Transforming Text: Four Valences of a Digital Humanities Informed Writing Analytics

Gregory J. Palermo, *Northeastern University*

Structured Abstract

- **Aim:** This research note narrates existing and continuing potential crossover between the digital humanities and writing studies. I identify synergies between the two fields' methodologies and categorize current research in terms of four permutations, or "valences," of the phrase "writing analytics." These valences include *analytics of writing*, *writing of analytics*, *writing as analytics*, and *analytics as writing*. I bring recent work in the two fields together under these common labels, with the goal of building strategic alliances between them rather than to delimit or be comprehensive. I offer the valences as one heuristic for establishing connections and distinctions between two fields engaged in complementary work without firm or definitive discursive borders. Writing analytics might provide a disciplinary ground that incorporates and coheres work from these different domains. I further hope to locate the areas in which my current research in digital humanities, grounded in archival studies, might most shape writing analytics.

- **Problem Formation:** Digital humanities and writing studies are two fields in which scholars are performing massive data analysis research projects, including those in which data are writing or metadata that accompanies writing. There is an emerging environment in the Modern Language Association friendly to crossover between the humanities and writing studies, especially in work that involves digital methods and media. Writing analytics

accordingly hopes to find common disciplinary ground with digital humanities, with the goal of benefitting from and contributing to conversations about the ethical application of digital methods to its research questions. Recent work to bridge digital humanities and writing studies more broadly has unfortunately focused more on territorial and usability concerns than on identifying resonances between the fields' methodological and ethical commitments.

- **Information Collection:** I draw from a history of meta-academic literature in digital humanities and writing studies to review their shared methodological commitments, particularly in literature that recognizes and responds to pushback against the fields' ostensible use of extra-disciplinary methods. I then turn to current research in both fields that uses and critiques computational techniques, which is most relevant to writing analytics' articulated focus on massive data analysis. I provide a more detailed explanation, drawing from my categorization of this work, of the conversations in digital humanities surrounding the digital archives that enable data analysis.

- **Conclusions:** A review of past and current research in digital humanities and writing studies reveals shared attention to techniques for tokenizing texts at different scales for analysis, which is made possible by the curation of large corpora. Both fields are writing new genres to compose this analysis. In these genres, both fields emphasize process in their provisional work, which is sociocognitively repurposed in different rhetorical contexts. Finally, both fields recognize that the analytical methods they employ are themselves modes of composition and argumentation. An ethics of data transformation present in digital humanities, however, is largely absent from writing studies. This ethics comes to digital humanities from the influence of textual studies and archival studies. Further research in writing analytics might benefit from reframing writing corpora as archives—what Paul Fyfe (2016) calls a shift from "data mining" to "data archaeology"—in its analyses. This is especially true for analyses of text, which in particular foreground writing and analysis of writing as acts of transformation.

- **Directions for Further Research:** I recommend that future efforts to find crossover between digital humanities and writing studies do so by identifying their common values rather than trying to co-opt language and spaces or engaging in broad definitional work. I further provide a set of guiding principles that writing analytics might follow in order to pursue research that draws upon and contributes to both digital humanities and writing studies. These research projects might consider and account for the silences of writing corpora—unseen versions of documents, and documents' elements not described in structured data—while attending to the silences that these efforts might in turn (re)produce.

Keywords: archives, definitions, digital humanities, methodologies, text analysis, transformation, writing analytics, writing studies

1.0 Background

At Modern Language Association 2017, writing assessment specialist and *Journal of Writing Analytics* Editor-in-Chief Norbert Elliot attempted to locate writing analytics and digital humanities together in CIP Code 23. The category in a scheme for the Classification of Instructional Programs, maintained by the National Center for Education Statistics, taxonomizes disciplines associated with English studies. Drawing from a definition offered in the second version of a Manifesto (2009) authored by some influential scholars in digital humanities, Elliot noted resonance between the two fields' programmatic emphases. "Manifesto 2.0" represents digital humanities' commitment to "co-creation" across disciplines using digital media. This shared emphasis with writing studies joins another on language and community, one that Elliot identified in recently departed MLA president Kwame Anthony Appiah's *Cosmopolitanisms* (2007): the book uses its eponymous philosophy to target the contemporary division of the world into creeds and cultures unfamiliar to and at odds with one another. A cosmopolitanist framework instead proffers inclusion by enunciating a common morality, with the goal of engendering relationships of mutual respect. During Appiah's term, Anne Ruggles Gere from writing studies was elected as Second Vice President, which puts her in line for the Association's presidency in 2018; Elliot portrayed this development as a cosmopolitanist uniting of divisions. Expecting an emerging space in the MLA friendly to crossover between humanities and writing studies signaled by this development, Elliot anticipated a future in which digital work in the humanities and digitally-based writing studies practices might inform one another.

For the purposes of this research note, one place to begin a discussion of crossover between digital humanities and writing studies in our current moment is Jim Ridolfo and Bill Hart-Davidson's 2015 *Rhetoric and the Digital Humanities*. Their collection hails itself as the first effort to explicitly "bridge scholarship in rhetorical studies and the digital humanities," which the editors hold to share both physical space (of conference venues) and cognitive space (of intellectual values) but rarely cite one another. Indeed, Ridolfo and Hart-Davidson's collection arises out of the long computers and writing tradition in writing studies. This tradition can be traced back to Cynthia Selfe's newsletter concerned with integrating computer technologies as tools in composition courses, which Selfe transformed in 1985 into the scholarly journal known as *Computers and Composition*. Following Gail Hawisher's 1989 criticism of the emerging field as too "technocentric" (p. 44)—focused on what writing studies might have to gain from using computerized tools, as opposed to critiquing the sociocultural environments that surround those tools—the journal's focus has expanded to include critiques of technology beyond software reviews. It became the seminal venue in writing studies (joined by *Kairos* in 1996) for discourse on the computer as a product and mediator of culture, in addition to a tool for composition, both in and beyond the classroom. An ecology has since grown around *C&C* that convenes on list-servs like TechRhet (est. 2000) and draws scholars interested in any and all intersections between writing and technology.

Its history runs parallel to the history of digital humanities (DH), a field that can trace its roots to the field of "humanities computing" that often finds its own narrative genesis in the mid-century punch-card computing work of Father Roberto Busa. From outside digital humanities, Busa's work automating the computational processing of "non-numerical, literary information" might tell the story of a field focused alone on what Susan Hockey has called "the applications of computing to research and teaching within subjects that are loosely defined as 'the humanities'" (Busa, 1974, p. 1; Hockey, 2004). This is the definition of DH that Ridolfo and Hart-Davidson invoke in their introduction. A characterization of digital humanities as "application," however, erases the degree of reflective attention that scholars in the humanities computing field that would become DH, like the scholars forming computers and writing, paid to digital methods and tools as they built and implemented them. The 1976 editorial that Busa wrote describing his research in the *Association for Literary and Linguistic Computing Bulletin* asks "Why Can a Computer Do So Little?" in its title. John Unsworth would later, in 2002, articulate humanities computing as "a way of reasoning and a set of ontological commitments" requisite for travel across sociocultural and institutional borders in a computational medium.

It is perhaps the ostensible newness of DH's focus on media, mediation, and materiality that compels scholars in computers and writing, including *Kairos*'s current editor Cheryl E. Ball, to report "relish[ing]" in computers and writing's antecedence while witnessing the hype[1] surrounding digital humanities[2] (Carter et al., 2015, p. 35). Ridolfo and Hart-Davidson accordingly spend much of the introduction lamenting the absence of the computers and writing tradition from digital humanities scholarship (p. 4), while at the same time unpacking the work that C&W has accomplished without needing to look to DH for guidance (p. 2). Maintaining this distance, *Rhetoric and Digital Humanities* is framed in terms of the benefits that rhetoric and composition scholars could glean from a strategic appropriation and application of digital humanities' spaces and language, what Joseph M. Moxley and Katie Walkup (2016) might recognize as "usability discussions" that exacerbate disciplinary tensions. Kevin G. Smith (2016) speculates that this pragmatism could be its own strategic move "to shake stubborn computer and writing scholars from their defensive positions into a more open dialogue."

The choice of a writing studies audience to the exclusion of digital humanists, however, unfortunately eclipses the efforts of the collections' contributors to exceed any "dispute over common territory" and build a bridge between writing studies and digital humanities. It further ignores a shared history of academic "marginalization" that contributors Shannon Carter, Jennifer Jones, and Sunchai Hamcumpai identify in the two fields. The authors cite C&W scholar Kathy Gossett and digital humanist Matthew K. Gold in the same space to remind us of a likeness between the fields' commitments (p. 37). This likeness should and does run deeper than a facile observation that they both make use of digital technologies, or even Ridolfo and Hart-Davidson's observation that both fields critique those technologies and build digital tools (p. 3). Writing studies and digital humanities share an explicit "focus on the sometimes unglamorous, hands-on activities such as writing, coding, teaching, and building" that is less common in the research programs of our colleagues, even as those faculty may contribute to some or all of these activities in other parts of their professional lives (Carter et al., 2015, p. 37). Both fields have had to negotiate our commitments to process, demonstrated in our pedagogy and research, with an ever "increasing pressure to find, present, and demonstrate results" from our institutions (Gold, 2015).

