The Early History of Digital Humanities

Chris Alen Sula
csula@pratt.edu
School of Information, Pratt Institute
United States of America

Heather Hill
hhill4@pratt.edu
School of Information, Pratt Institute
United States of America

Most commentators locate the origin of digital humanities (DH) in humanities computing of the mid-20th century. Dalbello (2011), for example, begins her account in 1946 with Roberto Busa's plans for the Index Thomisticus, a massive attempt to encode nearly 11 million words of Thomas Aquinas on IBM punch cards. This event (and the narrative that follows) is found throughout the literature, leading some to believe that early DH work "concentrated, perhaps somewhat narrowly, on text analysis (such as classification systems, mark-up, text encoding, and scholarly editing)" (Presner 2010, 6). Others seem convinced that DH is still only text analysis—and misguided in its approach (Fish 2012).

This paper presents an empirical perspective on the early history of digital humanities by tracing publications in two foundational journals (*Computers and the Humanities*, established in 1966, and *Literary and Linguistic Computing*, established in 1986), with particular emphasis on media types and authors' disciplines.

Background

Despite the variety and breadth of definitions of DH (e.g., Gold 2012 and Terras 2012), narratives of its history have been surprisingly homogenous. Hockey (2004) and later authors (Svensson 2009, 2010, 2012; Kirschenbaum 2010; Dalbello 2011) all ground DH in mid-20th century humanities computing, a view that is all but orthodox in short and anecdotal histories of the field. According to this narrative, DH begins in 1946 with the Index Thomisticus and proceeds through advances in corpus linguistics to the founding of the journal *Computers and the Humanities* (*CHum*) in 1966. These early projects are hindered by storage capacity, hardware costs, and processing limits; progress is slow. Though Svensson (2009) admits that not every article during this time is about text analysis, he notes that the field had narrowed enough by 1986 for *Literary and Linguistic Computing* (*LLC*) to supplant *CHum* as the premier humanities computing journal (note the journal titles). Hockey similarly describes the 1970s and 1980s as a period of "consolidation" of text analysis methods. As storage and processing capabilities increased from the late 1970s onward, structured electronic text and multimedia archives dominated the field, followed in the 1990s by Internet-enabled hypertexts, digital libraries, and collaborative editing. The overarching theme of this narrative is text, with the plot revolving around corpora of increasing size and susceptibility to machine analysis.

Though this account dominates historical views of the field, it raises four separate concerns. First, it privileges certain disciplines, projects, and tools at the expense of others (e.g., quantitative history, which is absent from the narrative). Second, it fails to chart an actual historical path from early work in text analysis to "big tent" DH (Jockers & Worthey 2011; Pannapacker 2011a, b), encompassing everything from digital archives and databases to GIS, network analysis, new publishing formats, digital pedagogy, and so on. Third, it precludes historicizing and contextualizing current work that falls outside of text analysis, which may lead to a lack of attention to method, its historical complexities, and points of convergence with related fields such as the social sciences. Finally, these histories all suffer from a lack of evidence; the narrative is assumed and applied rather than documented.

An alternative approach would attend to the various methods, platforms, and tools that animate current DH work and investigate their origins in the literature. Ball (2013), for instance, has drawn attention to longstanding interest in computers and technology within writing studies. Scheinfeldt (2014) has pointed to the historical importance of oral history in the history of DH. Significantly, Nyhan, Flinn, and Welsh's (2013) project, "Hidden Histories," collects and archives oral histories from those who worked in the field during its first decades. These efforts are in broad solidarity with the empirical history presented here and contribute to the growing number of heterodox histories of DH.

Methodology

The corpus for this study consisted of 1,334 research articles published in *Computing and the*

Humanities (1966–2004) and *Literary & Linguistic Computing* (1986–2004). This end date reflects the final issue of *CHum* and predates wide circulation of *A Companion to Digital Humanities* (2003), which was important in shaping DH in many ways. We omitted introductions, reviews, conference reports, and other articles that did not primarily present original research.

