University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications -- Department of English | English, Department of

7-1-2009

Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?

Kenneth M. Price
University of Nebraska - Lincoln, kprice2@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/englishfacpubs

Part of the English Language and Literature Commons

Price, Kenneth M., "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?" (2009). *Faculty Publications -- Department of English.* Paper 69.
http://digitalcommons.unl.edu/englishfacpubs/69

This Article is brought to you for free and open access by the English, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications -- Department of English by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

DHQ: Digital Humanities Quarterly
Preview
Summer 2009
Volume 3 Number 3

Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?

Kenneth M. Price <kprice_at_unlnotes_dot_unl_dot_edu>, University of Nebraska-Lincoln

Abstract

What are the implications of the terms we use to describe large-scale text-based electronic scholarship, especially undertakings that share some of the ambitions and methods of the traditional multi-volume scholarly edition? And how do the conceptions inherent in these choices of language frame and perhaps limit what we attempt? How do terms such as edition, project, database, archive, and thematic research collection relate to the past, present, and future of textual studies? Kenneth M. Price considers how current terms describing digital scholarship both clarify and obscure our collective enterprise. Price argues that the terms we use have more than expressive importance. The shorthand we invoke when explaining our work to others shapes how we conceive of and also how we position digital scholarship.

What are the implications of the terms we use to describe large-scale text-based electronic scholarship, especially undertakings that share some of the ambitions and methods of the traditional multi-volume scholarly edition? What genre or genres are we now working in? And how do the conceptions inherent in these choices of language frame and perhaps limit what we attempt? How do terms such as *edition*, *project*, *database*, *archive*, and *thematic research collection* relate to the past, present, and future of textual studies? Drawing on a range of resources, including the *Walt Whitman Archive*, I consider how current terms describing digital scholarship both clarify and obscure our collective enterprise. In addition, I will use the final term, *thematic research collection*, to discuss yet-to-be-developed parts of the *Whitman Archive* dealing with place-based cultural analysis and translation studies as a way to illustrate the expansive possibilities of this new model of scholarship. 1

Digital textual studies seem to me inadequately described by the terms now available. *Project* is amorphous; *archive* and *edition* are heavy with associations carried over from print culture; *database* is both too limiting and too misleading in its connotations; and *digital thematic research collection* lacks a memorable ring and pithiness. The terms we use have more than expressive importance. The shorthand we invoke when explaining our work to others shapes how we conceive of and also how we position digital scholarship. We need a new term that is vivid enough to be memorable, elastic enough to cover a class of like things, and yet restrictive enough to allow us to include some scholarly undertakings and not others. Ordinary readers and academics alike rely heavily on the work of editors, yet the standing of editors in the academy has for decades been shaky at best. For many people, electronic work is even more dubious: what relatively short history it has is marked by distrust, denigration, and dismissal. We all know the charges, however distorted they may be: digital work is ephemeral, unvetted, chaotic, and unreliable. When suspicion of the value of editing combines with suspicion of the new medium, we have a hazardous mix brewing. There is a danger that if humanities scholars do not undertake the key work of textual transmission, this work will be done by librarians and systems engineers — that is, it will be done by people with less specialized knowledge of the content. In the fraught circumstances of the academy, driven by a prestige economy, humanities scholars are well advised to be highly self-conscious about what we do and how we describe it. 2

Edition

What do we mean when we use the term *edition*? Even among print editions, there are a number of variations: selected editions, reader's editions, and some boldly claiming to be authoritative or definitive editions. The descriptive word "scholarly" has been applied to numerous approaches: authorial or social, critical or documentary, genetic, eclectic, or best text [Stauffer 2007]. Successful scholarly editions yield a text established on explicitly stated principles by a person or a group with specialized knowledge about textual scholarship and the writer or writers involved. What makes the edition scholarly, of course, is the rigor with which the text is reproduced or altered and the expertise deployed in the offering of suitable introductions, notes, and textual apparatus. 3

For those of us who work on prominent figures who have received previous treatment, our own textual work intervenes in an ongoing editorial tradition. A fundamental and often vexingly difficult question is, what should go in an edition? Like most digital editing endeavors, the *Walt Whitman Archive* must proceed with an awareness of the print past — in our case, especially of two significant attempts to present Whitman in scholarly editions: *The Complete Writings of Walt Whitman* (G. P. Putnam's Sons, 1902) and *The Collected Writings of Walt Whitman* (New York University Press, Peter Lang, and the University of Iowa Press, 1961-2004). This awareness produces competing impulses: we want to benefit from and respond to past work, but we also want to avoid constraints on thought and action that were a result of print-based limitations. As editors, we acknowledge the ways of knowing that are enabled by our predecessors — they are the cultural history we inherit — but our job is also to extend their efforts and to produce new ways of knowing that are responsive to cultural, critical, and technological changes (as well as the discovery of documents and the development of new biographical insights) that have happened in the interim. 4

The language of the *Walt Whitman Archive*'s first grant application to the National Endowment for the Humanities (NEH), drafted in 1999, shows how we were thinking of our digital work as in dialogue with the print past. We wrote, 5

