Database, Interface, and Archival Fever

JEROME McGANN

[I]

ED FOLSOM'S PRESENTATION OF WHITMAN'S work as many-faceted and multidimensional is true and important. "[H]is work resists the constraints of single book objects." Indeed. "[T]he entity we call *Leaves of Grass* is actually a group of numerous things. . . ." Just so. These are some of the characteristics not only of Whitman's work but of all imaginative works, which are by their nature multidimensional. Some—like Whitman's works—foreground their multidimensional qualities. Folsom and Ken Price undertook their project because they registered the truth of Whitman's flaunting declaration: "I am large, I contain multitudes."

But then Folsom, happy with the scholarly opportunities made possible by digital technology, goes on to construct a tale (dare I say a narrative?) about the *The Walt Whitman Archive* as an example of "a new genre, the genre of the twenty-first century." This genre is "database," and the Whitman archive is one of its incarnations: the "archive is, in actuality or virtuality, a database."

This statement is seriously misleading—more accurately, it is metaphoric, like Derrida's use of the term *archive* in his well-known book of 1995, *Archive Fever*, which has been so important for the story Folsom is telling. *The Walt Whitman Archive* is not—in any sense that a person meaning to be precise would use—a database at all. What Folsom calls the archive's "rhizomorphous" organization does not emerge from a database structure. It emerges from a core framework consisting of two parts: an inline markup structure (XML) and an XSL-generated interface. Together they allow users to access and—through an X-query-based search engine—manipulate *The Walt Whitman Archive* in the ways that Folsom rightly celebrates.

You will think I am being pedantic, and in a certain respect I am. But accuracy here is important. Folsom's central double theme—that database is a genre displacing book-based narrative genres and that *The Walt Whitman Archive* exhibits this displacement—misrepresents both the archive and the functional character of works of this kind, which are now fairly widespread and will only grow more so. No database can function without a user interface, and in the case of cultural materials the interface is an especially crucial element of these kinds of digital instruments. Interface embeds, implicitly and explicitly, many kinds of hierarchical and narrativized organizations. Indeed, the database—any database—represents an initial critical analysis of the content materials, and while its structure is not narrativized, it is severely constrained and organized. The free play offered to the user of such environments is at least as much a function of interface design as it is of its data structure—whether that structure be a database structure or, as in the case of *The Walt Whitman Archive*, a markup structure.

As humanities scholarship and its inherited archives migrate into their digital conditions and sets of practices, it's crucial to be clear about what is involved and how we want to shape the changes that are under way. I honor Folsom's enthusiasm about our "twenty-first century" opportunities and his adventurous

JEROME McGANN, the John Stewart Bryan University Professor, University of Virginia, is the director of NINES (Networked Infrastructure for Nineteenth-Century Electronic Scholarship [www.nines.org]) and the editor of *The Rossetti Archive* (www.rossettiarchive.org). His two most recent books are the companion volumes *The Scholar's Art: Literature and Scholarship in an Administered World* (U of Chicago P, 2006) and *The Point Is to Change It: Poetry and Criticism in the Continuing Present* (U of Alabama P, 2007).

https://doi.org/10.1632/S0030812900168555 Published online by Cambridge University Press

scholarly spirit in collaborating on the Whitman archive. But Folsom's essay introduces a loose way of thinking about our paper-based inheritance as well as about these new digital technologies, and that looseness endangers the work he has committed himself to.

This looseness does not originate in Folsom, however; its source is Lev Manovich's *The Language of New Media*, often cited by humanists who get excited about digital technology. Folsom extrudes his idea that the database is "the" genre of the twenty-first century from passages like the following:

> After the novel, and subsequently cinema, privileged narrative as the key form of cultural expression of the modern age, the computer age introduces its correlate—the database. Many new media objects do not tell stories; they do not have a beginning or end; in fact, they do not have any development, thematically, formally, or otherwise that would organize their elements into a sequence. Instead, they are collections of individual items, with every item possessing the same significance as any other.