We manually inspected each article for media type and applied one of six categories (e.g., text, image, sound, object, number, other). For articles that addressed more than one media type, we recorded 'multimedia'. After inspecting several hundred articles, we added 'technology' as a media type to accommodate articles primarily about technology (e.g., AI, databases, hardware), rather than its application to a particular media. Using all eight codes (see Table 1), we coded or re-coded all the articles.

Category	Example article topics
Text	markup for full-text publishing; database of the works of Pascal; use of software for teaching literature
Image	digitized map of land register of 1822; database of manuscript images
Sound	programs for reproduction of sounds; correcting errors in musical databases; analysis of pronunciation
Object	machine classification of features of archaeological objects; techniques for dating medieval inscriptions
Number	database of wages, money, and prices; report on statistics programs; using Mark IV to extract quantitative data from charters
Multimedia	digitizing *Beowulf* (manuscript images and text); recording live performance; video and speech generation
Technology	artificial intelligence; computer-assisted language learning; mainframe and microcomputer file formats; MS Word 3.0; PF474 string co-processor
Other	report on a center; advancements in publishing; reasoning with natural language

Table 1. Media types coded in this study

In addition to media type, we recorded information about each author's discipline(s) and country of institutional affiliation. In the case of faculty appointed to more than one department, we recorded the discipline as 'multiple,' reasoning that an interdisciplinary appointment is more than a simple conjunction of its constituent departments. For authors located outside of traditional academic departments, we used one of three codings, where appropriate: 'center', 'non-academic', or 'GLAM' (galleries, libraries, archives, and museums). The remaining cases were clustered into one of 21 broad disciplines spanning the humanities and other areas.

Finally, we recorded whether each article had a focus on teaching and learning (e.g., courseware, language learning software).

Data on media type, disciplines, and teaching and learning were visualized using the free software

Tableau Public and are available at http://bit.ly/earlydh.

Findings

The number of articles published each year varies from 6–50 (see Fig 1), the latter owing mainly to a double issue of *CHum* published in 1994/1995, which we recorded as 1995 because of its copyright date. Given the varying number of articles per year, we report several figures below as relative percentages each year (relative to the total number of articles that year). In cases where the two journals are compared, we also report relative percentages (relative to all articles from that journal in the corpus) because there are nearly twice as many total articles from *CHum* as compared to *LLC*, given their years of coverage.
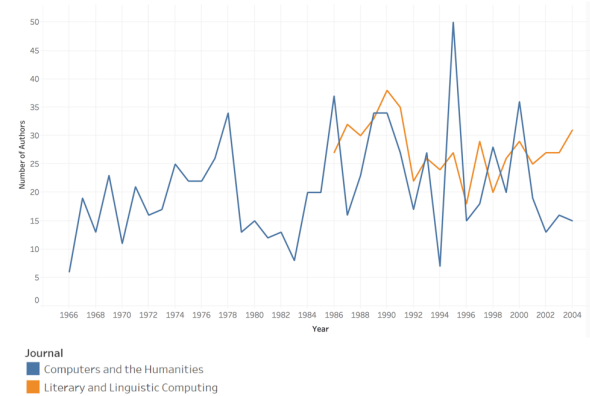