> Our goal has been to build upon the strengths of the *Collected Writings* edition, most volumes of which were supported by grants from the National Endowment for the Humanities. The amount of Whitman's work is so huge that no two scholars could hope to edit it effectively in a lifetime — fourteen scholars spent the better parts of their careers editing the materials that now make up the *Collected Writings*. But we do believe that developments in electronic scholarship have made it possible to enhance and supplement the *Collected Writings* by editing the materials that have not yet been included (and adding the materials that have come to light since the *Collected Writings* volumes were issued) and by digitizing and encoding the *Collected Writings* so that these disparate volumes — which often arrange material in confusing and contradictory ways — can function seamlessly and so that Whitman's materials can be presented effectively in any number of new configurations: by genre, by date, by keyword, by subject. The electronic environment can also allow us to make available not just printed transcriptions of Whitman's manuscripts, letters, and books, but to deliver actual facsimile images of the original documents. [Folsom and Price 1999]

It would be fair to acknowledge that Ed Folsom, co-director of the *Whitman Archive*, and I have had evolving views of the relationship between our undertaking and its most recent print predecessor, the *Collected Writings of Walt Whitman*. Our gradually shifting views have been shaped in part by discussions with publishers. At various times, we considered entering into agreements with two publishers — Primary Source Media and the University of Virginia Press — and in fact reached late stages of contract negotiations with each of them. Initially, we reasoned that if a publisher could secure the permissions for us to use the copyrighted material in the twenty-two volumes of the *Collected Writings* published by New York University Press, a significant amount of work, some of it meticulously done, could be preserved and extended.[1] Of course this line of thinking raised a key issue: if a new publisher had to pay for the permissions, the site, or some significant part of it, would need to be commercial in order to recover these and other costs, and perhaps make a profit as well. We were not absolute purists committed always to building a completely free site. In fact, there were extended periods when we were convinced that such an approach would not be possible for a poet like Whitman who left so much debris everywhere. We thought that editing such chaos would demand the combined resources and know-how of the scholarly, library and archival, and publishing communities. Gay Wilson Allen, a general editor of the *Collected Writings*, commented about editing Whitman, "Sometimes his exhausted editors almost wish that he had had two or three good house fires, and considering the houses he lived in, it is also astonishing that he did not" [Allen 1963, 8]. 6

For us, then, a key question emerged: would we conceive of the *Whitman Archive* primarily as being the remediation of the *Collected Writings*? We recognized that our relationship to the *Collected Writings* is problematic: we have a half-century of valuable editorial work collected there, but the limits of a print format make this edition a trial to use. The *Collected Writings* has been the standard edition, the edition cited by American literary scholarship over the past few decades, but much of the work needs to be done again and the presentation re-conceptualized. We struggled to come to terms with a giant from the print past. And yet this monumental edition was both enormous and characterized by some inexplicable omissions, most notably Whitman's revelatory poetry manuscripts. As our initial grant application pointed out, 7

> [W]e have Whitman's laundry lists in print; we have the business cards of his sidewalk repairman in print, but we don't have the manuscripts of "Song of Myself" in print... His poetry manuscripts and periodical publications reveal, among other surprising things, a Whitman who devoted extraordinary time and care to the creation of a poetry that appeared to be quick and spontaneous; his manuscripts expose an artist whose casual, loafing persona was in fact the result of intensive and obsessive artistic labor. [Folsom and Price 1999]

In retrospect, it is clear that we have responded to the *Collected Writings* not by "digitizing and encoding" it but by prioritizing work on material not included there: photographs, bibliography, full texts of various editions of *Leaves of Grass*, archival guides to manuscripts, transcriptions of manuscripts, contemporary reviews of Whitman's writings, and 8

so on. If we were the first editors of Whitman, this order of development for an online resource would have been peculiar. Certainly some of Whitman's prose, *Democratic Vistas* or *Specimen Days*, for example, or his correspondence might rank ahead of some of these items in most people's sequencing list. But of course we do work within a historical context, and what has seemed most pressing (and perhaps most fundable) have been those things altogether neglected or poorly treated by the *Collected Writings*.

Sometimes we learn to be thankful for our failures, and I am certainly grateful now that our negotiations with publishers always went bust. I think — because of a recent NEH challenge grant to be discussed later — that the *Whitman Archive* is in an unusual position: we now have a team of people and the resources in place so that, with reasonable luck, we ought to be able to achieve a more expansive *Whitman Archive* than the already quite extensive site, and to keep it freely available. There are of course examples of other large, not to say gargantuan, free sites. But we should not underestimate the challenges attendant on making vast amounts of material freely available since "free" means no cost to the end user, not the creators. 9

It is reasonable to wonder why Whitman needs to be edited if there have been two previous scholarly editions. And it is reasonable to acknowledge, in response, motivations that have nothing to do with the electronic medium specifically. Editorial work is one way to engage in historical criticism and to help bring the past into the present so it may live in the future. Although the shelf life of a scholarly edition far exceeds that of a monograph, scholarly editions begun half a century ago for Whitman in one case, or a century ago in the other, now seem inadequate. Their approaches require rethinking, not to mention the need to add material and convey new discoveries. Editions of modern writers are almost always selective. Still, a selection ought to include the most important items. If asked to pick Whitman's most important single text, many would name the first publication of *Leaves of Grass* (1855). Here Whitman was at his boldest and most experimental, and the book has elicited some memorable reactions over the past 150-plus years: Ralph Waldo Emerson found it to be "the most extraordinary piece of wit and wisdom that America has yet contributed" [Emerson 1938-1994, 446]. William Carlos Williams called the first *Leaves* "a book as important as we are likely to see in the next thousand years" (Williams, quoted in Hindus 1955, 3) . Clearly, this is a highly significant book. And we might expect the 1855 *Leaves* to be the highlight of an edition of Whitman's writings. Strangely enough, neither *The Collected Writings of Walt Whitman* nor the earlier *Complete Writings of Walt Whitman* bothered to include it. 10