This kind of talk debases our understanding of the matters being discussed, which are far more interesting and complex than such a pronouncement suggests. "Narrative," even "privileged narrative," is as ancient a form of cultural expression as we know. And so far as narrative goes, "the modern age"—presumably, here, the modernist twentieth century—is famous for the inventive ways it fractured and overthrew narrative, especially "privileged narrative." But Manovich needs an easy binary to install the progressivist story that underpins *The Language of New Media*.

For scholars interested in migrating our cultural inheritance to digital environments, databases are by no means the most useful tools for the task—or for the related critical tasks of investigating and rediscovering those materials. The inline markup approach of the Text Encoding Initiative (TEI [www.tei-c.org])—now evolved into XML—became a standard for digitizing literary works for a reason. There are good reasons why *The Walt Whitman Archive* is not a database.

Let's be clear. The TEI and XML do not adequately address the problem of knowledge representation that is the core issue here—that is, how do we design and build digital simulations that meet our needs for studying works like Whitman's?—but they get a lot further along with that task than do database models. They are better because they model some of the key forms of order that are already embedded in textual works like Whitman's. They are better because they understand that works like poems and novels are already marked data.

A deeper problem with Manovich's influential commentary comes from his ideas about the "privileged narrative" order of predigital works like poems and novels. So in place of "grand Narratives of Enlightenment" like, say, *Clarissa* or *Don Juan* or *War and Peace*, we are to imagine a future—a twenty-first century—democratically liberated from their single-minded clutches. Folsom's essay wavers on the question of whether our received literary works are "privileged narratives" requiring fractal redemption, as we see when he writes that "database begins to reveal that it has been with us all along, in the guises of those literary works we have always had trouble assigning to a genre—*Moby-Dick*, 'Song of Myself,' the Bible."

Perhaps there are sheep and goats, and these are examples from the sheepfold. But in this context we want to remember Walter Benjamin's trenchant remark "Every document of civilization is at the same time a document of barbarism." The point is that all our documents are always multiply coded and that scholarship preserves and studies the multiple meanings. If pressed, Folsom would surely agree that anyone could reach back into our cultural inheritance and pluck out, in place of his three examples, three others. For the truth is that imaginative work, as an imitation

https://doi.org/10.1632/S0030812900168555 Published online by Cambridge University Press

of life, is necessarily n-dimensional, protean, shifting: as another poet said, "Changeable too, yet somehow *idem semper*" (Byron 17.11). Is the "democratic beauty" of Whitman's work any more complex or open than the God-haunted and authoritarian Bible or than the savage and aristocratic beauty of the *Iliad*?

I pose that rhetorical question because it exposes a second large problem with Folsom's essay. Drawing on Derrida's representation of books and the archives that house them, Folsom contrasts what he sees as the flexibility of database with the rigidity of museums and libraries. Riffing on Derrida's "archive fever" as an infection spawned by the archive's physicalities, Folsom tells us that

> archives reify the period they record. They contain not only the records of a period but its artifacts as well, their dust the debris of toxins and chemicals and disease that went into making the paper and glue and inks, that went into processing the animal skins that wrap the books we open and, in the dusty light, read and inhale. When we emerge from an archive, we are physically and mentally altered.

Such fulsome prose is partly a Folsom jeu. But Folsom isn't *just* kidding around; this view of an archive as reified knowledge (and database as liberated knowledge) runs as a theme through his essay. Implicit in the idea is a now common but lamentable misunderstanding about libraries, museums, and the works they preserve and transmit. The misunderstanding is especially dismal in this context because we will not design and build effective digital tools and archival repositories—a task we now have clearly before us and that Folsom and Price have themselves embarked on—unless we work from an adequate understanding of our paper-based inheritance.