Figure 1. Number of articles published per year

Media type

Text is the most frequently studied medium (59% *CHum*, 72% *LLC*), but sound, multimedia, and reflections on technology are all present in the early literature (see Figs 2–3). These distributions vary by journal, with text being much more prominent in *LLC*. 'Other' is, admittedly, a rather large category at around 4% overall, but the heterogenous articles found there are not easily resolved into one or more media types, which is the focus of these codings, or even a primary theme, such as 'technology.' To some extent, many of these articles speak to the emergence of a field with its own meta-level discussions about theory and the production of knowledge. These articles are found throughout the early literature of DH and increase slightly around the end of the corpus, when "digital humanities" as such might be said to emerge.
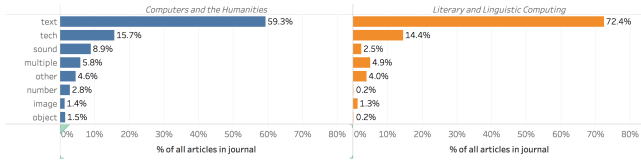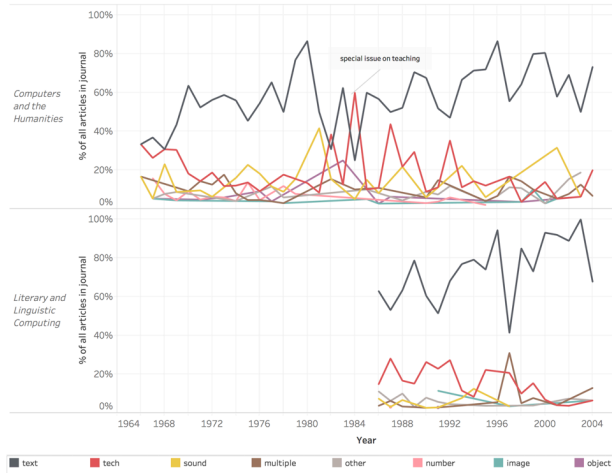
Figure 2. Articles by media type

Figure 3. Articles by media type over time

Disciplinarity

The distribution of authors' disciplines present in each journal is shown in Fig. 4. Computing and computer science is most frequent, largely because of the amount of coauthors from those areas. English language and literature is the most frequent humanities disciplines, commensurate with Kirschenbaum's claim that DH's "professional apparatus…is probably more rooted in English than any other departmental home" (2010, 55). However, authors from languages and literatures departments other than English are nearly as common, as are centers, labs, and non-academic affiliations.
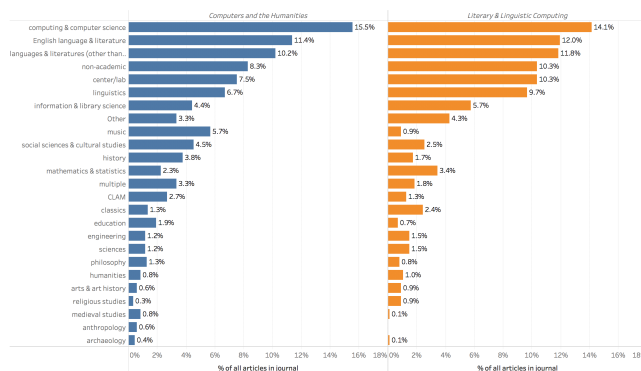

Figure 4. Articles by discipline

Some disciplines work with certain media types more than others (see Fig 5). For example, scholars of languages and literatures work almost exclusively with text, while art historians appear to favor multimedia.
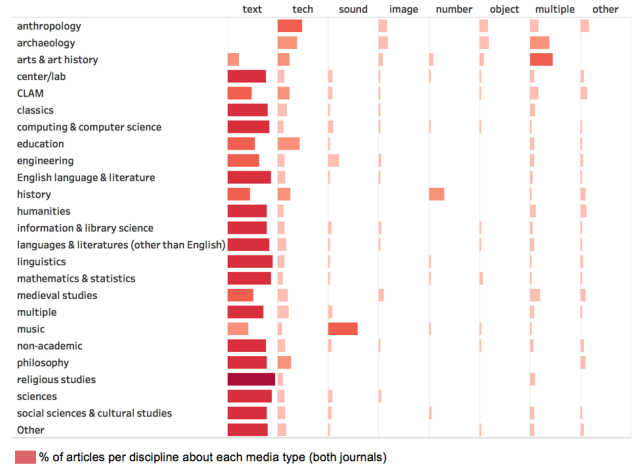