Appeals to usability take for granted common, if misdirected, knowledge in my home department of English: that writing studies and digital humanities provide the most opportunities among its "subfields" for institutional recognition. The same collaborative digital humanities projects and action research initiatives that we find less fruitful for tenure and promotion than single-authored monographs are, the story goes, paradoxically attractive to university

administrators. Grant money springs from our outwardly operationalizable and techno-progressive work—what Miriam Posner brands as "flash-in the pan Dean candy" (2015, p. 271). The association of our fields with continually-debunked[3] myths of utility and progressivist opportunity have fueled criticism of digital humanities and writing studies alike for being in bed with the neoliberal university and its systems of authority (Allington et al., 2016; Charney, 1996, p. 568). Not only are we supposedly overtaking what few tenure-track lines are being renewed, but we are also complicit in the trend towards adjunctification: the structures in which we produce our work invite ever-more-contingent "alt-ac" positions.[4] This is a minefield that scholars in both fields must navigate. Wendy Hui Kyong Chun adopts Lauren Berlant's phrase "cruel optimism" to describe the motivation for tactical embraces of technoprogressive institutional rhetoric, which she argues sustain us in the short term but destroy us in the long term (2015, p. 495-6). It may well be true, as Matthew Kirschenbaum has argued (2012), that leveraging hype can be beneficial. We must also contend with the reality that to do so—rather than objecting to the "inflated rhetoric" (Posner, 2016, p. 271)—only propagates the opportunity myths, reinforces the very structures that they claim we co-opt, and constructs us in opposition to our other colleagues as well as one another.

In this research note, I instead want to relate the points of methodological synergy between digital humanities and writing studies in the work most relevant to this nascent field of "Writing Analytics": that which employs and critiques the use of computational techniques for text analysis. To do so, I unpack the multiple valences of its name, which appears in the full title of this journal. While "writing" can be read as present participle modifying analytics, it can also be read as a gerund that denotes the action of writing. These valences, which I mean to overlap rather than demarcate, are as follows: (1) *analytics of writing*, the analysis we perform on and about corpora of writing using computational methods, which consider text as data to be tokenized at different scales; (2) *writing of analytics*, the practice of writing this analytical work in genres across media, especially in digital composing and visualization; (3) *writing as analytics*, which reflects our understanding of composition processes as forms of meaning making that are cognitive and analytical, processes that like computation simultaneously involve sociocultural and procedural scripts, scripting, and rewriting; and (4) *analytics as writing*, which captures how the data wrangling, data modeling, and data transformation that are part of analyzing data with computational methods are analogous to processes of writing.

As I conclude, it is this final valence—analytics as writing—that comprises what digital humanities at large can most lend to the development of writing analytics. I argue that while writing studies has established an ethics of

data *collection* from its methodological comfort with the social sciences, it might gain from digital humanities an ethics of data *transformation*. This ethics emerges from traditions in digital humanities influenced by textual studies, archival studies, and data analysis: they are each concerned with versions of textual data, keeping track of the versions, and documenting transformations between versions. I draw here from the literature on and my experience with curating and mining digital archives. Humanists' grappling with the responsible representing of changes in text—both by author and by editor—are typified in conversations about initiatives to digitize manuscripts like *Walden: A Fluid Text Edition* and the *Dickinson Electronic Archive*. As archivists, we arrange digital archives anticipating longevity and interoperable analysis (Bauman, 2011); as analysts, we intervene in the archive with the understanding that we incompletely access it, "contaminate" it, and remediate it. This transformation is especially apparent when we tokenize text in the pre-processing that allows for computational text analysis. We re-write the texts we analyze using code. Efforts to incorporate drafting practices in these analyses may bring our analyses closer to our understanding of writing as re-writing—and, further, expand conceptions of what constitutes "public" writing. The modeling and analyzing of always-already limited data, however, has the potential to perpetuate what Lauren Klein (2013) might call its silences. Reframing writing corpora as archives, and the analysis of corpora as archival intervention, transforms the work we have done, the work we are doing, and the work we might continue to do in the future, all of which draw from the concerns and techniques of multiple fields.

2.0 Resisting Disciplinary Definitions

The status of the digital humanities according to "Manifesto 2.0," as an amalgamation of practices across disciplines rather than a unified field, renders any attempt to locate digital humanities within a CIP a difficult, if not futile, task. Digital humanists work within a number of disciplines and sub-disciplines that already have their own CIP representations. We have coined the term "digital literary studies" to specify the digital humanities work realized in the literature subfields of English departments, which has been categorically redefined (the term is Stephanie L. Kerschbaum's, 2014, p. 10) to include humanistic critiques of digital methods as well as their application to answer literary research questions. This reciprocal move has facilitated the incorporation of some digital methods, and discourse on them, in disciplinary venues that have been more valid for literary scholars' tenure and promotion than our own niche venues. But there is digital humanities work closely aligned with literary studies that escapes this taxonomy, including research in humanities data modeling and digital scholarly

publishing; while relevant to English studies, this work is not necessarily directly applied to the reading of literary texts and cannot be wholly described by the "subfield's" name. Any similar effort at categorizing the digital work that resonates in digital humanities and writing studies, like the valences I am about to define, strategically circumvents disciplinary preconceptions rather than providing a representative or totalizing account of relevant work.

The "interdisciplinary origin stor[ies]" of writing studies and digital humanities identified by Carter et al., according to Alexander Reid, manifest in "identity challenges" that scholars in both fields have faced in a need to justify using methodologies that supposedly hail (from) fields not conventionally associated with humanism (2015, p. 15-16). Both fields, for example, have faced the challenge of making their strong traditions of empirical research legible to "traditional" humanities disciplines and subdisciplines, which have resisted them as a perceived encroachment of the social and physical sciences on the sanctity of English studies. Davida Charney (1996) has confronted criticism of technical writing for "not purging itself of lingering scientific propensities and for wavering resistance to dominant ideologies in the academy and the workplace" (p. 567-8). Charney blames writing studies' myopic view of the sciences for a persistent hostility in the field to supposedly objective[5] empirical methods, which would forgo the "self-critique, creative interpretation, and negotiation of meaning" intrinsic to humanistic inquiry and composition (p. 578). This mischaracterization has manifested more recently in a conversation "On the Value of 'Empirical Studies' in Rhetoric and Composition" through David Schwalm's Writing Program Administrator Listserv (WPA-L): the original poster repeats ad nauseam the same assumption that one "cannot study arts empirically" because "the contexts for writing, and especially concepts for observing how it is performed, are highly variable" and exceedingly complex (2015).[6]

A parallel is quite easily drawn here to disciplinary pushback against the digital humanities in literary studies. Computational text analysis, further reduced to Franco Moretti's implementation of a distant reading diametrically opposed to close reading, has been "the type of DH research that most often stands in as synecdoche for the larger whole of the digital humanities" (Gold, 2015). It is also a methodology that has been repeatedly portrayed in popular media as hypothesis-testing, and thereby the transformation of humanities into science (Leroi, 2015; Schulz, 2011). These editorials subsist on the humanities and sciences' apparent incompatibility. Literary studies has long stigmatized "scientific" quantitative methods, according to Julia Flanders, as "part of an ongoing activity of methodological self-definition which requires that the opponent be kept in view" (2005, p. 53). At the same time, provocateurs have channeled techno-utopian narratives of digital humanities to portray the field as "the next big thing" that

threatens to supplant "traditional" humanities scholars' work (Pannapacker, 2011). These narratives obscure the extent to which digital humanists make use of "traditional" humanistic practices and contribute to sustaining them. Instead, they offer humanists a choice between assimilation or creeping irrelevance. They echo the techno-utopian narratives surrounding Automated Writing Evaluation (AWE) Software (see Elliot et al., 2013), which allegedly threatens our shared positions as assessors of writing (Condon, 2013) and comes to stand in as effigy for the whole of what we might call writing analytics.

The definitional work in which we might indulge to dispute mischaracterizations of our fields or make them legible to institutional structures, however, is necessarily tendentious. Susan Hockey's definition of digital humanities by way of humanities computing, for example, shows its age of nearly two decades in what has been a high-velocity and accelerating field. Her attention, in her "History of Humanities Computing," to the application of procedural methodologies to humanistic questions was pivotal at the time of her chapter's publication: during the adolescence of the Text Encoding Initiative (TEI).[7] Hockey offers a narrative of the TEI's development as a counterpoint to the linguistic-focused concordance that was dominating computing applications "involving textual sources" since the mid-1980s, "leaving some more humanities-oriented conference participants out in the cold." She draws attention to the association of application with scientific methodology in a field that she recognizes as having needed "to embrace 'the two cultures.'" In order to combat this association, she stresses the TEI's humanistic emphases—despite its status as an application of language-level textual markup—and represents it as the pinnacle of humanities-focused computational scholarship. But the emphasis was never a definition—and, in fact, Hockey goes out of her way to disavow any definitional responsibility. Removed from its original context of its publication, her strategic focus on application can erase the critique that has been and continues to be important to digital humanities. Borrowing language from Unsworth's pioneering effort to give humanities computing boundaries might get us closer to a bipartite definition of digital humanities: one that accounts as much for what a humanistic approach offers our interfacing with data and machines as it does for the nontrivial application of those machines to teaching in learning in the humanities. Unsworth draws a distinction between the "tool from the various uses that can be made of it, if for no other reason than to evaluate the effectiveness of the tool for different purposes." This distinction, however, trades an over-emphasis on application for one that privileges digital humanities' 'maker culture'[8] over work in the field that does not involve tool or "thing" building.