How do we explain this omission? To a large extent, this odd result stems from twentieth-century editorial practices for establishing authoritative or definitive texts that encouraged the selection of a single text. The economics of print publishing — combined with the dominant editorial theories of the mid twentieth-century — made the so-called deathbed edition of *Leaves of Grass* the one most commonly featured in various commercial and scholarly editions. That final authorized printing of Whitman's book is in fact presented twice in the New York University Press edition: it serves as the basis of both the Comprehensive Reader's Edition and the *Leaves of Grass* Variorum. The deathbed edition is remarkable, but it could not be described as Whitman's most daring, most experimental, or even most coherent volume. 11

Print editions of Whitman tended to falter when dealing with multiplicity, whether of versions or of authorship. Whitman is well known as the writer who couldn't stop writing, revising, and reissuing *Leaves of Grass* (a book that appeared in six radically distinct American editions in his lifetime). Less well known is Whitman's involvement in collaborative enterprises. In fact, when we think of the great collaborators in literary history, Whitman hardly jumps to mind. Instead, we remember that Whitman was so self-reliant that for the first edition he more or less did everything: wrote the poetry, designed the book, set some of the type, distributed the book, and anonymously reviewed it. He appears to be dead set against even the largely invisible and ordinarily neglected forms of social authorship, a poet acting out the role of the solitary singer made famous in "Out of the Cradle Endlessly Rocking." Yet this poem also dramatizes collaboration, with one set of voices prompting another, bird song and human song, a single trill and the thousand responsive chords from a thousand different singers to follow. 12

Arguably the medium of print itself encouraged earlier editors to take a restricted view that often remained blind to the social aspects of textual production. It is easier, frankly, to exclude contributions made by book designers, copyeditors, typesetters, and others. Yet if we think longer and harder about Whitman's own career, the extent of his collaboration — almost entirely ignored by *The Collected Writings of Walt Whitman* — is striking. Whitman collaborated with typesetters, designers, and proofreaders, as he readily acknowledged,[2] and also in his journalism, both as editor and writer; in his extensive though anonymous contributions to the early Whitman biographies by R. M. Bucke and John Burroughs; in heretofore uncollected interviews (now being edited by Brett Barney); in his extensive conversations with Horace Traubel — a 5,000 page trove of information. In fact, his correspondence itself is fundamentally a collaborative undertaking involving (ordinarily) two-way engagements, though the strong authorial bias of the *Collected Writings* is clear in their featuring of just Whitman's outgoing correspondence. 13

Project

Project is a bigger, baggier term than edition and is far less specific in what it suggests about the type of work being undertaken. *Project* can describe everything from fixing a broken window on the back of a house to the Human Genome Project. In a literary context, *editions* and other results tend to emerge out of *projects*, but what constitutes the *project* is also the entirety of the undertaking: space, personnel, atmosphere, and the totality of all efforts. An *edition* might result 14

from a project, without being the *project*, which includes all of the work conducted and records produced. The *Whitman Archive*, when regarded as a project, encompasses the compiled email discussion list that fitfully records the building of the *Archive* and the thinking that has gone into it. The documentation of a project, in our case, includes the behind-the-scenes Works-in-Progress page, with its assortment of information, including grant proposals, minutes from Whitman planning meetings over the years, a manuscript tracking database, an image warehouse, and project-related humor.

Project is not a favored word in every context. When I sent drafts of a "We the People" challenge grant application to NEH program officers, I was struck by how forcefully they discouraged me from the using the word *project*, at least in the context of that competition. Their reasoning was that challenge grants were intended to fund permanent entities, unlike a *project* which they conceived of as having a finite temporal life. For me, "Whitman Project" and "*Whitman Archive*" were more or less interchangeable terms. I had to make a real effort to purge the document of all references to *project*. It was a neutral term to me: *project* was so natural as to be almost invisible in the drafts and certainly did not raise a red flag. 15

This story raises a larger issue: what happens when an undertaking becomes not just rhetorically but practically open-ended, when it has the good fortune or obligation to be an ongoing concern? We were successful with our challenge grant application, and we are now well along in building a $2 million permanent endowment for the *Whitman Archive*. Thanks to this remarkable turn of events, the *Whitman Archive* can now plan on an ongoing annual budget comparable to what one might expect annually from a major two- or three-year grant from a federal agency or foundation. And, remarkably, in this case, there is no end date to that support. 16