Figure 5. Media type by discipline

Location

Together, *CHum* and *LLC* represent nearly 50 different countries based on authors' institutional affiliations (see Fig 6).
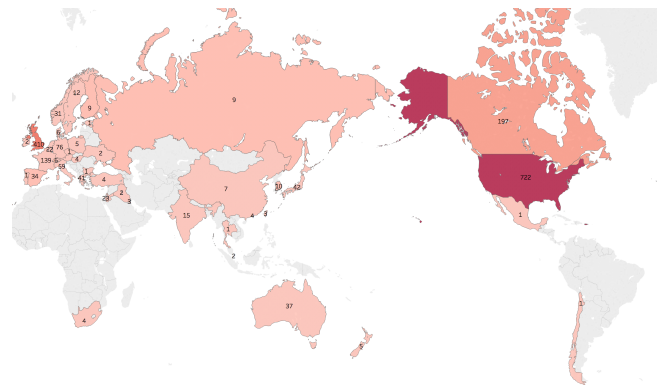

Figure 6. Location of authors

A small but appreciable portion of articles (5.6%, 75 articles) are international (i.e., with co-authors from institutions in different countries). However, the vast majority of authors in *CHum* and *LLC* hail from American and British institutions (respectively), though this predominance declines over the course of both journals (see Fig. 7). This data, as well as the largely Anglophone nature of these journals, presents a limited picture of early DH. A fuller analysis would include work published in other places and languages.
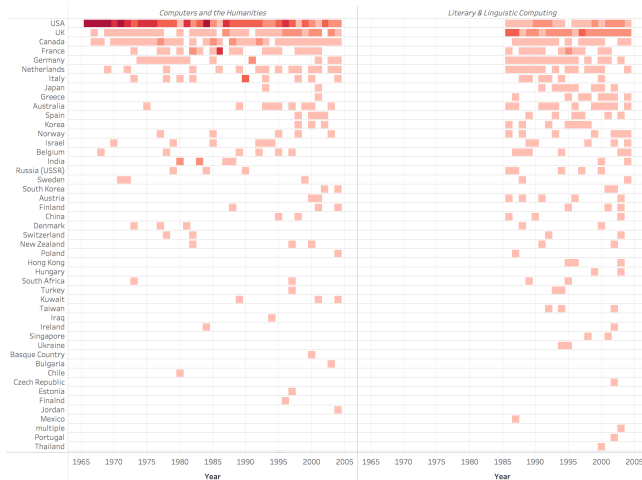
Figure 7. Location of authors over time

Teaching & Learning

There has been longstanding interest in teaching and learning in the field (as shown in Fig. 8), though less so within *LLC*. Peaks in each graph reflect special issues on teaching and learning published by each journal.
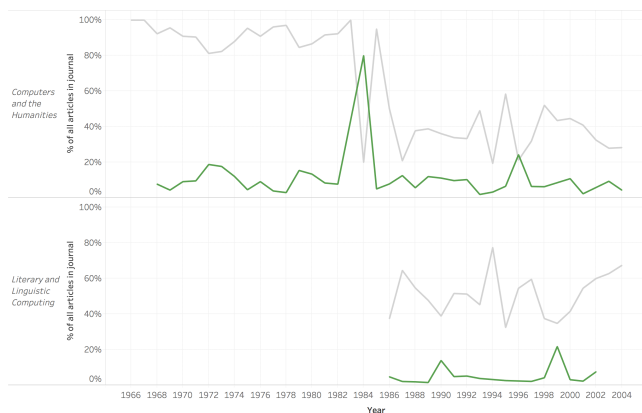

Figure 8. Articles about teaching and learning

Discussion and Future Directions

Rather than focusing on select disciplines, projects, tools, etc., this study includes the full range of early DH work (to the extent it appears in our corpus). The breadth of this picture helps set up the "big tent" view found in current accounts of the field. It also gives ground for historicizing and contextualizing the myriad forms of DH work today. One can imagine exploring this data to discover early DH articles about sound, in classics, from France, etc. and then consulting those primary source articles. Our study does provide some evidence for the claim that early DH work involves text experiments. Significantly, however, it documents the actual extent of that work (59%

CHum, 72% *LLC*), and in so doing, highlights other work in the early history of the field.

