

N. Katherine Hayles

Literature Program

Duke University

How We Think: The Transforming Power of Digital Technologies*

In starting with questions about publishing in the humanities, we may be casting our nets in a ditch while ignoring the ocean. Publishing, whether in crisis or not, cannot be considered in isolation from tectonic shifts within the humanities themselves. As Gary Hall argues in *Digitize This Book! The Politics of New Media, or Why We Need Open Access Now*, “the problem goes much deeper than that, to the extent that the digital model of reproduction raises fundamental questions for what scholarly publishing (and teaching) actually is; in doing so it not only poses a threat to the traditional academic hierarchies, but also tell us something about the practices of academic legitimation, authority, judgment, accreditation, and institution in general” (p. 70). Following a similar line of reasoning, I believe that the Humanities are undergoing a far-reaching transformation that fundamentally challenges the prevailing research paradigms. Just as earlier transformations such as the mid-century New Criticism and the poststructuralism of the 1970’s-1980’s revealed assumptions that had gone largely unnoticed or at least unquestioned, so the new paradigm exposes and subverts many of the assumptions foundational to the Humanities as they are presently configured. For example, the scholarly print monograph remains the gold standard for most of the Humanities, as indicated by the pervasive practice of considering it a prerequisite for tenure. And who produces these monographs? By an overwhelming majority, they are created by a single scholar writing on his or her own. However networked we all are now with ubiquitous email, conference travel on a scale unimaginable fifty years ago, and lively scholarly conversations that frequently span disciplinary boundaries, cultures, and countries, when it comes to writing that

* A significantly different version of this paper will be forthcoming in *Daedalus* in 2009.

all-important book, we tend to sit alone in our studies and think our individual thoughts. Against these normative practices, a new force has appeared on the horizon, tossing old assumptions around as if it were a hurricane headed for New Orleans. It is called the Digital Humanities.

A diverse field, the Digital Humanities include electronic editions and archives of material originally published in print, data-mining and text-mining algorithms that interrogate print corpora that have been digitized, works that are “digital born” such as websites that gather resources, and electronic creative literature including poetry, fiction, locative narratives and immersive art installations, among others. Humanities Computing can be considered a sub-field within Digital Humanities; John Unsworth argues that Humanities Computing at its most powerful not only uses search and pattern-recognizing algorithms but also allows users to modify the algorithms to suit their own purposes. The Digital Humanities has been around at least since the 1960’s, but it was not until the Internet and World Wide Web that it came into its own as an emerging field with its own degree programs, research centers, scholarly journals and books, and a growing body of expert practitioners. Coming to the scene with a background in scientific programming and long-standing interest in machine cognition, I was curious to see how engagements with digital technologies are changing the ways in which Humanities scholars think. The obvious and visible signs of a shift include the changing nature of research, the inclusion of programming code as a necessary linguistic practice, and the increasing number of large Web projects in nearly every Humanities discipline. This much I could learn online, but I was after something deeper and more elusive: how engagement with digital technologies is affecting the assumptions and presuppositions of Humanities scholars, including their visions of themselves as professional practitioners, their relations to the field, and their hopes and fears for the future.

To explore these issues, I conducted a series of phone and in-person dialogues with seventeen scholars at different stages of their careers and varying intensities of involvement with digital technologies. Lacking training in ethnography, I would not be so bold as to label these talks

ethnographies. Rather, they fall somewhere between an interview and a conversations, ranging in length from forty minutes to an hour. I generally kept my views out of the picture, but on occasion I would interject comments when not to do so would have made the conversation seem stilted or unnatural. The insights that my interlocutors expressed in these conversations exceeded my expectations. They shine remarkable light on the ways in which the Digital Humanities are transforming assumptions and posing significant challenges to the Humanities as they now exist. To measure the change, let us first evoke some commonly shared views about the Humanities. Many of us cherish a vision of the ideal Humanities scholar that goes something like this. Our exemplary scholar is someone who has spent years honing his or her sensibility with wide reading in primary texts, deep interrogation of some texts, and immersion in the important scholarly conversations of the time. Our model scholar is not afraid to take risks, drawing on intuition and tacit knowledge as well as encyclopedic knowledge to formulate significant questions and address them with thoughtful, well-reasoned arguments. The scaffolding for his or her work is likely to be the library and the archive, power haunts where our scholar finds the material that he or she will work and re-work into original, perhaps even path-breaking, insights. This figure, which as a teacher guides and inspires and as a professional sets the bar for others, I shall call the Traditional scholar, and the field within which he or she practices, the Traditional Humanities.