Because of the repeated definitional work over the course of digital humanities' history, the question of "What is DH?" is a fraught and somewhat

tired one in the field. We have had hard feelings in digital humanities over exclusionary politics grounded in definitional anxiety (Ramsay, 2011). We have only recently come to the end of what Bethany Nowviskie has called the field's "Eternal September," in which traditions of scholarship have been disregarded and trajectories repeated in service of hype driven by its perceived newness (2010). In a recent blog post, Ted Underwood (2017) notes that we are past this preliminary first-wave definitional stage, and are now at a point where digital humanities "is a semi-normal thing." It is now, he says, "just a matter of doing the work." So, with the understanding that there is value in locating the work we do in existing and changing institutional ecologies, I caution against grand definitional gestures. If there is one tenet that I would argue most binds digital humanities, it is the premise that information is mutable, that data is transformable. Like Underwood, I am now "moving to talk about one small subfield of DH rather than...the whole thing." I am also "mov[ing]," to quote Moxley and Walkup, "beyond usability discussions and stakeholder theory into questions of how WA can be used to benefit multiple domains" (4).

3.0 The Valences of Writing Analytics

3.1 Analytics of Writing *(computational analysis that transforms text at different scales)*

Our current moment sees a repose in the ire and buzz surrounding digital humanists working in literary studies, one that Underwood observes to be marked by an absence of "jeremiads" and the inclusion of macroscopic digital analyses in disciplinary journals. This development is concurrent with a second linguistic turn in subfields of writing studies like Rhetorical Genre Studies, characterized by work such as Laura Aull's (2015). Aull uses computational linguistics to draw conclusions from recurring patterns in large corpora of academic writing: with n-gram analysis that breaks up (or "tokenizes") text in differently-sized windows, she identifies common language-level constructions among first-year (FY) writers and "expert writers" that demystify what each group values. She makes the case that these currently rare large-scale analyses are both possible and valuable, especially for analyzing rhetorical features across writing contexts that small in-context samples may not provide (p. 6-8). Aull is clear that to "draw on corpus linguistic methodologies" is not to sacrifice "attention to socio-rhetorical context" (p. 10). Nor is it to "supplant close analysis of individual texts and contexts" (p. 13). Accordingly, she takes care when building her corpus to include metadata that includes institutional context and genre information (p. 54-8); it is this

metadata that allows for her to make context-specific claims from her empirical analysis.

Like writing studies, digital humanities has been concerned not only with how the text is constituted, but also to what degree the claims we make about text using empirical methods should be evidentiary or suggest new possibilities (Flanders, 2005, p. 41). This stance guides explicitly "speculative" computing and rhetoric, which engage in ludic analysis and play, for sure (e.g., Nowviskie, 2014; Ramsay, 2014; Reid, 2015); it also guides the best research that sets out with comparably defined research questions. Aull's work provides evidence for knowledge about FY writing that we may already intuit and take for granted (p. 87); it also allows space for preliminary findings to shape the directions that continued analysis takes.[9] Aull reports that "prompt distinctions...*emerged* as most significant for differentiating the FY writers across subcorpora" (p. 15, my emphasis). The "balance of possibility and certainty" that she attributes to expert writing is also a feature of expert data analysis (p. 16).

The emergence of new relevant subcorpora throughout analysis calls attention to the different scales at which we might break down our corpora. Our findings influence both how we subset our corpora and how we choose to tokenize their text, which reveal additional research questions in an iterative and recursive process—one that Frederick W. Gibbs and Daniel J. Cohen call "a conversation with [the] data" (2011, p. 70). Concern over how to assemble, wrangle, and subset humanistic "big data" (although Aull's corpora might be considered "medium data") is nothing new. Louis T. Milic imagines "The Next Step" in 1966, in the premier "digital humanities" journal *Computers and the Humanities*, for dealing with the "information explosion" caused by manually generated concordances and collates (p. 3). Milic meditates on the size and dimension of patterns that can be perceived without and with the aid of digital technologies (p. 5). Susan Lang and Craig Baehr, in a September 2012 special issue of *College, Composition, and Communication* on research methodologies, provide an overview of the ways that data and text mining could be used as research methodologies in writing studies. Using "scale" to mean 'scope' rather than 'granularity,' Lang and Baehr reckon on the potential for "scalable investigations" provided by large electronic archives of writing as sources of data (p. 172). The computational techniques that facilitate these methods assume that text is "massively addressable at different levels of scale" (Witmore 2012). Flanders (2013) has called for scholarship in digital humanities that varies the "z-axis" between close and distant analyses, which the field has responded to in its meta-academic conversations. Gold and Klein's *Debates in the Digital Humanities 2016* includes a "Forum" section on "Text Analysis at Scale." The question of scale in computational analyses has infiltrated disciplinary journals as

well, exemplified by a special issue in *Modern Language Quarterly* on "Scale and Value: New and Digital Approaches to Literary History." Among the methodological concerns addressed are meditations on how we may break corpora up into "documents," consisting of works, clusters of works, chapters, paragraphs, or other arbitrarily-bounded sections.

The humanities—and, I contend, writing studies—would like to imagine that the transition between these scales is seamless and that insights from analysis applied at one scale will easily apply at others. The example research questions that Lang and Baehr provide would not require particularly large datasets because they are asked of a single writing program (p. 173). These questions in their current form would not necessarily scale up, either, across writing programs. Anna Lowenhaupt Tsing reminds us that most complex phenomena are nonscalable because of their heterogeneity. This feature of our multivalent corpora, however, is one that she argues we should embrace. This is because "contact across difference"—what she labels "friction"—"can produce new agendas" (2012, p. 510). Ryan Cordell (2015) portrays choices of often irreconcilable scales as "act[s] of deformance" in a passage that invokes the language of quantum mechanical collapse:

> [. . .] each act of measurement—each time we freeze the textual system in place in order to make an observation—is an act of deformance. We address this scene, this theme, this argument, this vocabulary, in order to better know this poem, this book, this oeuvre, this corpus. In doing so we learn something true, but we also distort the system, lending outsized importance to our object at the expense of those textual features outside our purview.

While literary studies and writing studies alike have tended to focus on small samples to describe macro-level constructs, work like Aull's instead focuses on micro-level features across a swath of samples. She finds it necessary to return to the language of documents in order to establish patterns common to them. Given Tsing and Cordell's critiques, we might interrogate her tendency to 'zoom in' in granularity even as she 'zooms out' in scope. Can writing analytics extend Aull's framework to additionally identify the macro-level structural patterns (e.g., argument, ethos, structure, voice) that she claims have been the domain of writing assessment, making genre-based claims about them across corpora (p. 8)? In other words, how might micro-level features differently signal these larger structures depending on their local contexts of use, and what might the friction between the two reveal? To address these questions would be to attempt to reconcile the

distant quantitative analysis that allows for scale with a commitment to genre-based rhetorical claims.

3.2 Writing of Analytics *(composition of analysis that transforms genres and media; digital composing)*

It is appropriate that this quantitative research has emerged in fields that have both been instrumental to and benefitted from new media theory and ecological frameworks, which have foregrounded materiality and reframed genres as cultural artifacts that are sites of social action. Because writing, especially in the digital age, is a social activity that "blur[s] the boundaries between writer and audience" (Lunsford, 2015, p. 21), these fields turning a critical eye to new media are more likely to experiment with interactive genres or participate in collaborative work (Reid, 2010)—work that doesn't always fit the bounds of traditional academic publishing and isn't always legible to disciplinary and administrative structures. In recent work that regards quantitative analysis as, to quote Underwood, "itself a mode of interpretation," we create propulsive "friction" (Horner & Lu, 2014, p. 119) when translating the visual and reactive affordances of computational work to the existing refereed genres for relaying research.