[II]

In a late lecture, "What's Past Is Prologue," D. F. McKenzie speculated briefly on computerization and textual criticism. His remarks addressed two ways that scholars were using digital tools: for electronic storage of large corpora and for the dynamic modeling of textual materials. McKenzie saw modeling as the more interesting prospect, even if it would "represent a radical departure" from his central "article of bibliographical faith": "the primacy of the physical artifact (and the evidence it bears of its own making)" (259).

McKenzie was a great theorist of the archives in which he spent his radiantly dryasdust life as a scholar. "Rigidity is a quality of our categorical systems . . . ," Folsom tells us, and in celebrating the idea of a transgeneric database he looks to escape those categorical imperatives. But databases and all digital instruments require the most severe kinds of categorical forms. The power of database—of digital instruments in general—rests in its ability to draw sharp, disambiguated distinctions.

Libraries and museums—let's call them archives—also deploy categorical systems and subsystems ("cross-references"). No more than databases do these complex systems exhaust, or define, the multiple possible paths through which we may negotiate and (so to say) narrativize our way(s) through these great towers of Babel. The power of a database is a function of its elementary abstract structure. But therein lie the advantage and the disadvantage of a database compared with an indexing system like a card catalog. The physicality of an archive's categorical system shows a flexibility that a database does not have, because a card catalog is itself an interfaced database.

Moreover, the physicality of the card catalog allows useful interventions in the "rigidity" of the library's categorical substructure. The notations, typed or written, added to hand catalogs graphically demonstrate the historical dimensions licensed by these traditional archival systems. *Leaves of Grass* will have many card entries in the catalog, and each of those cards will not only carry basic metadata, each will carry as well cross-references and the

https://doi.org/10.1632/S0030812900168555 Published online by Cambridge University Press

notations of various archivists. In addition, because even the most well-established notation systems undergo changes over time, the cards and entries bear the evidence of their historical passage and making. Of course, we have to learn to use such instruments, as we have to learn how to design and use databases. But that only brings us back to the basic point: these tools are prosthetic devices, and they function most effectively when they help to release the resources of the human mind—in short, when their interfaces are well designed. Archival-system design must build interfaces that allow user-initiated annotations to enrich the underlying data structure without compromising its formal stability.

In considering how to design and build effective digital systems, we want to think back through the physicality of card catalogs to the materials these catalogs are designed to organize for our use. The dust and toxins and chemicals—every material aspect of "the records of a period [and] its artifacts"—are the minutest surviving particulars of the historical process "that went into making" the preserved work. And from that level we move up to higher levels of historical facticity—for example, to the histories of the depositories and of those who have made and used them. Any system that intends to preserve and organize materials for critical analysis must do everything it can to "save these appearances" (see Barfield), integrate them, and make them accessible for critical study. Databases are useful parts of the digital systems we are moving toward. Like pawns in chess, they are essential elements of the game.

Everyone is impressed—or should be—by the n-dimensionality of literary works, and we are always developing tools, digital or not, to analyze how they work, to help us think about them critically. McKenzie understood, better than most, that the n-dimensionality of a literary work is a function of its historical character and that its historical dimensions are coded in the work's material circumstances. If anything threatens to "reify" the human materials we organize through systems like databases, it is the latter. The threat is avoidable, or can be mitigated, if we think carefully about the character of the materials we are trying to model. A network of devices is needed—not just hypermedia environments, imaging software, markup systems, databases, and searching and data-mining tools but the complex administrative apparatuses that will control, as much as possible, the limitations as well as the capacities of these devices. *Leaves of Grass* is many-splendored because of its complex production and reception histories, because it has been repeatedly mediated and remediated. "It" is more than one thing because people, including Whitman, have continually sought and found different ways to use it and read it.

