled-4e08b137-aec5-49a4-83c0-38258425f145/section/5dcc1fee-caef-4c10-aa3c-08a9bbf0b68b#ch06.

Ernst, Wolfgang. "Archives in Transition: Dynamic Media Memories." *Digital Memory and the Archive*, edited by Jussi Parikka, U of Minnesota P, 2012, pp. 95–101.

Evans, Elizabeth F., and Matthew Wilkens. "Nation, Ethnicity, and the Geography of British Fiction, 1880–1940." *Journal of Cultural Analytics*, 13 July 2018, doi:10.22148/16.024.

Fiormonte, Domenico. "Toward a Cultural Critique of Digital Humanities." Gold and Klein, pp. 438–58.

Foucault, Michel. *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*. Edited by Colin Gordon, translated by Gordon et al., Pantheon, 1980.

Galina Russell, Isabel. "Geographical and Linguistic Diversity in the Digital Humanities." *Literary and Linguistic Computing*, vol. 29, no. 3, Sept. 2014, pp. 307–16, doi:10.1093/llc/fqu005.

"Generative Adversarial Network." *Wikipedia*, 20 Mar. 2019, 09:18 UTC, en.wikipedia.org/w/index.php?title=Generative_adversarial_network&oldid=888618653.

Gerner, Deborah J., et al. "Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions." *Computational Event Data System*, eventdata.parusanalytics.com/papers.dir/gerner02.pdf.

Gold, Matthew K., and Lauren F. Klein, editors. *Debates in the Digital Humanities 2016*. U of Minnesota P, 2016.

Goldstone, Andrew, and Ted Underwood. "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us." *New Literary History*, vol. 45, no. 3, 2014, pp. 359–84.

Grossner, Karl. "Event Centrality." *Kgeographer*, 9 July 2016, kgeographer.com/event-centrality/.

———. "Topotime: Qualitative Reasoning for Historical Time." *Kgeographer*, 29 July 2013, kgeographer.com/topotime-qualitative-reasoning-for-historical-time.

Guldi, Jo. "Time Wars of the Twentieth Century and the Twenty-First Century Toolkit: The History and Politics of *Longue-Durée* Thinking as a Prelude to the Digital Analysis of the Past." *Between Humanities and the Digital*, edited by Patrik Svensson and David Theo Goldberg, MIT Press, 2015, pp. 253–65.

Guldi, Jo, and David Armitage. *The History Manifesto*. Cambridge UP, 2014.

Heap, Bradford, et al. "A Joint Human/Machine Process for Coding Events and Conflict Drivers." *Advanced Data Mining and Applications: ADMA 2017*, edited by Gao Cong et al., Springer, 2017, pp. 639–54, doi:10.1007/978-3-319-69179-4_45.

Jockers, Matthew, and Glen Worthey. "Welcome to the Big Tent." Introduction. *Digital Humanities 2011 Conference Abstracts*, pp. vi–vii. *Stanford University Library*, dh2011.stanford.edu/wp-content/uploads/2011/05/DH2011_BookOfAbs.pdf.

Keyes, Os. "Counting the Countless: Why Data Science Is a Profound Threat for Queer People." *Real Life*, 8 Apr. 2019, reallifemag.com/counting-the-countless/.

Kizhner, Inna, et al. "Accessing Russian Culture Online: The Scope of Digitization in Museums across Russia." *Digital Scholarship in the Humanities*, 19 Sept. 2018, doi:10.1093/llc/fqy035.

Kraus, Richard. "Let a Hundred Flowers Blossom, Let a Hundred Schools of Thought Contend." *Words and Their Stories: Essays on the Language of the Chinese Revolution*, edited by Ban Wang, Brill, 2010, pp. 249–62.

Kräutli, Florian. *Visualising Cultural Data: Exploring Digital Collections through Timeline Visualisations*. 2016. Royal College of Art, PhD dissertation. *Florian Kräutli*, www.kraeutli.com/index.php/2016/04/15/visualising-cultural-data.

Kyriakodis, Harry. "Before the Pope, the Days of the Revivalists." *Hidden City Philadelphia*, 26 Aug. 2015, hiddencityphila.org/2015/08/before-the-pope-the-days-of-the-revivalists/.

Lee, James Jaehoon, et al. "Linked Reading: Digital Historicism and Early Modern Discourses of Race around Shakespeare's *Othello*." *Journal of Cultural Analytics*, 25 Jan. 2018, doi:10.22148/16.018.

Lewis, Vivian, et al. *Building Expertise to Support Digital Scholarship: A Global Perspective*. Council on Library and Information Resources, 2015, www.clir.org/pubs/reports/pub168/pub168.