By comparison, the scholar working in the Digital Humanities may also have wide experience, but at least part of his or her time is typically spent building and implementing digital tools rather than reading texts. In addition, he or she is likely to spend as much time in a laboratory working as part of a collaborative team as alone in the study. Whereas the Traditional scholar almost always produces a print monograph or essay, the Digital Humanities scholar is apt to have a dual career in print and online, often following both paths at once or sometimes alternating between them. These differences in professional practice are the background from which emerged the major themes of the interviews. Although each interview contributed unique insights, there

were also remarkable similarities. By no means, however, did everyone agree. Indeed, several areas of contestation led to vigorous disagreements. Nevertheless, even these tended to center on a set of common topics. They can be summarized as follows: scale; visualization and multimodality; data streams in contrast to hermeneutic interpretation; collaboration; cumulative research; language and code; and new alliances between Humanities scholars and the general public; and language and code. Taken together, these themes constitute a paradigm shift that leaves almost no area of the Traditional Humanities untouched.

Perhaps the single most important issue is the matter of scale. Gregory Crane (2008a) estimates that the upward bound for the number of texts anyone can read in a lifetime is 25,000 (assuming one reads a book day from age fifteen to eighty-five). By contrast, the digitized texts that can be searched, analyzed, and correlated by machine algorithms numbers in the hundreds of thousands or millions, limited only by ever-increasing processor speed and memory storage. Consequently, machine reading allows questions to be asked that would simply be impossible to implement by hand calculation. Timothy Lenoir, for example, devised algorithms to search patents on Radio Frequency Identification Tags, which are embedded in databases that contain six million five hundred thousand patents. Even when hand searches are theoretically possible, as with the online archive of British literature from 1800-1829 containing 2,272 works of fiction, the number and kinds of queries that one can implement electronically is exponentially greater than would be practical by hand.

To see how scale can change long-established truisms, consider the way in which literary canons typically function within disciplinary practice, for example in a graduate program that asks students to compile reading lists of sixty to a hundred works for the preliminary examination. Most, if not all, of these works are drawn from the same group of texts that populate anthologies, dominate scholarly conversations, and appear on course syllabi,

presumably because these texts are considered to be especially significant, well-written, or interesting in other ways. Almost by definition, then, they are not typical of run-of-the mill literature. Someone who has read only these texts will likely have a distorted sense of how “ordinary” texts differ from canonized works. By contrast, as Gregory Crane observes, machine queries—or what Franco Moretti (2000) calls “distant reading”—enables one to get a sense of the background conventions against which memorable works of literature emerge. Remarkable works endure in part because they complicate, modify, extend and subvert conventions, rising about the mundane works that surrounded them in their original contexts. Scale changes not only the amounts of texts, then, but also the contexts and contents of the questions.

As scale grows exponentially larger, visualization tools become increasingly necessary. Machine queries frequently yield masses of information that are incomprehensible when presented as tables or databases of results. Visualization, drawing on our evolutionary heritage of negotiating complex three-dimensional environments, helps sort the information and make patterns visible. Once the patterns can be discerned, the work of interpretation can begin. Here disagreement among my respondents surfaces. Some argue that the discovery of patterns is sufficient, without the necessity to link them to meaning. Timothy Lenoir’s observation (2008) forcefully articulated this idea: “Forget meaning,” he proclaimed. “Follow the datastreams.” Others, like Stephen Ramsay (2008), argued that the data must lead to meaning for it to be significant. If the Digital Humanities cannot do this, Ramsay declared, “then I want nothing to do with it.” The issue is central because it points to how the Digital Humanities should be articulated with the Traditional Humanities. This articulation has implications for the future: will the Digital Humanities become a separate field whose interests are increasingly remote from the Traditional Humanities, or will it on the contrary become so deeply entwined with questions

of hermeneutic interpretation that no self-respecting Traditional scholar could remain ignorant of its results? If the Digital Humanities were to spin off into an entirely separate field, almost everything would be affected. Obviously, this is a political as well as an intellectual issue, including the possibility of turf battles, competition for funding, disciplinary boundaries, and academic prestige.