Toward the end of his essay, Folsom remarks on his "surprising realization" that a "less visible database, the database of users" has been growing along with the archive's core data content. I don't know if this "database of users" is a fact or another figure of speech for *The Walt Whitman Archive*. The last time I looked, the archive had not set up a database to track its users and their types of use, though such a database would be an excellent addition. Because the Whitman archive participates in the Networked Infrastructure for Nineteenth-Century Electronic Scholarship (NINES [www.nines.org])—an online, peer-reviewed aggregation of nineteenth-century British and American scholarship—it belongs to a digital environment designed to integrate users into the intellectual life of a larger system, which necessarily includes the intellectual life of *The Walt Whitman Archive*. NINES materials exist in a distributed network of servers, not a central location, but its design is such that (a) all these materials are aggregated for searching, collection, analysis, and remediation and (b) the individuals using NINES and its materials are formally looped into the system so that their activities can also be searched, collected, analyzed, and remediated.

https://doi.org/10.1632/S0030812900168555 Published online by Cambridge University Press

These critical operations are enabled not by a database or a set of databases but by an open-source toolset, Collex, that represents data as a function of the histories of their use.

Reflecting on digital technology, McKenzie saw that its simulation capacities were forcing him to rethink a "primary article of [his] bibliographical faith," the material self-identity of the archival object. He did not live to undertake an editorial project in digital form. Had he done so, he would have found that his "social text" approach to scholarly work was greatly and practically advanced by the resources of digital technology. He would have seen and embraced these technologies because he understood the dynamic structure of all archives and all their materials.

Editors and scholars engage with works in process. Even if only one textual witness were to survive—say that tomorrow a manuscript of an unrecorded play by Shakespeare were unearthed—that document would be a record of the process of its making and its transmission. Minimal as they might seem, its user logs have not been erased, and they are essential evidence for anyone interested in engaging with the work. We are interested in documentary evidence because it encodes, however cryptically at times, the evidence of the agents who were involved in making and transmitting the document. Folsom is right when he says that "*Leaves of Grass* is actually a group of numerous things. . . ." This is why databases cannot model such complex works. Scholars do not edit or study self-identical texts. They reconstruct a complex documentary record of textual makings and remakings, in which their own scholarly investments directly participate.

Works Cited

Barfield, Owen. *Saving the Appearances: A Study in Idolatry*. New York: Harcourt, 1965.

Byron. *Don Juan*. *The Complete Poetical Works*. Ed. Jerome McGann. Vol. 5. Oxford: Clarendon, 1980–93. 7 vols.

Derrida, Jacques. *Archive Fever: A Freudian Impression*. Trans. Eric Prenowitz. Religion and Postmodernism. Chicago: U of Chicago P, 1996.

Manovich, Lev. *The Language of New Media*. Cambridge: MIT P, 2001.

McGann, Jerome. "Marking Texts in Many Dimensions." *A Companion to Digital Humanities*. Ed. Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell, 2004. 198–217.

McKenzie, D. F. "'What's Past Is Prologue.'" *Making Meaning: "Printers of the Mind" and Other Essays*. Ed. Peter D. McDonald and Michael F. Suarez. Amherst: U of Massachusetts P, 2002. 259–75.

Remediating Whitman

MEREDITH L. McGILL

Ed Folsom's prediction that digital databases will produce an "epic transformation" of archives is based on his firsthand knowledge of the benefits that new-media projects such as *The Walt Whitman Archive* offer to scholars and critics: unprecedented access to rare or inaccessible materials; comprehensiveness—that is, their seemingly infinite capacity to collect scattered texts and commentary, a capacity so much vaster than a book's that

MEREDITH L. McGILL, associate professor of English at Rutgers University, New Brunswick, is the author of *American Literature and the Culture of Reprinting, 1834–1853* (U of Pennsylvania P, 2003) and the editor of *The Traffic in Poems: Nineteenth-Century Poetry and Transatlantic Exchange* (Rutgers UP, 2008). She is working on a study of the circulation of poetry in the antebellum United States.

https://doi.org/10.1632/S0030812900168555 Published online by Cambridge University Press