Lewis-Kraus, Gideon. "The Great AI Awakening." *The New York Times Magazine*, 14 Dec. 2016, www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

Li, Tao, and M. Ogihara. "Music Genre Classification with Taxonomy." *Proceedings: (ICASSP '05): IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*, vol. 5, IEEE, 2005, pp. 197–200. *IEEE Xplore Digital Library*, doi:10.1109/ICASSP.2005.1416274.

Liu, Alan. "Digital Humanities Diversity as Technical Problem." *Alan Liu*, 15 Jan. 2018, doi:10.21972/G21T07.

———. "Drafts for *Against the Cultural Singularity* (Book in Progress)." *Alan Liu*, 2 May 2016, doi:10.21972/G2B663.

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

———. *Friending the Past: The Sense of History in the Digital Age*. U of Chicago P, 2018.

Manovich, Lev. "What Is Visualization?" *Manovich*, 2010, manovich.net/index.php/projects/what-is-visualization.

"Mapping and Analysing the Geography of Fiction with Interactive Tools." *A Literary Atlas of Europe*, 2019, www.literaturatlas.eu/en/.

Maria de la Cruz, Dulce, et al. "Mapping Cultures in the Big Tent: Multidisciplinary Networks in the *Digital Humanities Quarterly*." *Repository for IVMOOC 2015: Visualizing the Digital Humanities Project*, 8 Dec. 2015, github.com/jkaupp/DHQ.

Meisner, Maurice. *Mao's China and After: A History of the People's Republic*. 3rd ed., Free Press, 1999.

Mimno, David, et al. "Polylingual Topic Models." *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, vol. 2, pp. 880–89. *Association for Computational Linguistics*, dl.acm.org/citation.cfm?id=1699627.

"Mining the Renaissance." *Digital Bridges for Humanistic Inquiry*. iowadigitalbridges.com/about/projects/dilley-lee-english-renaissance-latin-text-mining/.

Moffat, David Festus Charles, et al. "Unsupervised Taxonomy of Sound Effects." *Semantic Scholar*, Allen Institute for Artificial Intelligence, www.semanticscholar.org/paper/Unsupervised-Taxonomy-of-Sound-Effects-Moffat/e44e2f2c60e78b930eb64bdb583654309a9b1af6.

Murthy, Venkatesh N., et al. "Automatic Image Annotation Using Deep Learning Representations." *Proceedings of the Fifth ACM on International Conference on Multimedia Retrieval*, Association for Computing Machinery, 2015, pp. 603–06. *ACM Digital Library*, doi:10.1145/2671188.2749391.

"Names, Dates, People, and Places." *P5: Guidelines for Electronic Text Encoding and Interchange*, version 3.4.0, Text Encoding Initiative, 23 July 2018, www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html.

"Neural Network." *Wikipedia*, 20 Mar. 2019, 19:25 UTC, en.wikipedia.org/w/index.php?title=Neural_network&oldid=888691390.

Ortega, Élika. "*Zonas de Contacto*: A Digital Humanities Ecology of Knowledges." *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein, U of Minnesota P, 2019, dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/aeee46e3-dddc-4668-a1b3-c8983ba4d70a#ch15.

Ozouf, Mona. *Festivals and the French Revolution*. Translated by Alan Sheridan, Harvard UP, 1988.

Perkins, Barbara, and George B. Perkins, editors. *Kaleidoscope: Stories of the American Experience*. Oxford UP, 1993.

Pratt, Mary Louise. "Arts of the Contact Zone." *Profession*, 1991, pp. 33–40. *JSTOR*, www.jstor.org/stable/25595469.

Pressman, Jessica, et al. *Reading Project: A Collaborative Analysis of William Poundstone's Project for Tachistoscope {Bottomless Pit}*. U of Iowa P, 2015.

"PROV-Overview: An Overview of the PROV Family of Documents: W3C Working Group Note." *W3C*, 30 Apr. 2013, www.w3.org/TR/prov-overview/.

Ramsay, Stephen. "Who's In and Who's Out." *Stephen Ramsay*, 8 Jan. 2011. *Internet Archive*, web.archive.org/web/20121001072912/http://lenz.unl.edu/papers/2011/01/08/whos-in-and-whos-out.html.

Raymond, Eric S. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media, 2001.

Rico, Barbara Roche, and Sandra Mano, editors. *American Mosaic: Multicultural Readings in Context*. Houghton Mifflin, 1991.