For the 2007 Digital Humanities conference at the University of Illinois, Urbana-Champaign, Matt Kirschenbaum coordinated a panel called "Done. Finished Projects in the Digital Humanities." He asked, "How do we decide when we're done? What does it mean to finish something? How does the 'open ended nature of the medium' (a phrase we all pay lip service to) jibe with the reality of funding, deadlines, and deliverables? What can we learn from finished projects, both successful and unsuccessful? For that matter, how do we define success and failure? Are 'we' the ones who ought to be defining it? If not, who?" These are good questions, and at the *Whitman Archive* we find ourselves concerned with them even as we face different considerations as well. What happens when work plans realistically *could* continue over generations? What is the best way to plan for that type of future?[3] A theoretical possibility of digital scholarship — the indefinite expansibility — has become a lived reality in our case. We are only now absorbing the meaning of this grant, but one implication is that it provides us with the license, perhaps even the charge, to be as bold and ambitious as our talents and energies allow. 17

Database

How adequate is the term *database* for describing the type of large scale electronic projects we have been considering? Throughout this essay, I have used the *Walt Whitman Archive* as a testing point and illustrative example. To discuss the *Whitman Archive* in terms of database is especially timely now because *PMLA* recently featured an article about the *Walt Whitman Archive* by Ed Folsom, "Database as Genre: The Epic Transformation of Archives," and included a handful of responses (along with Folsom's response to the responses). The ensuing discussion made clear that people understand the term *database* in a variety of ways and attach different connotations to the word. These differences arise mainly from a distinction between 1) a strict definition of *database* — as a technical term in an electronic context database refers primarily to a collection of structured data that is managed by a database management system, most commonly based on a relational model; and 2) a looser use of *database* that employs the term on a more metaphorical level. 18

As the *PMLA* discussion of the *Whitman Archive* indicates, *database* can be a suggestive metaphor because it points to the re-configurable quality of our material (and that of similar sites). The term also conveys simultaneously "finished" and "unfinished" qualities; while a *project* can be logically thought of as "done" or "not yet done," we usually conceive of a database as usable as soon as it begins to exist, and we take as a given that the data will continue to proliferate, potentially indefinitely. The *Whitman Archive* resembles a database in that its content is discrete computer files that function atomistically: as functional units within a computing system each item is just as important as every other item. 19

If the *Walt Whitman Archive* resembles a database (without meeting the specifications of a technical or a literal definition), so, too, does Whitman's own process of composition. As Folsom notes, "Whitman formed entire lines as they would eventually appear in print, but then he treated each line like a separate data entry, a unit available to him for endless reordering, as if his lines of poetry were portable and interchangeable, could be shuffled and almost randomly scattered to create different but remarkably similar poems" [Folsom 2007, 1574-75]. At times, it almost seems as if Whitman were anticipating Raymond Queneau's *Cent Mille Milliards de Poèmes* [One Hundred Thousand Billion Poems], a fascinating book in which the pages are cut horizontally so that each verse in each sonnet of the collection can be turned separately and all combinations of choices are poetically grammatical. (Queneau estimated that a reader would have to spend two hundred million years, working twenty-four hours a day, to read every combination.[4]) Whitman's own cutting and pasting of lines, and his rearranging of poems to make other poems is not this extreme — nor is it as extreme as Samuel Beckett's experiments in *Lessness*[5] — though there is some resemblance to both. Finally, though, what may appear random ordering in Whitman is best understood as restless experimentation, a 20

combinatory and recombinatory poetics, guided by Whitman's recurrent drive to improve the effectiveness of his poems. Here, for those willing to use the term *database* metaphorically and to recognize non-electronic forms of databases, we can think of database as a key tool for Whitman himself: his storehouse of poetic lines, in both manuscript and print, was his working database for future compositions, one that he had always only partial access to because of the scattering of his documents but that nonetheless served as a means of composition.

If we turn to more literal uses of the word database and think about the *Whitman Archive*, we see that it is a complex composite structure that includes numerous databases and XML files. Folsom's description of the *Whitman Archive* as "a huge database" is illuminating when taken metaphorically, though it is less helpful when taken literally, because the entirety of the *Whitman Archive* is not a single database any more than it is, as Jerome McGann asserts, merely XML files plus XSLT. In fact, the *Walt Whitman Archive* is comprised of numerous databases (some public and some not) along with many XML files including TEI, EAD, and XHTML files.[6] McGann goes on to claim that the XML and XSLT work together to "allow users to access and—through an X-query-based search engine—manipulate *The Walt Whitman Archive* in the ways that Folsom rightly celebrates" [McGann 2007, 1588]. Ironically, though, in the course of denying the applicability of *database* as a term suitable to the *Whitman Archive*, McGann overlooks that our search engine is entirely dependent on translating the XML files into database form. At a more general level, McGann is perceptive in noting that any database represents an initial interpretation of the material. A database is not an undifferentiated sea of information out of which structure emerges. Argument is always there from the beginning in how those constructing a database choose to categorize information — the initial understanding of the materials governs how more fine-grained views will appear because of the way the objects of attention are shaped by divisions and subdivisions within the database. The process of database creation is not neutral, nor should it be. 21

Archives and Digital Thematic Research Collections

Having discussed *edition*, *project*, and *database* separately, I now turn to consider the final two terms together, *archive* and *digital thematic research collection*. In the past, an archive has referred to a collection of material objects rather than digital surrogates. This type of archive may be described in finding aids but its materials are rarely edited and annotated as a whole. In a digital environment, *archive* has gradually come to mean a purposeful collection of surrogates. As we know, meanings change over time, and *archive* in a digital context has come to suggest something that blends features of editing and archiving. To meld features of both — to have the care of treatment and annotation of an edition and the inclusiveness of an archive — is one of the tendencies of recent work in electronic editing. One such project, the *William Blake Archive*, was awarded a prize from the Modern Language Association recently as a distinguished scholarly edition.[7] 22