Our next steps include exploring additional sources to expand our corpus. In part, this includes investigating disciplinary journals for early DH articles. We might also identify such articles or journals by mining citations in our *CHum*/*LLC* corpus or by consulting sources such as the Companion. There are existing lists of early DH books as a starting point for monographs.

In addition, the full text of our corpus presents several possibilities for analysis, including a citation study that might address questions of transference between disciplines and the degree to which corpus articles cite each other (forming their own scholarly discourse) as compared to literature outside of core DH journals.

Bibliography

Ball, C. (2013). "Digital Publishing in the Tradition of Making Within Writing Studies." *DH 2013.* Lincoln, Nebraska. http://prezi.com/8hrhbrqfw4fs/digital-publishing-in-the-tradition-of-making-within-writing-studies.

Dalbello, M. (2011). "A Genealogy of Digital Humanities." *Journal of Documentation* 67 (3) (April 26): 480–506. doi:10.1108/00220411111124550.

Fish, S. (2012). "Mind Your P's and B's: The Digital Humanities and Interpretation." New York Times, January 23, 2012. http://opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation.

Gold, M. K, ed. (2012). *Debates in the Digital Humanities.* Minneapolis: Univ Of Minnesota Press.

Hockey, S. (2004). "The History of Humanities Computing" in *A Companion to Digital Humanities.* Oxford: Blackwell, 1–19. http://onlinelibrary.wiley.com/doi/10.1002/9780470999875.ch1/summary.

Jockers, M. and Worthey, G. (2011). "Introduction: Welcome to the Big Tent." *DH2011 Conference.* Stanford University. http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-005.xml.

Kirschenbaum, M. G. (2010). "What Is Digital Humanities and What's It Doing in English Departments?" *ADE Bulletin* 150: 55–61.

Nyhan, J., Flinn, A. D., Welsh, A. (2013). "Oral History and the Hidden Histories Project: Towards Histories of

Computing in the Humanities." *Literary and Linguistic Computing.*http://llc.oxfordjournals.org/content/early/2013/07/30/llc.fqt044.short.

Pannapaker, W. (2011a) "'Big Tent Digital Humanities,' a View From the Edge, Part 1," *Chronicle of Higher Education,* 31 June 2011, http://chronicle.com/article/Big-Tent-Digital-Humanities/128434.

Pannapaker, W. (2011b) "'Big Tent Digital Humanities,' a View From the Edge, Part 2," *Chronicle of Higher Education,* 18 Sept 2011, http://chronicle.com/article/Big-Tent-Digital-Humanities-a/129036.

Presner, T. (2010). "Digital Humanities 2.0: A Report on Knowledge." http://cnx.org/content/m34246/1.6.

Scheinfeld, Tom. 2014. "The Dividends of Difference: Recognizing Digital Humanities' Diverse Family Tree/s" Found History. http://www.foundhistory.org/2014/04/07/the-dividends-of-difference-recognizing-digital-humanities-diverse-family-trees.

Svensson, P. (2009). "Humanities Computing as Digital Humanities" *Digital Humanities Quarterly* 3 (3). http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html.

Svensson, P. (2010). "The Landscape of Digital Humanities" *Digital Humanities Quarterly* 4 (1). http://www.digitalhumanities.org/dhq/vol/4/1/000080/000080.html.

Svensson, P. (2012). "Beyond the Big Tent." In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 36–49. Minneapolis: Univ Of Minnesota Press. http://dhdebates.gc.cuny.edu/debates/text/22.

Terras, M., Nyhan, J., and Vanhoutte, E. (2013). *Defining Digital Humanities: A Reader.* Ashgate Publishing, Ltd.