Risam, Roopika. "Beyond the Margins: Intersectionality and the Digital Humanities." *Digital Humanities Quarterly*, vol. 9, no. 2, 2015, www.digitalhumanities.org/dhq/vol/9/2/000208/000208.html.

Ritzer, George, and Jeffrey Stepnisky. *Contemporary Sociological Theory and Its Classical Roots: The Basics*. 5th ed., SAGE Publications, 2017.

Rosenberg, Alfred, and Wilhelm Weiß. "Nuremberg 1927: Torchlight Procession, Dedication of Banners, Mass March." Excerpt from *Reichsparteitag der NSDAP Nürnberg 19./21. August 1927*, Verlag Franz Eher, 1927, pp. 47–52. *German Propaganda Archive*, research.calvin.edu/german-propaganda-archive/rpt27d.htm.

Saltzer, J. H., et al. "End-to-End Arguments in System Design." *ACM Transactions on Computer Systems*, vol. 2, no. 4, Nov. 1984, pp. 277–88. *ACM Digital Library*, doi:10.1145/357401.357402.

Schmidt, Benjamin M. "Vector Space Models for the Digital Humanities." *Ben's Bookworm Blog*, 25 Oct. 2015, bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html.

———. "When You Have a MALLET, Everything Looks Like a Nail." *Sapping Attention*, 2 Nov. 2012, sappingattention.blogspot.com/2012/11/when-you-have-mallet-everything-looks.html.

Smithies, James. *The Digital Humanities and the Digital Modern*. Palgrave Macmillan, 2017.

———. "Speaking Back to America, Localizing DH Postcolonialism." *Jamessmithies.org*, 4 Sept. 2013, jamessmithies.org/blog/2013/09/04/speaking-back-to-america-localizing-dh-postcolonialism/.

Stallman, Richard. "Why Open Source Misses the Point of Free Software." 2011. *GNU Operating System*, Free Software Foundation, 2016, www.gnu.org/philosophy/open-source-misses-the-point.en.html.

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press

Svensson, Patrik. "Beyond the Big Tent." Gold and Klein, pp. 36–49.

———. *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. U of Michigan P, 2016, doi:10.3998/dh.13607060.0001.001.

Tahmasebi, Nina, et al. "A Convergence of Methodologies: Notes on Data-Intensive Humanities Research." *Nina Tahmasebi*, tahmasebi.se/publication/2019-aconvergenceofmethods/.

Terras, Melissa. "On Changing the Rules of Digital Humanities from the Inside." *Melissa Terras' Blog*, 2013, melissaterras.blogspot.com/2013/05/on-changing-rules-of-digital-humanities.html.

———. "Peering inside the Big Tent: Digital Humanities and the Crisis of Inclusion." *Melissa Terras' Blog*, 26 July 2011, melissaterras.blogspot.com/2011/07/peering-inside-big-tent-digital.html.

"347: Varieties of Digital Humanities." *MLA 2018 NYC*. Modern Language Association, mla.confex.com/mla/2018/meetingapp.cgi/Session/2517.

"The Translation Toolkit." *Global Outlook::Digital Humanities*, go-dh.github.io/translation-toolkit/.

Underwood, Ted, and Jordan Sellers. "The *Longue Durée* of Literary Prestige." *Modern Language Quarterly*, vol. 77, no. 3, Sept. 2016, pp. 321–44, doi:10.1215/00-267-929-3570634.

Urberg, Michelle. "Pasts and Futures of Digital Humanities in Musicology: Moving towards a 'Bigger Tent.'" *Music Reference Services Quarterly*, vol. 20, nos. 3–4, Oct. 2017, pp. 134–50, doi:10.1080/10588167.2017.1404301.

Varieties of Digital Humanities. MLA Annual Convention, 5 Jan. 2018, New York Hilton Midtown.

Walker, Scott, editor. *Stories from the American Mosaic*. Graywolf Press, 1990.

Wang, Chong, et al. "Continuous Time Dynamic Topic Models." *ArXiv.org*, 13 June 2012, arxiv.org/abs/1206.3298.

Wang, Xuerui, and Andrew McCallum. "Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends." *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2006, pp. 424–33. *ACM Digital Library*, doi:10.1145/1150402.1150450.

Watrall, Evan. "Archaeology, the Digital Humanities, and the 'Big Tent.'" Gold and Klein, pp. 345–58.

Wölfflin, Heinrich. *Principles of Art History: The Problem of the Development of Style in Later Art*. Translated by M. D. Hottinger. 1932. Dover Publications, 1950.

https://doi.org/10.1632/pmla.2020.135.1.130 Published online by Cambridge University Press