Responding to what Risa Applegarth labels "rhetorical scarcity" (2014, p 29), both fields are founding new journals and building new platforms for them that facilitate composition of these methods' processes as well as their products. Andrew Piper started the *Journal of Cultural Analytics* shortly before the *Journal of Writing Analytics* was founded; both are online, open-access publications that define new genres to include computational techniques and code. Matt Gold and Douglas Armato's *Manifold*, recently released in beta, promises to provide a venue for interactive and iterative monographs that incorporate multiple forms of media and social annotations. All of these comprise efforts to balance our fields' articulated open missions (Bazerman et al., 2008; Spiro, 2012, p. 24-25) with efforts to make incremental counting work count. While building and using these platforms, we take into account research on interface theory (e.g., Drucker, 2011a; 2013). This work brings the usability concerns of the human-computer interaction (HCI) field, taken up by writing studies in early C&W research (Selfe & Selfe, 1994), to critique and guide how we present data and its analysis to readers (e.g., Drucker, 2011b).

These tools and platforms with which we compose come to mirror the sociohistorical grammars that govern our use of technologies for quantitative analysis: Hadley Wickham's "ggplot2" R package (2010), for example, adopts Leland Wilkinson's *Grammar of Graphics* (2005) to allow programmers to stack

graphical elements in their plots. Digital humanists doing analytical work have embraced Donald E. Knuth's literate programming, which embeds the code enacting our transformation and visualization in order to narrativize it. The effort to operationalize and explain what pieces of our code are doing, when the genre is appropriate, fulfills part of digital humanities' commitment to an open research exchange: not only open access to research findings, but open access to platforms and code. As Knuth writes, "instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do" (1984, p. 1). Writing analytics might thus devote additional attention to publishing the code underpinning tools as it builds tools for writing assessment and publishes findings using those tools. As it continues to adopt and create platforms that permit the inclusion of code, writing analytics might contextualize textual transformations for human audiences, anticipating how others might themselves re-write these transformations to learn and use them.

3.3 Writing as Analytics *(writing as socio-cognitive, scripted, and transformative processes)*

We now take for granted that we write to learn as much as we write to communicate; writing is then an opportunity to "grapple with complex ideas, test out [our] understanding, and incorporate new knowledge into old" as much as a rhetorical act. This reframing offers one point of narrative entry into our firmly-established pedagogical emphasis on process(es) over product—on "writing as input" rather than an output with the sole purpose of communicating already-held knowledge or assessing literacy (Anson, 2015, p. 207-8). Digital humanities has a long tradition of "grey literature" that documents the incremental thinking and decisions of scholars in community spaces outside (often prior to, and sometimes in lieu of) academic publication (Visconti, 2016). Deborah Brandt (2015) addresses the extent to which writers draw on their knowledge to take advantage of their multiple literacies to assess rhetorical situations in what she calls the "deep writing" necessary in a digital age. Writing, Brandt argues, has become the analytical skill of consequence because our era requires us to "choos[e] language to mediate multiple interests, audiences, regulations, and knowledge, ventriloquizing and twisting these systems" (p. 160). As we expand our notion of what constitutes composing (e.g., Shipka, 2011), we become aware of issues surrounding our diverse interactions with technology that surpass "technological issues affecting access, funding, and literacy" (Selfe, 1999, p. xvi) to include sociocognition. Robert J. Mislevy (2016), for example, brings a situative/sociocognitive perspective to question validity in writing assessment, an

approach that considers the interaction between writers' social practices and the linguistic, cultural, and substantive patterns of their environments.

Joseph Harris (2006) has more explicitly represented this cognitive endeavor of writing as one of re-writing and transformation: "com[ing] to terms with a text by translating its words and ideas into your own language, making them part of your own prose." In the language of mathematics, this transformation is not a bijection, or what we could call a paraphrase; "your focus as a writer," Harris continues, "soon shifts away from simply restating what that text has to say and toward the uses you can make of its concepts and phrases, or toward the gaps and problems you encounter in trying to do so" (p. 32). Harris continues on to foreground the technologies we employ to enact this transformation: re-writing is "not only re-presenting the work of another writer but also, at times, actually retyping it" to situate it in one's own context. Jason Swarts (2009) studies the rhetorical purposes of content reuse in technical communication, classifying reuse strategies into situations in which chunks of text are moved versus those in which text is changed "in meaning, appearance, or function" (p. 137). Kevin Roozen (2010) attends to how disciplinary writers more broadly repurpose practices from extra-disciplinary contexts across media. To write is to reorder and inject or surject. Independent of whether or not it may be algorithmically executed or processed, writing is thus a form of codification, or coding.

3.4 Analytics as Writing *(computational analysis as its own transformative process of composition)*

The slippage between codification and computer coding is significant to writing analytics. To quote Jeff Rice, "a program is a script" (2013, p. 370). If we accept that writing is a process of sociocognitive scripting that rewrites existing scripts, what does this reveal about the communication of scripted instructions to a computer? Or, given Knuth's understanding of literate programming, what does this reveal about the communication of a set of computer instructions communicated to an audience of people according to certain conventions? More succinctly, if writing is a form of coding, how might coding be a form of writing? Annette Vee traces the comparison between programming and literacy "echoed so frequently that it is more than just rhetorical flourish" (2013, p. 43). Just as Roozen notes that writers transform writing practices as they repurpose them, Vee notes that as code is adapted for use in fields outside of computer science, programmers ascribe new values to code in new contexts (p. 56). Vee does not claim that coding's compositional status constitutes an equivalency between natural and programming languages, any more than Shipka might argue that composing with text and dancing are identical communicative practices. The

discreteness and explicitness of coding languages makes coding not a literacy practice, "but a (potential) literacy on its own, with a complex relationship to textual literacy" (p. 47). But the constraints of code don't constrict programmers; code has style, which is critiqued by the somewhat niche field of critical code studies. Vee reframes computer programming as a "computational literacy" to "open up our concepts of writing to include programming" and use it more meaningfully in writing pedagogy (p. 60). Kevin G. Smith has already begun to answer this call, using XML-markup in the classroom to teach writing in order to "defamiliarize students' notions of composing by changing their primary compositional tool." Smith "hope[s] that having students apply markup (which is essentially metadata) to their texts will promote a metacognitive awareness of the (often implicit) rhetorical choices that are made during the composition process" ("(Re)orienting").

Once we become familiar with coding as a literacy, we might consider the analytics for which code provides as a literacy. In other words, the transformations enacted by coding processes are practices of meaning making that also involve rhetorical choices of scale and codification. Katie Rawson and Trevor Muñoz (2016) problematize the phrase "data cleaning" to mean the preparation of data for analysis, often performed with algorithmic processes, since it presumes an ideal eventual form according to "scripts, linguistic rules, and machine learning." Rawson and Muñoz argue that such standardization is reductive and precludes data diversity: it may be of research interest, for example, to retain spelling and formatting variations, both of which would otherwise introduce friction that keeps data from being easily tallied and filtered. The data we assemble have semantics. In addition to the grammar of graphics package, Hadley Wickham has designed a suite of R packages called the `tidyverse` that facilitate the importing, tidying, and transforming of what is called "tidy" data. These are data which "take on a consistent form that matches the semantics of the dataset with the way it is stored" (Grolemund & Wickham, 1.1). Other elements of the "tidy" coding philosophy impose grammars upon not only the shape and semantics of our data, but our composition of code that wrangles that data. The package also, for example, introduces a 'pipe' function that extends and restructures the language's functional syntax: while R usually wraps functions that act on the results of other functions around each other recursively, piping allows for a comparatively sequential (operational) writing and processing of stacked transformations.

While making one's data "tidy" may seem to preclude the diversity of data for which Rawson and Muñoz call, the architecture of the tidyverse also provides for a responsible "cleaning" that leaves the data in the state in which it was taken up, by stacking functions that enact pattern-based cleaning. This paradigm mirrors

non-destructive editing practices by creative image professionals that enact transformations on 'raw' files (to the extent that files can be "raw") in real time from metadata. Rawson and Muñoz offer the alternative of changing data to allow for filtering not with code that standardizes but with additional data structures. An index, what Janet Giltrow (2002) might recognize as a meta-genre, would account for variation and context by pointing to a more homogenous set of names or categories "around and atop the existing [heterogenous] data." The results of our analytical transformations are then new compositions that provide for repurposing through analysis. Digital humanities has been comfortable with this idea that data structures and platforms themselves make rhetorical arguments—as Tom Scheinfeldt puts it succinctly in a tweet, "DH arguments are encoded in code" (2011). Alan Galey and Stan Ruecker (2010) make the case that a digital prototype conveys an argument about how interfaces should be designed and data should be modeled, with the goal of opening up digital prototypes for scholarly peer review. These data structures—whether the corpora that comprise datasets we build independently, or more conspicuously-curated collections—are dynamic rather than static. In a recent Twitter conversation with Alex Gil, Humanities Data Curator Thomas Padilla expressed a desire to take up dynamic data citation work by computer scientists, including Andreas Rauber and Stephan Pröll, in a Research Data Alliance Working Group on Data Citation. These scholars are researching scalable methods that would allow for analysts to indicate the changing nature of source data and the shape they take (2015).