Also in play are research practices. Given the predominance of machine queries and the size of projects in the Digital Humanities, collaboration is the rule rather than the exception. Often laboratories are formed in which teams of researchers collectively conceptualize, implement, and disseminate their research. For example, the Humanities Lab at Stanford University, directed by Jeffrey Schnapp (2008), self-consciously models itself on “Big Science,” seeking to implement research paradigms that Schnapp calls “Big Humanities.” Realizing such projects requires diverse skills, including Traditional scholarship but also programming, graphic design, interface engineering, sonic art, and other humanistic, artistic, and technical skills. Almost no one possesses all of these skills, so collaboration becomes a necessity; in addition, the sheer amount of work required makes sole authorship of a large project difficult if not impossible. Unlike older (and increasingly untenable) practices where a Humanities scholar conceives a project and then turns it over to a technical person to implement (usually with a power differential between the two), these collaborations “go deep,” as Tara McPherson comments on the work that has emerged from the online multimodal journal *Vectors*. Conceptualization is intimately tied in with implementation, design decisions often have extensive theoretical consequences, algorithms embody reasoning, and navigation carries interpretive weight, so the humanities scholar, graphic designer and programmer work best when they are in continuous and respectful communication with one another.

As a consequence of requiring a clear infrastructure within which diverse kinds of contributions can be made, “Big Humanities” projects make it possible for students to make meaningful contributions, even as undergraduates. As I write these words, thousands of undergraduates across the country are engaged in writing essays that only their teacher will see (and, if they are lucky, a few other classmates)—essays that will have no life once the course ends and whose sole purpose is the student’s education. As Jeffrey Schnapp (2008) and Gregory Crane (2008a) note, students can complete small parts of a larger project—encoding metadata tags for digital texts, for example—and complete them under faculty supervision, the results of which are incorporated into the project and continue to have a life long after the student has graduated. Mark Amerika has instituted a similar research practice at the University of Colorado, supervising undergraduate contributions to a large database that continues to grow through generations of students, becoming richer and more extensive as time goes on. One of the perennial difficulties of teaching composition has been getting students to envision an audience for their work, a task that often can be accomplished only in fantasy. With Digital Humanities projects, however, there is no need to strive to create an imaginary audience for the student, because he or she is virtually guaranteed an audience when the project goes online.

While scale and collaboration transform the conditions under which research is produced, digital tools work both at the macrolevel of conceptualization and the microlevel of fashioning individual sentences and paragraphs. David Lloyd (2008), a scholar working in Irish literature at the University of Southern California, recounted how he worked with a print essay on Irish mobility in the 19th century to re-envision it for digital publication in the multimodal online journal *Vectors*. Working with the flexible database that Tara McPherson and her co-editor Steven Anderson had devised, Lloyd re-wrote his text, removing all the structures of

coordination and subordination. The fragments were then entered into the database in a recursive process that started with specifying beginning categories and then modifying and changing them as the work proceeded. Lloyd pointed that in cutting out subordination and coordination, something was lost—namely the coherence of his argument and the beautifully crafted prose of his original essay, a loss he felt acutely when he was in the process of cutting it up. But something was gained as well. The effect of the database format, Lloyd said, was to liberate contradictory and refractory threads in the material from the demands of a historically-based argument, where they were necessarily smoothed over in the interest of coherent argumentation. By contrast, the database elements could be combined in many different ways, depending on how a reader wanted to navigate the interface. In collaboration with designer Erik Loyer, Lloyd and Loyer (2006) visualized the topics as potatoes in a field, and the reader navigates by “digging” them. Instead of an argument as such, one potato, two potatoes, dozens or hundreds of potatoes. While some research is no doubt best presented in print, where linear argument reigns, Lloyd was sufficiently impressed with the advantages of a database structure that he wants to use it for future poetic writing.