Digital archives are often notable for their depth and breadth of coverage of whatever the stated thematic interest is. Such scope has not been common in editing. Indeed it is possible to see a tension in the very term *collected edition* because collecting and winnowing are two very different activities. Thomas Wentworth Higginson, in a review of the *Complete Writings of Walt Whitman*, might have been commenting on the *Whitman Archive* when he wrote, "[T]he present editors do not shrink from inserting not only the details of every change, but even the unprinted variations which have hitherto existed in manuscript only" [Higginson 1903, 400]. Of course, the more inclusive an edition becomes the more it may be dominated by the surviving "discarded" writings, especially for writers who kept many documents [Folsom 1982, 374]. Some feel that we do violence to the wishes of writers when we make their second-rate material available to the public, while others celebrate what they believe is made possible by inclusive editions: a new, deepened, and enriched sense of the artist's process of composition, preoccupations, and achievements. Ultimately, the whole question of what is in keeping with the wishes of a writer is beside the point. We do not edit for writers themselves but for our own purposes as scholars and readers. 23

Peter Shillingsburg expresses skepticism about the advantage of the archival approach: 24

> The computer makes possible, we are told, the juxtaposition of all the relevant texts in their linguistic and bibliographic variant forms. Thus a library of electronic texts, linked to explanations and parallels and histories, becomes accessible to a richly endowed posterity. To the extent that such archives contain accurate transcriptions, high resolution reproductions, precise and reliable guides to the provenance and significance of their contents, and the extent to which they are comprehensive, to that extent they are "definitive" — until the next generation of critics and scholars with new interests notices some other aspect of texts that scholarly editors of the past (by then that will be us) took for granted and ignored. But already, information overload has set in. The comprehensiveness of the electronic archive threatens to create a salt, estranging sea of information, separating the archive user from insights into the critical significance of textual histories. [Shillingsburg 2006, 165]

Shillingsburg focuses on the limits of a form still being developed as opposed to the potential of that form. Nothing in the archive form intrinsically requires it to be "estranging" or alienating, of course. An electronic archive can be as welcoming as fresh water and as rewarding as the wit of its creators can make it. Having a lot of information is not

inherently more estranging than having less information. Nothing guarantees the effectiveness of selective treatment accompanied by "textual histories," and nothing guarantees effectiveness of more comprehensive treatment accompanied by textual histories. In each case, everything depends on the quality of the editorial work. Digital and print scholarship are equally embedded in history, and both share a vulnerability to aging.

Another term that is more or less synonymous with electronic *archive* is *digital thematic research collection*.[8] Some prefer this term because it may avoid some of the misleading connotations of *archive* — ordinarily people assume that materials in a traditional print-based archive are unedited.[9] Carole Palmer writes about *thematic research collections*, 25

> Collections of all kinds can be open-ended, in that they have the potential to grow and change depending on commitment of resources from collectors. Most thematic collections are not static. Scholars add to and improve the content, and work on any given collection could continue over generations. Moreover, individual items in a collection can also evolve because of the inherent flexibility (and vulnerability) of "born digital" and transcribed documents. The dynamic nature of collections raises critical questions about how they will be maintained and preserved as they evolve over time. [Palmer 2004, 351]

Archive is a self-designated term, one adopted by the creators of resources. In contrast, *digital thematic research collection* is a term used by people describing the work created.

Thematic research collection may be the most accurate term for what many of us are attempting, but it has not gained currency because it is neither pithy nor memorable. Carole L. Palmer notes that a *digital thematic research collection* is the closest thing to the laboratory that we have in the humanities — the place where necessary research materials are amassed. I have argued elsewhere that in a "digital context, the 'edition' is only a piece of the 'archive', and, in contrast to print, 'editions', 'resources', and 'tools' can be interdependent rather than independent" [Price 2007, 435]. 26

Does collecting — the emphasis in Palmer's description — qualify as research, as a scholarly genre? A digital thematic research collection possesses the virtues of a traditional scholarly edition while containing much more. We may nonetheless wonder about how helpful the term *digital thematic research collection* is to the uninitiated. Nothing in the term indicates editorial rigor and nothing points to the value added by scholarly introductions, annotations, and textual histories. The only thing that seems to separate it from a mass digitization project is the "thematic" element. However, one can imagine a mass digitization project that is thematic and that lacks editorial supervision and intervention in the reader's experience of the text. Can we find a better term that indicates this difference? Does *digital thematic research collection* communicate its meaning adequately? 27

If literary scholars who are assembling electronic texts are becoming fundamentally or solely "literary-encoders" and "literary-librarians," then, despite my own recognition of the inseparability of interpretation and encoding, I fear for the standing of their work when judged by faculty in humanities departments (Schreibman, as quoted in [Palmer 2004, 352]). Without care and forceful practical examples and theoretical essays, the same prejudices and misunderstanding that drove editing and bibliography from the center to the periphery of literary studies will continue to prevail. We also need descriptions of digital thematic research collections that highlight the editorial work and other types of scholarly value that are added to the raw materials populating the collection. In many circles, editing — whether it is print-based or electronic — is regarded as pre-critical work. Some editorially related tasks are fairly routine and do not require scholarly expertise (the same is true of critical work as well). And yet others clearly do, and we need to find ways to clarify how historical knowledge, theoretical sophistication, and analytical strengths are necessary to the creation of a sound text or texts and accompanying scholarly apparatus in a successful edition. 28