4.0 Versions in and of the Archive

One of the goals of digital archival work and corpus-building alike is the potential structured data that they provide for such analytics. Flanders (2007) portrays the goal of digital collections to "capture bibliographic codes and textual materiality in ways which can represent them usefully to readers: not simply as visual cues but as data which can give one leverage on the text." Flanders has elsewhere described how Archival encoding projects have:

> ...provided the underlying data standards for an entire new research infrastructure of digital editions, digital archives, digital corpora, digitized special collections, and digital 'projects' whose common element—now more visible than ever in the era of large-scale data—was the way they managed and exposed large numbers of texts. (2014, p. 164)

These collections are modeled to connect items together by links to common data sources, shared systems of annotations, and between multiple

versions of the same work (p. 169). A focus on versionality has been a long-time hallmark of a strand of digital humanities scholarship coming out of textual studies, book historical scholarship, and other work by scholars concerned with the materiality of texts. This influence has afforded digital humanities with an interest in the technologies for curating, disseminating, and circulating text, of which the computer is only the most recent. This research reanimates the embedding of textual genres in processes of composition and production, which might complement work in rhetorical genre studies since the coining of "genre sets" (Devitt, 1991) and "genre systems" (Bazerman, 1994): it situates a work of a certain genre in relation to other works, materials, and roles in the ecology that produces it. Digital humanities has additionally looked to "old-school" bibliography for methods to describe the different versions of a single work over the course of its production, consistent with writing studies' historical attention to processes of composition. A project at my alma mater, for example, has built on the work of Ronald E. Clapper's 1967 dissertation, which transcribes the additions and deletions between manuscript versions of *Walden*, by encoding the different versions in the TEI (Schacht). The Versioning Machine tool that pulls these encoded documents allows its users to compare encoded versions of the manuscripts with one another on the fly, such that they might experience what John Bryant has called a "fluid text moment" for themselves: a moment in which "there is an apparent slippage between your text of a literary work and someone else's," which reminds you that there are multiple editions of texts and also of the labor put into their production (2002, p. 64).

The encoding work that enables this tool, like the qualitative coding familiar to writing studies, is labor dependent on the interests and fastidiousness of the individual encoder. Thus, encoding creates its *own* versions. When encoding is scaled up to include multiple authors and encoders, version information lives in structured metadata; it accompanies the items in digital archives and aspires to transparently and ethically account for editorial intervention that represents choices and may include "errors."[10] Any emendations are captured in encoding to keep them from being "silent alterations" (Flanders, 2007). Martin Paul Eve (2016) argues that this keeping track of versions should inform and provide fodder for further humanistic research: in an article tracking version variation in *Cloud Atlas* throughout its publishing history, Eve cites the claims of textual scholars like Jerome McGann that "trans-textual variance should be considered in the act of interpretation" (p. 27). Eve points to McGann's work on the Rossetti Archive as the "pioneering work" of constructing critical (as opposed to definitive) digital editions that embrace their incongruities (p. 5). Transitioning from the construction of archives to corpus-building, Eve argues

that distant analyses of texts should be as meticulous about versions as the textual encoding initiatives that we acknowledge allow for them (p. 28-9).

The text that would be encoded, of course, is not always composed of the machine-readable text, whether transcribed or 'born-digital,' that would enable computational text analysis. Despite the criticism Eve acknowledges of textual scholarship—to be "complicit in the rise of 'neoliberal' management practices within the academy"—because of its narrow definitions of what constitutes a text, textual studies has indeed had a more inclusive scope (p. 5). These projects do not quite meet the mission of fields like literacy studies in decentering alphabetic text if only because their encodings, at the end of the day, are written in a text-based language. They have, however, attempted to assemble and encode the multimedia compositions of artists long editorialized as authors composing exclusively in text (e.g., The *William Blake Archive*). Further, and more directly relevant to the purposes of writing analytics, digital projects in the tradition of textual studies have considered authors' text-based compositions themselves multimodally. McGann's Rossetti archive includes the graphic design of Rossetti's text, and "not simply its linguistic elements," to be "among [their] primary concerns" of study: McGann counts these among "concurrencies" that include "rhetorical structures, grammatical, metrical, sonic, [and] referential." The project has confronted the challenges of using the textual markup of the TEI to encode the "visibilities of [Rossetti's] expressive media" in machine-readable text. Deciding that the TEI's practices for encoding non-textual phenomena were too limited for their "documentary demands," McGann and his team took up the task of modifying the more generalized SGML (from which XML, on which the TEI is based, is derived) to accommodate the project's needs; McGann narrates their grappling with the affordances and limitations of SGML's structure for encoding the concurrent rhetorical features of Rossetti's pieces (2004).

What, then, to quote poet and critic Susan Howe, is potentially "lost in the typeface?" Howe's discussion of manuscript variation and editorial intervention alike in Emily Dickinson's poetry are a challenge to the "ur-text" status of definitive editions in favor of Dickinson's original manuscripts. Howe is critical of the arrangement by editors of Dickinson's poetry into hymn-like stanzas that erase the variations in form and cadence that she achieves with her line breaks; for Howe, the footnotes that trace editorial interventions center the editor's own constraints and "mask" their authorial role (1987). Howe returns to facsimiles of Dickinson's handwritten manuscripts for a more encompassing interpretation of the "aesthetic function" of her compositions: she argues that the shapes and letters of Dickinson's calligraphy riff on one another and influence her meaning. Howe sees these features as integral to Dickinson's poetics: "[i]n the precinct of poetry," Howe says, "a word, a space around a word, each letter, every mark silence or

sound, volatizes an inner law of form; moves on a rigorous line." This is what Virginia Jackson refers to as "Howe's now famous suggestion that [Emily] Dickinson's manuscripts are themselves 'artistic structures'" (2005, p.37).

The promise of digital projects like the *Dickinson Electronic Archive* is to provide access to images of Dickinson's manuscripts, which might allow for an interpretation that takes into account elements of the drafting process lost in previously edited editions. As Jackson notes, the logic is that the "web images...give a better idea of Dickinson's compositional practices than any previous edition could" (p. 52). In addition to the "handwriting [that] allows more interpretive range than does printing," the manuscript facsimiles provide ostensibly "unmediated and immediate" access to "the visual intentionality of Dickinson's spaces, marks, crosses, dashes, etc." (p. 47-8; Howe, 1986). Howe clocks the moment at which mediation of the private manuscript occurs as the moment when the manuscript is compiled and circulated for public consumption:

> Poetry is never a personal possession. The text was a vision and gesture before it became sign and coded exchange in a political economy of value. At the moment these manuscripts are accepted into the property of our culture, their scholar-author escapes the ritual of framing--symmetrical order and arrangement. (Howe, 1986)

Jackson portrays the romantic access that might circumvent this mediation as an "exposure of [the] private hand to the public gaze" of the type that "has thrilled readers since the nineteenth century" (p. 51). "The new media," Jackson argues, "return the problem of genre in Dickinson to an old division between private and public temporality" that fetishizes the private draft (p. 47). Jackson critiques Howe's vision of an archive that provides access to private materials as a techno-utopian and "progressive narrative of ever greater public access to those papers in the locked box"[11] (p.46). Howe's vision might otherwise call to mind Selfe's early C&W scholarship on access. But while Selfe might recognize that the web is its own mediated venue, "[w]eb publication is cast here" by projects digitizing Dickinson's manuscripts "as the liberation of Dickinson's writing from the policing gaze of the print public sphere." This would-be liberation is from the editorial constraints that have defined Dickinson's compositions as lyric poetry, in the medium of the book, since their initial publication shortly after her death (p. 46). Jackson, however, reminds us that Dickinson's poetry is embedded in genre systems that always-already consider her manuscripts as lyric poetry, consistent with our understanding in writing studies that ostensibly private communication is not private, but public, with an "intertextual memory" (Freadman, 2002, p. 48).

"Most readers," Jackson writes, "have found it impossible to read Dickinson's manuscripts as if they had not already been printed as poems" (p. 20).

The vestiges of editorial choices therefore shape how our digitized collections are modeled. Because we must make choices of scale, the manuscript versions we choose to structure and make public freeze particular moments in the private composition process. In its effort to offer a text that "is not that which the public read," for example, the *Fluid Text Edition* of *Walden* mentioned earlier chooses discrete versions of *Walden*'s manuscript to compare, labeling them Version A through Version G. These correspond with the seven draft versions held at the Huntington Library in San Marino, California, a scheme accepted and normalized by Ron Clapper in the 1960s and still in use by scholars of Thoreau. In making an otherwise invisible work of editorial assemblage visible, the project is transparent about the extent to which the tool that the *Edition* has created inherits a predefined apparatus. Users, in turn, take on the responsibility of deciding what comparison of manuscript versions constitutes a span of alterations meaningful enough to study, highlighting those versions for comparison and representing them among the choices available.