Another advantage of databases is the ability to craft different kinds of interfaces, depending on what users are likely to find useful or scholars want to convey. Given a sufficiently flexible structure, a large archive can have elements coded into a database for which different scholars then construct multiple interfaces. As Tara McPherson points out, the same repository of data elements can thus serve different purposes to different communities. An example is Kim Christen’s *Mukurtu: Wampurrarni-kari* website on aboriginal artifacts, histories, and images. She provided aboriginal users with a different interface offering more extensive access than the general public sees, giving different functionalities to each. In other instances, teams of

collaborators might work together to create a shared database, with each team creating the interface best suited for its research purposes. Thus each team's efforts are leveraged by the magnitude of the whole, while still preserving the priorities of its own needs and criteria.

Underlying machine queries, database structures, and interface design is a major assumption that characterizes the Digital Humanities as a whole: that human cognition is collaborating with machine cognition to extend its scope, power, and flexibility. The situation requires both partners in the collaboration to structure their communications so as to be legible to the other. For humans, this means writing executable code that ultimately will be translated into a binary system of voltages; for the machine, it means a “tower of languages” mediating between binary code and the diverse kinds of displays the user sees. Multiple implications emerge from this simple fact. If the transition from handwriting to typewriting introduced a tectonic shift in discourse networks, as Friedrich Kittler (1992) has argued, the coupling of human intuition and machine logic leads to specificities quite different in their effects from those mobilized by print. On the human side, the requirement to write executable code means that every command must be explicitly stated in the proper form. One must therefore be very clear about what one wants the machine to do. For Tanya Clement (2008), a graduate student at the University of Maryland working on a digital analysis of Gertrude Stein's *The Making of Americans*, this amounts to an exteriorization of desire, allowing implications to be brought to light, examined and modified in ways that may not happen with print.

The necessity for executable code also creates new requirements for digital literacy. Not every scholar in the Digital Humanities needs to be an expert programmer, but to produce high quality work, they certainly need to know how to talk to those who are expert programmers. The Digital Humanities scholar is apt to think along two parallel tracks at once: what the surface

display should be, and what kinds of executable code will be necessary to bring it about. This puts subtle pressure on the writing process, which in turn also interacts with the coding.

Reminiscent of David Lloyd's excision of structures of coordination and subordination, many writers who move from print to digital publication notice that their writing style changes. In general, the movement seems to be toward smaller blocks of prose, with an eye toward what can be seen on the screen without scrolling down, and toward short conceptual blocks that can be rearranged in different patterns. The effects spill over into print as well. Alexander R. Galloway and Eugene Thacker's *The Exploit: A Theory of Networks* (2007), a print text about digital networks, parses the argument through statements in bold type followed by a short explanatory block of prose, so that the book can be read as a series of major assertions (by skipping the explanations), short forays into various questions (by picking and choosing among blocks), or straight through in traditional print reading fashion.

A major issue that inevitably appears when scholars in the Digital Humanities present their work for tenure and promotion review is evaluation. Aside from questions about collaboration (what part did Professor X contributed to this project?), many digital projects, especially large team collaborations, do not undergo peer review in the usual sense, for example *The William Blake Archive*, *the Rossetti Hypermedia Archive*, and so on. In addition, not all online journals undergo peer review, nor do digital repositories always require that contributors undergo peer review before being allowed to contribute.ⁱ Some journals finesse the problem by sending out submissions for peer review exactly as a print journal would. *Postmodern Culture* does this, and so does *Vectors*, even though Tara McPherson (2008), one of the founding co-editors, thinks that peer review is "broken".ⁱⁱ The problems are not insurmountable. For example, it would be possible to establish a Board of Peer Review for large web projects, in which the applicant would pay a token fee—say, \$100—to have the project reviewed and evaluated, with the overall score to be posted at

the website. The fee would pay for the infrastructure of the Board (staff to correlate applications , arrange for expert reviewers, and present the reports to the Board), and the Board would consist of Humanities scholars in various disciplines, much as university press Boards operate presently.