Some components of a digital thematic research collection or archive may stretch ordinary understandings of *edition*. Many thematic research collections or archives aim toward the ideal of being all-inclusive resources for the study of given topics. A good thematic research collection might begin with an edition conceived in inclusive terms. Digital thematic research collections go far beyond traditional editions in their presentation of many types of materials. They are often even more "organic" than print editions (despite their technological aspects) — that is, they grow, evolve over time, based very much on immediate circumstances. For the *Walt Whitman Archive*, new work on the Civil War is now underway because an expert on Abraham Lincoln at the University of Nebraska-Lincoln, Kenneth J. Winkle, and I perceived a scholarly need and are interested in collaborating on this undertaking. New work on translation — I will say more about both of these new endeavors later — developed because Matt Cohen, already associated with the *Whitman Archive*, was interested. Being published online but being simultaneously a work-in-progress allows for a flexibility in the *Whitman Archive* that print editions could never have. New scholars with new ideas may emerge at any time, creating new and unexpected additions to our work. 29

I mentioned earlier that the theoretical possibilities of digital scholarship might oblige us to boldness — the present moment, when electronic scholarship is still nascent and the boundaries are still capable of being moved, provides a mandate to innovate and expand possibilities. Ideally, a digital thematic research collection would also allow for the study of cultural contexts. In the case of Whitman, we might want to study him as a city poet. He once said that *Leaves of Grass* "arose out of my life in Brooklyn and New York from 1838 to 1853, absorbing a million people, for fifteen years 30

with an intimacy, an eagerness, an abandon, probably never equaled" (quoted in Reynolds 1995, 83) . A life-long city-dweller, his work also emerged out of New Orleans, Washington DC, and Philadelphia/Camden, New Jersey. We would like for the site to enable and to promote interpretations of place-based writing that were not possible before. It would be useful to be able to study all of these areas with dynamic maps containing detail down to the block level. Period maps exist for Washington, DC, New York, Brooklyn, Philadelphia, and New Orleans. New discoveries will emerge once we can ask different questions because of having a great deal more information from census records, maps, health records, police reports, possibly even information on sexual subcultures, and so on.

I have recently begun work on a digital undertaking that may or may not become part of the *Whitman Archive*. Whether the project ultimately is folded into the *Archive* or remains a separate, stand-alone collection, it certainly grew out of my work on the *Archive*. We might think about it as budding off of an existing digital thematic research collection and taking on a life of its own. The project "Civil War Washington: Studies in Transformation" draws on the methods of many fields — literary studies, history, geography, computer-aided mapping — to create an experimental digital resource. The President and the poet both experienced the War from vantage points in the nation's capital, Lincoln striving to reunite the divided nation and Whitman caring for tens of thousands of wounded soldiers. Their activities and perspectives chronicle the War and provide insights into the large and complex forces that transformed Washington from a sleepy Southern town to the symbolic center of the Union and nation. 31

We are gathering uncollected factual data about an urban space that served as the center both of the Union's War effort and of a divided nation, where hospitals arose overnight, wounded men moved in and out, "contraband camps" of fugitive slaves developed, and temporary shelters were erected to house the city's swelling population, which tripled during the four years of the War. Washington was a noisy city during these years: the noise in the city was of construction as work on the Capitol continued; the noise just outside the city was of destruction as the Confederate army worked to tear it down. Even as bridges were defended and a ring of forts made this space the most heavily defended city on earth, Washington fostered vibrant life. 32

"Civil War Washington: Studies in Transformation" will situate Lincoln and Whitman in the midst of a rich field of geo-spatial and temporal data. At the heart of the project will be richly layered, interactive maps plotting both geographic and temporal data that clarify the transformation of Washington, DC. The maps and underlying databases will make it possible to analyze change over time as structures grew and the population swelled and developed a new ethnic and racial mix. We will make possible multifaceted and dynamic studies of Lincoln's and Whitman's activities during the War years, based on textual and statistical evidence and using the power of maps and graphs to illustrate historical change. Lincoln's and Whitman's routes can be plotted on a daily and sometimes hourly basis. We believe that by providing a rich backdrop of census, health, and hospital records; theater schedules; horsecar routes; and other factual data, we will make possible a better understanding of Lincoln's and Whitman's lives and their roles in the transformation of the nation and its capital. 33

Another extension of the *Whitman Archive* now being undertaken serves to expand trans-linguistic, cross-cultural understandings. Whitman scholarship offers rich opportunities because *Leaves of Grass* has been translated into every major language. One of the *Archive*'s objectives is to present editions of Whitman's work key to literary, cultural, and historical study of the poet and his work's effects. Thus Matt Cohen has taken the lead in tackling a digital edition of the first extensive translation of his work into Spanish. Álvaro Armando Vasseur's 1912 selection from *Leaves of Grass* is the work of a Uruguyan poet who translated Whitman not directly from English but via an earlier Italian translation. This fascinating text tells us a lot about the circulation of culture. Making a version of *Leaves* available to the Hispanaphone world seems fitting given current trends in U.S. demographics and in light of the many calls to internationalize American studies. 34