Thus, these vestiges of editorial choices are unwittingly propagated in our analysis if we fail to document them. "New media," to quote Paul Fyfe, "is always in the process of constituting itself as new, erasing the legacies of its entanglements and the continuous work of its propagation" (2016, p. 546). Fyfe's most recent article "concerns the largely invisible corporate histories of digital scholarly resources and the question of how (or even if) we might recover, reconstruct, and interpret them" (p. 551). Lauren Klein places the responsibility for this reparative work on the critic, whose "involvement in the design and implementation—or at least, the selection and application—of digital tools demands an acknowledgement of his or her critical agency" (2013, p. 668). While we have a tendency to think of automated techniques as automatic, even given the difficulty of getting them to do what we want them to do, we must not let a complacency follow writing code that works. Cordell (2017, forthcoming) thinks of the optical character recognition technologies we use to create machine-readable text from digitized manuscripts as compositors, "pressur[ing] the distinction between OCR as an 'automatic' process and composing type as a 'human' process." The immediacy of keyword searches then masks "the underlying data structures on which they rely." Cordell argues that facsimiles are their own versions of digitized materials, and thus cannot serve as surrogates for their originals, because they are mediated by historically-embedded processes that are the products of human authors. The immediacy of computational text analysis similarly masks the degree of transformation it enacts on the text's "original" data structures. Fyfe reframes projects of "data mining" as projects of "data

archaeology" that instead of "extract[ing] meaning" might "recover and reconstitute media objects within their changing ecologies" (p. 551).

5.0 Conclusion: Writing Corpus Analysis

If we are to practice a writing analytics that both draws upon and contributes to both digital humanities and writing studies, it would benefit from the former's ethics of data transformation. We could orient part of our research agendas towards recovering silences in our corpora. Scholars contributing to the formation of writing analytics have already started "captur[ing] writing development in all of its complexity" (Wardle and Roozen, 2012). Rudniy and Elliot (2016) are performing an n-gram analysis of instructor and student comments on their *My Reviewers* platform. Ross et al. (2016) are using topic modeling to analyze peer review feedback. Holcomb and Buell (2016) have prototyped software to observe revision practices in large corpora of student writing. By reframing these corpora as archives, writing analytics might even more responsibly account for how we intervene in the corpora we build. I have assembled below some guiding principles for practices informed by digital humanities that writing analytics might adopt in its future analyses that make private composition practices public.

We could reframe computational analysis as transformations of our corpora. We could additionally pay attention to the scale of our analyses and the degree to which they are non-scalable, since choices of scale may obscure other vantages. We could teach analysis and writing in terms of one another, which would resonate with what we already teach: that to write is to manipulate text according to and in friction with existing grammars.

We could, when possible, publish representative code that built the tools with which we draw our conclusions. If coding is a literacy we value, we must recognize that what we code is an argument that makes certain sets of assumptions that change with context when repurposed, is put out for public critique, and also is the basis for others' learning and repurposing. We could continue to build, support, and use platforms that provide for an open exchange of code in addition to data and written findings.

We could include elements normally outside our purview of analysis. Of particular interest could be the "silent alterations" of a composition by the genre systems in which they are embedded, as well as potential drafting data encoded by writers but erased in their pieces' final published forms. These data could provide for analyses of drafting practices and citation practices across technologies and media. We could look for opportunities to create structured data, when ethically

appropriate, that anticipates analysis when authors compose in otherwise unstructured media.

We could understand that our efforts to recover silent portions of the composition process may re-inscribe other silences or create new silences. We could research ways to document in metadata our own interventions in archival material when preparing them for analysis. Further, we could document how our corpora change with time. We could research methods to cite changes in others' corpora they have documented. This would provide for transparency and the basis for further, ongoing writing analytics.

Author Biography

Gregory J. Palermo is a doctoral student in English at Northeastern University. His research spans writing studies and digital humanities, with a specific focus on disciplinary formation and computational text analysis. He has published in *Genders* on the limitations of the archival record and serves as a Managing Editor of *Digital Humanities Quarterly*. Among his duties, he supports digital projects as part of the Digital Scholarship Group in Northeastern University Libraries.

Acknowledgements

References

About the archive: Editorial principles. *The William Blake archive*. Retrieved from http://www.blakearchive.org/staticpage/archiveataglance?p=editorialprinciple s

Allington, D., Brouillette, S., & Golumbia, D. (2016, May 1). Neoliberal tools (and archives): A political history of digital humanities. *Los Angeles Review of Books*. Retrieved from https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/

Appiah, K. A. (2017). *Cosmopolitanisms*. New York: NYU Press.

Applegarth, R. (2014). *Rhetoric in American anthropology: Gender, genre, and science* (1st edition). Pittsburgh, Pa: University of Pittsburgh Press.

Anderson, C. (2013, May 2). 20 years of Wired: Maker movement. *Wired*. Retrieved from http://www.wired.co.uk/article/maker-movement

Anson, C. M. (2015). Crossing thresholds: What's to know about writing across the curriculum. In L. Adler-Kassner & E. Wardle (Eds.), *Naming what we know: Threshold concepts of writing studies* (pp. 203–219). Boulder, CO: University Press of Colorado. https://doi.org/10.7330/9780874219906.c0013

Aull, L. (2015). *First-year university writing: A corpus-based study with implications for pedagogy* (1st ed. 2015 edition). Basingstoke: Palgrave Macmillan.

Bauman, S. (2011). Interchange vs. interoperability. In *Proceedings of the Balisage: The Markup Conference 2011.* (Vol. 7). Montréal, Canada. https://doi.org/10.4242/BalisageVol7.Bauman01

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 79-101). Bristol: Taylor and Francis.

Bazerman, C., Blakesley, D., Palmquist, M., & Russell, D. (2008). Open access book publishing in writing studies: A case study. *First Monday*, *13*(1). Retrieved from http://uncommonculture.org/ojs/index.php/fm/article/view/2088

Brandt, D. (2014). *The rise of writing: Redefining mass literacy*. Cambridge, United Kingdom: Cambridge University Press.

Bryant, J. (2002). *The fluid text: A theory of revision and editing for book and screen*. Ann Arbor: University of Michigan Press.

Busa, R. (1976). Guest editorial: Why can a computer do so little? *Association for Literary and Linguistic Computing Bulletin*, *4*(1), 1-3.

Carter, S., Jones, J., & Hamcumpai, S. (2015). Beyond territorial disputes: Toward a "disciplined interdisciplinarity" in the digital humanities. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 33–48). Chicago and London: University of Chicago Press.

Charney, D. (1996). Empiricism Is not a four-letter word. *College Composition and Communication*, *47*(4), 567–593.

Chun, W. H. K., Grusin, R., Jagoda, P., & Raley, R. (2016). The dark side of the digital humanities. In M. Gold & L. F. Klein (Eds.), *Debates in the digital*

humanities 2016 (pp. 493–509). Retrieved from
http://dhdebates.gc.cuny.edu/debates/part/14

Clapper, R. E. (1968). *The development of Walden: A genetic text*. ProQuest
Dissertations and Theses.

Condon, W. (2013). Large-scale assessment, locally-developed measures, and
automated scoring of essays: Fishing for red herrings? *Assessing Writing*,
18(1), 100–108.

Cordell, R. (2015, November 5). Scale as deformance. Retrieved from
http://ryancordell.org/research/scale-as-deformance/

Cordell, R. (Forthcoming, 2017). "Q i-jtb the raven": Taking dirty OCR seriously.
Book History, *20*.

Devitt, A. (1991) Intertextuality in tax accounting: Generic, referential, and
functional. In C. Bazerman & J. Paradis (Eds.), *Textual dynamics of the
professions: Historical and contemporary studies of writing in professional
communities* (pp. 335-57). Madison: U of Wisconsin Press.

Drucker, J. (2011a). Humanities approaches to interface theory. *Culture Machine*, *12*.
Retrieved from
http://svr91.edns1.com/~culturem/index.php/cm/article/download/434/462

Drucker, J. (2011b). Humanities approaches to graphical display, *5*(1). Retrieved
from http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Drucker, J. (2013). Performative materiality and theoretical approaches to interface,
7(1). Retrieved from
http://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html

Elliot, N. (2017). *Sociocognitive domains of reader-response theory: Locating
writing analytics*. Presented at the MLA 2017, Philadelphia PA.

Elliot, N., Gere, A. R., Gibson, G., Toth, C., Whithaus, C., & Presswood, A. (2013).
WPA-CompPile Research Bibliographies. Retrieved from
http://comppile.org/wpa/bibliographies/Bib23/AutoWritingEvaluation.pdf

Eve, M. P. (2016). "You have to keep track of your changes": The version variants
and publishing history of David Mitchell's Cloud Atlas. *Open Library of
Humanities*, *2*(2), 1. https://doi.org/10.16995/olh.82

Fitzpatrick, K. (2015, March 11). No tenure: Ends, futures, and transforming the
academy. Retrieved from http://mediacommons.futureofthebook.org/alt-
ac/pieces/no-tenure-ends-futures-and-transforming-academy

Flanders, J. (2005). Detailism, digital texts, and the problem of pedantry. *TEXT Technology*, (2), 41–70.