These procedures, however, may miss the larger picture, which is that peer review itself should be re-thought, along with the institutions of authority that undergird them. As Cathy Davidson comments, “The very concept of peer review needs to be defined and interrogated. We use the term as if it were self-explanatory and unitary, and yet who does and does not count as a peer is complex and part of a subtle and often self-constituting (and circular) system of accrediting and credentialing” (2008, p. 707). Gary Hall, following Samuel Weber’s arguments in *Institution and Interpretation*, argues that disciplines establish their right to legitimacy (including peer review) through the violence of exclusion, by drawing a boundary between them and competing interests. Although Hall does not expound on this point, it is not difficult to think of historical examples: astronomy versus astrology, biology versus botany, modern cosmology versus religion, medicine versus witchcraft, surgeons versus midwives. These boundaries are often reinforced through “science wars” that seek to tar the opposition with labels such as a lack of rigor, a lapse of unreliability, a reliance on posturing rather than scientific practice, etc. The Traditional Humanities have also had their “culture wars.” While it may be unclear who won the “culture wars,” it is very clear who lost them: the Traditional Humanities. The Traditional Humanities have never suffered from lower public esteem than they do presently. In the national press, they are routinely viewed as frivolous, obscure, unimportant, and indulging in completely opaque discourses that no one else can understand. Meanwhile, within its borders, the Traditional Humanities turn to ever-earlier “professionalization” that trades on such words as “rigor,” “specialization,” and “peer review.” These practices frequently legitimate expertise that counts for less and less. According to Gary Hall, estimates are that the essay published in a print journal attracts on the average between three and seven readers (2008, p. 43). And this is not even taking into account other problems

with peer review, such as the tendency to replicate existing paradigms, competition between the reviewer and the aspiring author, bias in the reviewer's views, and so on.

The site in the digital domain where these problems surface with special force is *Wikipedia*. Many professors forbid their students to cite *Wikipedia* as a source, and some even go so far as to forbid their students to consult it (here we may hear an echo of Weber's argument that a discipline defends its borders with the violence of exclusion). Even Larry Sanger, one of the co-founders of *Wikipedia*, finally rebelled against its open content policies and deserted to start *Citizendium*. Although space precludes me from going into detail about these events, there are a number of important points to be made about the legitimating strategies of *Wikipedia*. As James J. Brown (2008) has pointed out, authority within the *Wikipedia* community is gained through citations; if a contribution or edit is contested, the contributor will quickly be challenged to back up his or her claims with citations. The overwhelming majority of these citations come from print sources—that is, from the very print journals and books that professors command their students to cite instead of *Wikipedia*. *Wikipedia* and print scholarship therefore are not so much mutually exclusive territories as they are partners in a trading zone, with each importing some knowledge from the other. Equally important as quality is quantity. By providing a framework to which ordinary people can contribute their knowledge, *Wikipedia* uses “crowd sourcing” to accomplish an eye-popping explosion of available, accessible, and searchable knowledge. *Wikipedia* is now widely acknowledged as the authoritative source for many areas of popular culture, such as rock bands, TV shows, etc., because someone who is obsessed with a particular cultural artifact or performance has contributed an article on it.

Such contributors deserve to be called expert amateur: I want to insist this is not an oxymoron. The engineer who spends his evenings reading about the civil war; the accountant who knows everything about Prussian army uniforms; the programmer who has extensively studied telegraph code books and collected them for years—these people acquire their knowledge not as

professionals practicing in the Humanities but as private citizens with passion for their subjects. For the first time in human history, digital technologies provide a means by which this knowledge can be tapped and made available for research on a global scale. The Digital Humanities therefore have an unprecedented opportunity to build bridges between expert scholars and expert amateurs. An example is the *Hypermedia Berlin* project that Todd Presner (2008) directs, which provides open source software through which community people can contribute their narratives, images, and memories to the growing databank of information about Berlin in various historical periods. The project includes setting up kiosks in neighborhoods where citizens can go to record information about that area from a historical perspective, say of World War II. Presner, along with many international collaborators, has recently started the *Hypercities* project (2008b), in which not only Berlin but cities across the globe can use open source software to create data repositories

This model also lends itself to new practices of evaluation that might be called open review. In contrast to peer review, the open review process invites readers at large to read and comment on a work, and frequently a work in progress, not only a completed manuscript. MacKenzie Wark, for example, put the work-in-progress that would be published in print as *Gamer Theory* up on the Web as GAM3R TH3ORYT, and many of the suggestions and comments he garnered there resulted in revisions, some of them significant. Noah Wardrip-Fruin likewise put up on the Web his book in progress, *Expressive Processing: Digital Fictions, Computer Games, and Software*, for what he calls blog-based peer review. He reasoned that the expert audience here would include not only expert scholars but also the people who played the games (that is, expert amateurs). Both Wark and Wardrip-Fruin also had their books undergo peer review in the usual sense, although in both cases, expert amateurs also helped to shape the final print books.