We supplement the translation with a critical introduction and a sample back-translation into English in order to give those unable to read Spanish an opportunity to see how the text was altered in the process of translation. For example, consider the following lines as given by Whitman: 35

> The disdain and calmness of martyrs,
> The mother of old, condemn'd for a witch, burnt with dry wood, her children gazing on,
> The hounded slave that flags in the race, leans by the fence, blowing, cover'd with sweat,

And here is how Vasseur rendered these lines as revealed in a literal back translation: 36

> The mother of old condemned as a witch and burned over dry firewood, before her children's eyes,
> The slave, persecuted like an imprisoned woman, who falls mid-flight, all atremble and sweating blood.

Vasseur's direct comparison of the slave to a woman presumably is based on their common lack of power, but it also creates some cross-gendered possibilities that turn the passage in new ways. Whitman had distinct units — separate lines — for the witch and the hounded slave. An association could be made between them because of their juxtaposition, but that association is hardly insisted on in the English original. Vasseur turns the suggestion of a link into an unmistakable link. Now racial slavery has become associated with the irrationality of the inquisition and serves to remind the reader of the widespread support of slavery by the church in the U.S. (and in South America). While this 37

reading is only barely available in Whitman's original, in Vasseur's translation it appears on the surface. This passage clarifies that translating a text is interpreting it in another language. To ignore such interpretations is to ignore an enormous part of Whitman's reception in the world.

We have either in progress or the planning stages work on Whitman and other languages (German, Russian, Ukrainian, Portuguese, and Chinese). This will begin to better place him in a world context rather than situtating him solely in Anglophone culture. The work will provide valuable texts to further Whitman studies and through associated commentary reflect the social, historical, and linguistic milieus of the nations in which the translations were done, thereby once again stretching the bounds of what a digital thematic research collection originally envisioned within much narrower parameters can do. These possibilities, the ever-emerging questions and new directions, go far beyond the ordinary edition in the pre-digital age. 38

As I have indicated, we do not have an adequate term to describe the digital scholarly work now underway in numerous projects. What is it that we want our descriptive word to capture: is it the physical thing? Digital sites, contrary to popular (and sometimes scholarly) opinion, are physical things after all — they take up space, can be created and destroyed, and so on. Is it the nature of the content? If so, we need a word that suggests what can be an infinitely extensible resource. Or should we emphasize, primarily, the way we make the thing, the collective that has come together in order to do work on a new scale in humanistic study? 39

Importantly, we should not strive to fit our work to one or another existing term but instead expect that, in time, terms will alter in meaning — or new ones will come into existence — so as to convey the characteristics of a new type of scholarship. I strongly agree with Peter Shillingsburg that a new term is needed, though I am not enthusiastic about his proposed term: *knowledge site*. (So many places and institutions could justifiably be called *knowledge sites* that the term seems unlikely to become identified with a particular genre of electronic scholarship.) I propose instead a not-immediately-intuitive but perhaps ultimately more promising alternative: *arsenal*.[10] The online etymological dictionary helps explain the appeal of the term: 40

> *arsenal*
>
> 1506, "dockyard," from It. arzenale, from Ar. dar as-sina'ah "house of manufacture, workshop," from sina'ah "art, craft, skill," from sana'a "he made." Applied by the Venetians to a large wharf in their city, which was the earliest meaning in Eng. Sense of public place for making or storing weapons and ammunition is from 1579.

I like the emphasis on workshop since these projects are so often simultaneously products and in process. I also like the stress on craft and skill, a reminder that editing is not copyist work. The "public place for making" suits current aspects of the genre under discussion and will no doubt characterize it even more in this age of social networking. The dockyard connotations of *arsenal* are helpful in suggesting a kind of inclusiveness about all the vessels, sloops, ketches, and yawls that can hook up to it. (The wharf and dockyard are places of multilingual exchange.) The obvious objection to the term *arsenal* is that it seems militaristic in current usage. Yet we should recall that *magazine* once primarily meant a storehouse for weapons and ammunition. If the primary meaning of *magazine* can shift from being a storehouse of weapons to a storehouse of mixed content for periodical publication, who knows what could happen with *arsenal*?[11] We are, for better or worse, always entangled with force and power: the Internet itself has its origins in the military. Perhaps one step toward turning swords into plowshares is to seize a word like *arsenal* and make it our own. Can we imagine a world in which what is emphasized is not the created thing so much as the group of people who are now joined together for a common purpose?

Notes

[1]After New York University Press published twenty-two volumes of *The Collected Writings of Walt Whitman*, the publishing house of Peter Lang published two additional volumes of Whitman's journalism, and the University of Iowa Press published a single supplemental volume of Whitman's correspondence.