Flanders, J. (2007, October 31). Electronic textual editing: The women writers project: A digital anthology. Retrieved from http://www.tei-c.org/About/Archive_new/ETE/Preview/flanders.xml

Flanders, J. (2014). Rethinking collections. In K. Bode & P. L. Arthur (Eds.), *Advancing digital humanities* (pp. 163–174). London: Palgrave Macmillan UK.

Freadman, A. (2002). Uptake. In R. M. Coe, L. Lingard, & T. Teslenko (Eds.), *The rhetoric and ideology of genre: Strategies for stability and change*. Cresskill, N.J.: Hampton Pr.

Fyfe, P. (2016). An archaeology of Victorian newspapers. *Victorian Periodicals Review*, *49*(4), 546–577.

Galey, A., & Ruecker, S. (2010). How a prototype argues. *Literary and Linguistic Computing*, *25*(4), 405–424. https://doi.org/10.1093/llc/fqq021

Gibbs, F. W., & Cohen, D. J. (2011). A conversation with data: Prospecting Victorian Words and Ideas. *Victorian Studies*, *54*(1), 69–77. https://doi.org/10.2979/victorianstudies.54.1.69

Giltrow, J. (2002). Meta-genre. In R. M. Coe, L. Lingard, & T. Teslenko (Eds.), *The rhetoric and ideology of genre: Strategies for stability and change* (pp. 187-205). Cresskill, N.J.: Hampton Pr.

Gold, M. (2012, January 7). DH and Comp/Rhet: What we share and what we miss when we share. *The Lapland Chronicles*. Retrieved from http://blog.mkgold.net/2012/01/07/dh-and-comprhet-what-we-share-and-what-we-miss-when-we-share/

Gold, M. (2015, April 17). Facts, patterns, methods, meaning: Public knowledge building in the digital humanities. *The Lapland Chronicles.* Retrieved from http://blog.mkgold.net/2015/04/20/facts-patterns-methods-meaning-public-knowledge-building-in-the-digital-humanities/

Grolemund, G., & Wickham, H. (n.d.). R for data science. Retrieved from http://r4ds.had.co.nz/

Harris, J. (2006). Coming to terms. In *Rewriting: How to do things with texts*. Logan, Utah: University Press of Colorado. Retrieved from http://www.jstor.org/stable/j.ctt4cgnf2

Hawisher, G. (1989). Research and recommendations for computers and composition. In G. Hawisher & C. Selfe (Eds.), *Critical perspectives on computers and composition studies* (pp. 44-69). New York: Columbia UP.

Hockey, S. (2004). The history of humanities computing. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *Companion to digital humanities (Blackwell companions to literature and culture)* Oxford: Blackwell. Retrieved from http://www.digitalhumanities.org/companion/

Holcomb, C., & Buell, D. (2016*). First-year composition as "Big Data": Examining student revisions at scale*. Presented at the 9th International Conference on Educational Data Mining (EDM 2016), Raleigh, NC, USA. Retrieved from https://pdfs.semanticscholar.org/de36/8a95f91b5bef34443d2f2edc0c7e99b0bdb9.pdf

Horner, B., & Lu, M.-Z. (2014). Toward a labor economy of literacy: Academic frictions. In J. Duffy, J. N. Christoph, E. Goldblatt, N. Graff, R. S. Nowacek, & B. Trabold (Eds.), *Literacy, economy, and power: Writing and research after literacy in American lives* (pp. 111–126). Carbondale, IL: Southern Illinois UP.

Howe, S. (1993). These flames and generosities of the heart: Emily Dickinson and the illogic of sumptuary values. In *The birth-mark: Unsettling the wilderness in American literary history* (EPC Edition). Middletown, CT: Wesleyan. Retrieved from http://writing.upenn.edu/library/Howe/

Howe, S. (1986). Some notes on visual intentionality in Emily Dickinson. *HOW(ever)*, 3(4), 11-13. Retrieved from https://www.asu.edu/pipercwcenter/how2journal/archive/print_archive/alertsvol3no4.html#some

Jackson, V. (2005). Dickinson undone. In *Dickinson's misery* (pp. 16–67). Princeton University Press. Retrieved from http://www.jstor.org/stable/j.ctt5hhq3d.7

Kerschbaum, S. L. (2014). *Toward a new rhetoric of difference*. Urbana: National Council of Teachers of English.

Kirschenbaum, M. (2012). Digital humanities as/is a tactical term. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 415–428). Minneapolis: Univ Of Minnesota Press.

Kirschenbaum, M. (2014). What Is "digital humanities," and why are they saying such terrible things about it? *Differences*, 25(1), 46–63. https://doi.org/10.1215/10407391-2419997

Klein, L. F. (2013). The image of absence: Archival silence, data visualization, and James Hemings. *American Literature*, *85*(4), 661–688. https://doi.org/10.1215/00029831-2367310

Knuth, D. E. (1984). Literate programming. *The Computer Journal*, *27*(2), 97–111.

Lang, S., & Baehr, C. (2012). Data mining: A hybrid methodology for complex and dynamic research. *College Composition and Communication, 64*(1), 172-194.

Leroi, A. M. (2015, February 13). Digitizing the humanities. *The New York Times*. Retrieved from http://www.nytimes.com/2015/02/14/opinion/digitizing-the-humanities.html

Lockett, A. (2015, April 3). On the value of "empirical studies" in rhetoric and composition. *Writing Program Administrator ListServ (WPA-L)*. Retrieved from https://lists.asu.edu/cgi-bin/wa?A2=WPA-L;8efce113.1504

Lunsford, A. A. (2015). Writing addresses, invokes, and/or creates audiences. In L. Adler-Kassner & E. Wardle (Eds.), *Naming what we know: Threshold concepts of writing studies* (pp. 20–21). Boulder, CO: University Press of Colorado.

Lynne, P. (2004). *Coming to terms: A theory of writing assessment* (1st edition). Logan, Utah: Utah State University Press.

MacDonald, S. P. (2007). The erasure of language. *College composition and communication*, *58*(4), 585–625.

McGann, J. (2004). Imagining what you don't know: The theoretical goals of the Rossetti Archive. In R.Modiano, L. Searle, & P. L. Shillinsburg (Eds.), *Voice, Text, Hypertext: Emerging Practices in Textual Studies* (pp. 378–97). New York, NY: Palgrave Macmillan.

Milic, L. T. (1966). The next step. *Computers and the Humanities*, *1*(1), 3-6.

Mislevy, R. J. (2016). How developments in psychology and technology challenge validity argumentation. *Journal of Educational Measurement*, *53*(3), 265–292. https://doi.org/10.1111/jedm.12117

Moxley, J. M., & Walkup, K. (2016). *Mapping writing analytics*. Presented at the 9th International Conference on Educational Data Mining (EDM 2016), Raleigh, NC, USA. Retrieved from https://pdfs.semanticscholar.org/de36/8a95f91b5bef34443d2f2edc0c7e99b0b db9.pdf

Nowviskie, B. (2010, October 15). Eternal September of the digital humanities. Retrieved from http://nowviskie.org/2010/eternal-september-of-the-digital-humanities/

Nowviskie, B. (2014). Ludic Algorithms. In K. Kee (Ed.), *Pastplay: Teaching and learning history with technology*. Ann Arbor, MI: University of Michigan Press. Retrieved from http://quod.lib.umich.edu/d/dh/12544152.0001.001/1:5/--pastplay-teaching-and-learning-history-with-technology?g=dculture;rgn=div1;view=fulltext;xc=1

Padilla, T. [thomasgpadilla]. (2017, March 27). currently nerding out about dynamic data citation, living that data life [Tweet]. Retrieved from https://twitter.com/thomasgpadilla/status/846359202968715265.

Pannapacker, W. (2011, January 8). Pannapacker at MLA: Digital humanities triumphant? Retrieved from http://www.chronicle.com/blogs/brainstorm/pannapacker-at-mla-digital-humanities-triumphant/30915

Posner, M. (2016). Here and there: Creating DH community. In M. Gold & L. F. Klein (Eds.), *Debates in the digital humanities 2016* (pp. 265–273). Minneapolis; London: University of Minnesota Press. Retrieved from http://dhdebates.gc.cuny.edu/debates/

Ramsay, S. (2011, January 11). On building. Retrieved from http://stephenramsay.us/text/2011/01/11/on-building/

Ramsay, S. (2014). The hermeneutics of screwing around; or What you do with a million books. In K. Kee (Ed.), *Pastplay: Teaching and Learning History with Technology*. Ann Arbor, MI: University of Michigan Press. Retrieved from http://quod.lib.umich.edu/d/dh/12544152.0001.001/1:5/--pastplay-teaching-and-learning-history-with-technology?g=dculture;rgn=div1;view=fulltext;xc=1

Rawson, K., & Muñoz, T. (2016). Against cleaning. Retrieved from http://www.curatingmenus.org/articles/against-cleaning/

Rauber, A., & Pröll, S. (2015, March 23). Scalable dynamic data citation - RDA-WG-DC Position Paper. Retrieved from https://www.rd-alliance.org/groups/data-citation-wg/wiki/scalable-dynamic-data-citation-rda-wg-dc-position-paper.html

Reid, A. (2010, October 17). The creative community and the digital humanities. Retrieved from http://alex-reid.net/2010/10/the-creative-community-and-the-digital-humanities.html

Reid, A. (2015). Digital humanities now and the possibilities of a speculative digital rhetoric. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 15–19). Chicago: University of Chicago Press.