Along with accessing a huge body of knowledge not normally accessible within the policed boundaries of the academy, open review also provides a constructive common project in which expert scholars and expert amateurs engage together. I recall an occasion when I heard a historian

speak with disdain about “history buffs.” Like history, most professional fields in the Humanities have their shadow fields, for example, people who want to argue that Shakespeare did not write the works for which he is credited. Scholars often regard such activity as a nuisance because it is not concerned with questions that the scholar regards as important or significant. But this need not be the case. Working together within a shared framework of assumptions, expert scholars and expert amateurs can build databases accessible to all and enriched with content beyond what the scholars can contribute. In addition to contributions to scholarship, such projects would create new networks between scholars and amateurs, from which may emerge, on both sides of the disciplinary boundary, a renewed respect for the other. I think this kind of model could significantly improve the standing of the Humanities with the general public.

One of the questions I asked my respondents was what percentages of scholars in the Humanities are presently seriously engaged with digital technologies. Fairly enough, many respondents asked what I meant by “serious.” They pointed out that in a sense, virtually everyone in the Humanities is engaged with digital technologies through email, Google searches, web surfing and so on. But if we take “seriously” to mean engagements that go further into web authoring and the construction of research projects using digital tools, the percentages were generally low, especially if averaged over the humanities as a whole. My own estimate is about 10%. But this figure is misleading, for as everyone agreed, the numbers are generationally skewed, rising quickly within the younger ranks and even more so with graduate students. Many people estimated 40-50% of younger scholars are seriously engaged. This demographic suggests that the Digital Humanities will continue to increase in the coming years, perhaps hitting about 50% when those who are now assistant professors become full professors in ten to fifteen years. The prediction suggests that the scholarly monograph will not continue indefinitely to be the only gold standard, and that Web publishing will not only be commonplace but will attain equal standing with print.

It would be naïve to think that this boundary-breaking trajectory will happen without contestation. Nevertheless, there is strong probability that the future of publishing will increasingly involve the Digital Humanities and electronic publication. The issues discussed here—scale, collaboration, cumulative scholarship, visualization and multimodality, datastreams versus hermeneutic interpretation, database structures and rhetoric, and bridges between expert scholars and expert amateurs—will affect the structures through which knowledge is created, contextualized, stored, accessed and disseminated. These developments offer unprecedented opportunities for rethinking the premises and assumptions that too often go unexamined when someone talks about a “crisis in Humanities publishing.” I do not doubt print publication will continue indefinitely into the future and that many important works will find their most suitable venue in print. What cannot continue, however, is the tacit acceptance of unexamined assumptions about what constitutes “rigor,” “expertise,” “peer review,” and so forth. The Traditional Humanities are in trouble, and the scope of this trouble includes but is by no means limited to publishing. Entangled with this trouble is an opportunity. I am reminded of a line from Tom Stoppard’s *Arcadia*, when one of the characters proclaims, “What a wonderful time to be alive, when everything you thought you knew is wrong!” The Digital Humanities are not so much proving us wrong as changing the terms in which right and wrong are defined, assessed, and evaluated. This ethical dimension makes the Digital Humanities more than a technology, more than a tool. Rather, it represents a paradigm shift that will help to define Humanities scholarship, Traditional and Digital, in the 21st century and beyond.

Endnotes

ⁱ Gary Hall in *Digitize This Book!* (pp. 45-48) goes into detail on the policies of Stevan Harnad's arXiv.org E-Print Archive and Cogprints, a site that archives submissions sent in by authors of scientific pre-prints. Since the contributions are not required to be already published, some of the submissions may not in fact ever see publication in print. Thus peer review operates as a *future expectation* (that is, that it will be carried out in the future and will either approve the submission or negotiate for required revisions). According to Hall, Harnad argues that this expectation is enough to ensure that authors will take peer review into account, a position that seems vulnerable to critique, to put it mildly.

ⁱⁱ Tara McPherson's critique of peer review goes considerably deeper than Lindsay Water's (2004) claim in *Enemies of Promise: Publishing, Perishing, and the Eclipse of Scholarship* that peer review is compromised because departments have ceased to review the work of their colleagues themselves and take the imprimatur of a prestige university press as doing that work for them.