[2]After the publication of the 1881-1882 *Leaves of Grass*, Whitman remarked, "All this is not only my obligation to Henry Clark, but in some sort to all proof-readers everywhere, as sort of a tribute to a class of men, seldom mentioned, but to whom all the hundreds of writers, and all the millions of readers, are unspeakably indebted. More than one literary reputation, if not made is certainly saved by no less a person than a good proof-reader. The public that sees these neat and consecutive, fair-printed books on the centre-tables, little knows the mass of chaos, bad spelling and grammar, frightful (corrected) excesses or balks, and frequent masses of illegibility and tautology of which they have been extricated" [Whitman 1978, 256]

[3]Issues involving long-term preservation come to mind, of course. A simple curation is not viable. That is, we cannot hand over to a library the present-day *Whitman Archive* and expect people fifty years from now to find its interface and technical underpinnings particularly easy to use. This is in sharp contrast to a book published fifty years ago and

deposited in a library. For digital scholarship, we cannot foresee how maintenance, updates, and migration will work in the future.

[4]"C'est somme toute une sorte de machine à fabriquer des poèmes, mais en nombre limité; il est vrai que ce nombre, quoique limité, fournit de la lecture pour près de deux cents millions d'années (en lisant vingt-quatre heures sure vingt-quatre)" [Queneau 1961, n. p.]

[5]Beckett's short story *Lessness* was first published in French as *Sans*. In his enigmatic story Beckett experimented with random ordering of sentences in the making of fiction.

[6]In a "Reply" to those who commented on his essay, Folsom observed that the *Whitman Archive* is in fact "several databases."

[7]It should be noted that my view of *archive* differs here from that of some commentators. Peter Shillingsburg, for example, remarks that "the level of critical intervention is miniscule in the electronic archive" [Shillingsburg 2006, 156]

[8]In a series of talks in the 1990s John Unsworth and Daniel Pitti began applying the term *thematic research collection* to the type of scholarship under discussion in this paper.

[9]Recent work in archival theory by Heather MacNeill, Elizabeth Yakel, and Michelle Light and Tom Hyry, emphasizes the non-neutral nature of archives themselves and urges the adoption of language such as "archival representation" to highlight the mediating role archivists as they order, interpret, and develop information architectures within socially constructed practice.

[10]Henry Wadsworth Longfellow famously explored the possibility transforming an arsenal into pipe organs of love. See his poem "The Arsenal at Springfield."

[11]I am less concerned that *arsenal* catches on than I am that we recognize the fresh features of new work underway and that we are self-conscious about what we want any new term to convey.

Works Cited

Allen 1963 Allen, Gay Wilson. "Editing the Writings of Walt Whitman: A Million Dollar Project without a Million Dollars." In *Arts and Sciences* 2 (Winter 1963): 8.

Beckett 1969 Beckett, Samuel. *Sans*. Paris : Les Éditions de minuit, 1969.

Beckett 1970 Beckett, Samuel. *Lessness*. London, Calder & Boyars, 1970.

Emerson 1938-1994 Emerson, Ralph Waldo. *The Letters of Ralph Waldo Emerson*. New York: Columbia University Press, 1938-1994.

Folsom 1982 Folsom, Ed. "The Whitman Project: A Review Essay." In *Philological Quarterly* 61 (Fall 1982): 369-394.

Folsom 2007 Folsom, Ed. "Database as Genre: The Epic Transformation of Archives." In *PMLA* 122 (October 2007): 1571-79.

Folsom and Price 1999 Folsom, Ed, and Kenneth M. Price. "The Walt Whitman Archive." Grant Proposal to the National Endowment for the Humanities, 1999.

Higginson 1903 Higginson, Thomas Wentworth. Review of *The Complete Writings of Walt Whitman*. In *The Nation* 76 (May 14, 1903): 400-401.

Hindus 1955 Hindus, Milton. "The Centenary of Leaves of Grass." In *Leaves of Grass: One Hundred Years After*. Palo Alto: Stanford University Press, 1955, 3-21.

Light and Hyry 2002 Light, Michelle and Tom Hyry. "Colophons and Annotations: New Directions for the Finding Aid." *The American Archivist* 65 (Fall / Winter 2002): 216-30.

MacNeil 2005 MacNeil, Heather. "Picking Our Text: Archival Description, Authenticity, and the Archivist as Editor." *The American Archivist*. 68 (Fall / Winter 2005): 264-78.

McGann 2007 McGann, Jerome. "Database, Interface, and Archival Fever." In *PMLA* 122 (October 2007): 1588-92.

Palmer 2004 Palmer, Carole L. "Thematic Research Collections." In *A Companion to Digital Humanities*. Oxford: Blackwell Publishing, 2004, 348-65.

Price 2007 Price, Kenneth M. "Electronic Scholarly Editions." In *A Companion to Digital Literary Studies*. Oxford: Blackwell Publishing, 2007, 434-50.

Queneau 1961 Queneau, Raymond. *Cent mille milliards de poèmes*. [Paris]: Gallimard, 1961.

Reynolds 1995 Reynolds, David S. *Walt Whitman's America: A Cultural Biography*. New York: Knopf, 1995.

Shillingsburg 2006 Shillingsburg, Peter. *From Gutenberg to Google: Electronic Representations of Literary Texts*.

Stauffer 2007 Stauffer, Andrew. "Digital End of the Scholarly Edition." Paper delivered at the Society for Textual Scholarship meeting, New York, March 2007.

Whitman 1978 Whitman, Walt. *Daybooks and Notebooks*. 3 vols. Ed. William White. New York: New York University Press, 1978.

Yakel 2003 Yakel, Elizabeth. "Archival Representation." *Archival Science* 3 (2003): 1-25.