Rice, J. (2013). Occupying the digital humanities. *College English*, *75*(4), 360–378.

Ridolfo, J., & Hart-Davidson, W. (2015). Introduction. In *Rhetoric and the digital humanities* (pp. 1–12). Chicago: University of Chicago Press.

Risam, R. (2013, September 15). Where have all the DH jobs gone? Retrieved from http://roopikarisam.com/uncategorized/where-have-all-the-dh-jobs-gone/

Roozen, K. (2010). Tracing trajectories of practice: Repurposing in one student's developing disciplinary writing processes. *Written Communication*, *27*(3), 318–354. https://doi.org/10.1177/0741088310373529

Ross, V., Liberman, M., Ngo, L., & LeGrand, R. (2016). *Weighted log-odds-ratio, informative Dirichlet prior method to compare peer review feedback for top and bottom quartile college students in a first-year writing program.* Presented at the 9th International Conference on Educational Data Mining (EDM 2016), Raleigh, NC, USA. Retrieved from http://ceur-ws.org/Vol-1633/ws2-paper4.pdf

Rudniy, A., & Elliot, N. (2016). *Collaborative review in writing analytics: N-gram analysis of instructor and student comments.* Presented at the 9th International Conference on Educational Data Mining (EDM 2016), Raleigh, NC, USA. Retrieved from https://pdfs.semanticscholar.org/cca9/f952af2bde66dedd7ce6c609f5cbc3476e53.pdf

Schacht, P. (Ed.). (2016). Introduction. In H. D. Thoreau, *Walden: A fluid text edition*. Retrieved from https://digitalthoreau.org/fluid-text-toc/

Scheinfeldt, T. [foundhistory]. (2011, November 10). DH arguments are encoded in code. I disagree with the notion that those arguments must be translated / re-encoded in text [Tweet]. Retrieved from https://twitter.com/foundhistory/status/134808062283354112.

Schulz, K. (2011, June 24). The Mechanic Muse - What Is Distant Reading? *The New York Times*. Retrieved from http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html

Selfe, C. L. (1999). *Technology and literacy in the twenty-first century: The importance of paying attention*. Carbondale, IL: Southern Illinois University Press.

Selfe, C. L., & Selfe, R. J. (1994). The politics of the interface: Power and its exercise in electronic contact zones. *College Composition and Communication*, *45*(4), 480–504. https://doi.org/10.2307/358761

Shipka, J. (2011). *Toward a composition made whole*. Pittsburgh, Pa: University of Pittsburgh Press.

Smith, K. G. (2016). A writing studies review of Jim Ridolfo and William Hart-Davidson, editors, Rhetoric and the Digital Humanities. *Digital Humanities Quarterly*, *10*(4). Retrieved from http://www.digitalhumanities.org/dhq/vol/10/1/000226/000226.html

Smith, K. G. (n.d.). (Re)orienting TEI in composition. Retrieved from http://kevingeraldsmith.com/work/reorienting-tei-in-composition/

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

Swarts, J. (2010). Recycled writing: Assembling actor networks From reusable content. *Journal of Business and Technical Communication*, *24*(2), 127–163. https://doi.org/10.1177/1050651909353307

Tsing, A. L. (2012). On nonscalability: The living world is not amenable to precision-nested scales. *Common Knowledge*, *18*(3), 505–524.

Underwood, T. (2017, March 30). Digital humanities as a semi-normal thing. Retrieved from https://tedunderwood.com/2017/03/30/digital-humanities-as-a-semi-normal-thing/

Unsworth, J. (2002, November 8). What is humanities computing and what is not? Retrieved from http://computerphilologie.uni-muenchen.de/jg02/unsworth.html

Vee, A. (2013). Understanding computer programming as a literacy. *Literacy in Composition Studies*, *1*(2), 42–64.

Visconti, A. (2016). A digital humanities what, why, & how. In *Literature Geek*. Retrieved from http://literaturegeek.com/2016/07/21/dlf-digital-humanities-what-why-how

Wardle, E., & Roozen, K. (2012). Addressing the complexity of writing development: Toward an ecological model of assessment. *Assessing Writing*, *17*(2), 106–119. https://doi.org/10.1016/j.asw.2012.01.001

Wickham, H. (2010). A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, *19*(1), 3–28. https://doi.org/10.1198/jcgs.2009.07098

Wilkinson, L. (2005). *The grammar of graphics* (2nd ed.). New York: Springer.

Witmore, M. (2012). Text: a massively addressable object. In Matthew Gold (ed.), *Debates in the Digital Humanities*, 324-327.

---

[1] Matthew Kirschenbaum's "What Is 'Digital Humanities,' and Why Are They Saying Such Terrible Things about It?" might remind us that this ruckus surrounding DH, far from always celebratory, is often quite negative. Kirschenbaum outlines some of the false characterizations, including some that might carry resonance for scholars in writing studies: that "like Johnny, digital humanities can't read," that "digital humanities is managerial," and that "digital humanities is something separate from the rest of the humanities, and—this is the real secret—digital humanities wants it that way" (2012, p. 50).

[2] Kirschenbaum also cites the "tensions" surrounding the field's name—"is it *the* digital humanities or just 'digital humanities,' is it capitalized or not capitalized"—that he holds emblematic of both the definitional debates in the field and, relevant in the context of writing analytics, the field's concern with how a cogent keyword-searchable identity (or lack thereof) might affect research and metrics (2014, p. 51). "The," here, would suggest a plurality of practices across fields rather than a defined field. A lack of capitalization would ostensibly forsake the abbreviation potential for consistency with other field labels. I am electing to use "digital humanities" because I am, for the purposes of this research note, talking about the field as a relatively cohesive community of practice alongside writing studies.

[3] The myth, for example, that there is some superabundance of tenure-track jobs in digital humanities was debunked by Roopika Risam (2013) during the height of the DH rumor-mill's most productive season. Her search of the MLA's Job Information List revealed that the number of literary studies and rhetoric/composition job listings that appeal to digital qualifications is over-exaggerated; and, further, that there are actually fewer standalone listings in digital humanities than any other "subfield" associated with English, including rhetoric and composition.

[4] The task-orientation and contingency of the alt-ac work that has provided an alternative to the faculty job market—work in libraries, on digital projects, in writing centers, and in writing program administration—is a point of conversation in both fields in part because it directly affects us. An embrace of alternatives to the traditional faculty model can be conscious and critical rather than acquiescent. See, for example, Kathleen Fitzpatrick's 2015 post about the possibilities for intellectual autonomy and structural reform that alt-ac employment may offer in a profession with already eroding job security.

[5] In a context perhaps even closer to home for scholars analyzing and assessing writing with quantitative empirical methods, Patricia Lynne has attempted to re-label assessment practices to merely sound less objective. She claims that the use of "terms developed within the objectivist paradigm of educational measurement theory and still carrying positivist baggage" subjugates the

principles of social constructionist composition scholars in favor of "scholars working in a different and incompatible paradigm" of testing associated with big business (2004, p. 6-7).

[6] A discussion of the full epistemological and political implications of the original poster's argument are beyond the scope of this paper. They, however, include that empiricism, as a product of enlightenment thought, denies, nevertheless supports, and replicates systems of power and advantage. Given this argument, the application of empirical methods is then at best an exercise in futility, and at worst an exercise of institutional power that disavows racial and gender-based oppressions.

[7] The Text Encoding Initiative (TEI) is an XML-based markup language that is the de-facto standard in the humanities for text-based encoding to represent texts in digital form. For more information, visit http://www.tei-c.org.

[8] The term is Dale Dougherty's, coined when he started the quarterly DIY journal *Make* magazine in 2005. Maker culture's values of materiality and collaboration are palpable in DH. The movement's facilitation by, and dependence on, increasingly-accessible computational power and mass-scale production are brought to bear in techno-progressive portrayals of the field, as well as in the critiques they inspire. See Anderson (2013) for a popular narrative of the rise of the maker movement.

[9] Aull notes that Gere, as her dissertation advisor, "encouraged" her "early on to see the generative ways that linguistic and rhetorical approaches come together to inform FY writing research and instruction" (p. 10).

[10] Bibliographer Donald Francis MacKenzie (1999), for example, argues that these types of misreadings are historical documents in their own right. Julia Flanders (2007) outlines the Women Writers Project's rationale for preserving "errors" in digital transcriptions regardless of perceived intentionality: "Our premise here is, first, that errors may be significant, whatever their source: they are part of the information that circulated to readers when the text was first published...And secondly,...it may be difficult to say with confidence that a given reading is an error."

[11] Howe reports that the use of Dickinson's manuscripts previously required permission from and fees directed to Harvard College and the Trustees of Amherst College (1987).