ABSTRACT

Title of Document: DIGITAL (IN)HUMANITIES: RE-READING DIGITAL ARCHIVES AS A FORM OF CULTURAL EXPRESSION

Aaron P. Dinin, Master of Arts, 2009

Directed By: Professor, Martha Nell Smith, Department of English

A 2007 PMLA article discussing the *Walt Whitman Archive* juxtaposed narrative and database as competing forms of cultural expression. This article incited a flurry of responses which continued to use the database and narrative comparison. Dinin, in his article "Digital (In)Humanities," reassesses the terms of the digital archive debate, arguing that the terms "narrative" and "database" are both constricting and misleading. The juxtaposition shouldn't be database versus narrative to see which one becomes the dominant form of cultural expression because narrative, he argues, is a form of database. The more proper juxtaposition, as presented by the paper, is one that places "digital archive" alongside "narrative" because both are products of database and both are forms of cultural expression. Dinin, in his article, then goes on to explore the potential of digital archives as a form of cultural expression.

Section 1: Extracting the Human from Digital Humanities

Ancient Greek mythology tells the story of the Cumaean Sybil. According to the legend, the Sybil at Cumae was a woman who once asked that Apollo grant her immortality; however, she neglected to ask for immortal youth. Apollo, being of the notoriously playful Greek god type, granted the woman's wish for immortality without giving her immortal youth. As the years passed, the Sybil's body shrank until she was no larger than the small jar in which she lived, and she became a recluse hiding in a mountainside cave.

Apart from shrinking her, immortality had another effect – having lived in the world for many centuries, the Sybil at Cumae learned to recognize the patterns of life, and as a result, she gained a reputation as a sort of prophet. People from all parts of Greece would travel to the Sybil's cave with questions about their futures and seeking answers to important decisions. They would leave their questions outside the Sybil's cave and retreat to the woods – she did not like to be seen – where they would await her response.

When the Sybil had made her prediction – a prediction based on past experiences – she would scrawl her answer on palm leaves, writing one word per leaf and aligning them in front of her cave in the appropriate order. Only when the leaves were in place and the Sybil had retreated to her cave was it appropriate for the questioner to approach. But before he could reach the leaves, a gust of wind would inevitably arise, blowing the leaves from their intended order. The questioner would be able to collect all the leaves, but he was forced to guess at the Sybil's original proclamation. So desperate for an answer to a question he recognized as vitally important to his future, the asker would

presume a word order that made the most sense to him, follow that advice, and often it would lead to despair, downfall, and ruin.

Destitute and disenchanted, the questioner would return to the Sybil's cave and call upon her to explain why she had given him such poor advice. Certain she had not been wrong in her prediction, the Sybil would reproduce the palm leaves and ask him to order them in the message he thought she had supplied. After seeing the message, the Sybil would discover a few errant word orderings, rearrange the leaves into the original prophecy, and had the advice seeker abided by those words he would have obtained success, wealth, riches, and whatever other happiness he had originally desired.

I begin my discussion of digital literary archives with this ancient tale because, at the present moment, many digital humanities pioneers are in the position of questioners. So determined are they to immediately divine the answer of how digital media might re-imagine their discipline, many have taken all the pieces and hastily arranged them into an answer. While following such answers will not likely lead to any ultimate ruin, as the desperate obeying of the Cumaean Sybil's errantly ordered words often did, it can, and already has framed the debate in limiting terms.

One such controversial answer currently limiting the discussion of digital humanities and digital archives appears in *PMLA*. Extolling the virtues of the *Whitman Archive*, one of its co-editors, Ed Folsom, provocatively wonders "if narrative itself is under threat" (1576). Folsom's musing stems from a realization about the expansiveness of database. As a result of new technologies (like the ability to present all of Walt Whitman's manuscripts, via database, in a central, easily accessed location) the "details of the database quickly [exceed] any narrative we might try to frame the data with"

(1576). For Folsom, database's expansive possibilities threaten to make narrative at the very least unwieldy, and at its most destructive, obsolete. To paraphrase his concern, technological advances stand to remove what Jerome McGann, in a later response to Folsom, describes as being "as ancient a form of cultural expression as we know" (1589). Folsom's answer is, in a literary tradition sense, apocalyptic – database, electronic archives, digital media, and their similar technologies are going to supplant mankind's most ancient cultural expression.

This "doomsday for narrative" approach does not extend solely from Folsom. The editor of the *Whitman Archive* bases his concerns on the argument of Lev Manovich, who, in *The Language of New Media*, proclaims, "Database and narrative are natural enemies. Competing for the same territory of human culture, each claims an exclusive right to make meaning out of the world" (225). But Manovich's proclamation is founded on an errant principle. Narrative is not the only tool of cultural expression. What of the lyric? What of music? What of paintings and sculptures and any number of other artistic and professional disciplines? By framing the digital debate as a contest between database and narrative, Manovich and Folsom neglect that narrative already "competes" (to appropriate their terminology) with other forms of cultural expression. If database is, indeed, a new form of cultural expression, why would it dislodge narrative and no other? Just as narrative coexists with other forms of cultural expression, database should, if it is indeed a form of cultural expression, also coexist.

The metaphor of "natural enemies" Folsom adopts in his article to describe the relationship between database and narrative is Folsom rearranging his palm leaves until he comes to an answer that sounds plausible. In a lively debate, five of Folsom's peers

bristle at his suggestion and refute him through the course of five brief article responses. However, these responses are equally as limiting, framing themselves in the same database and narrative terminologies. As a result, instead of defending his acceptance of the database/narrative enemies metaphor, Folsom, in a final response to the discussion incited by his original article, accepts an entirely different metaphor equally as limiting. Folsom writes:

> To describe the relation between narrative and database, N. Katherine Hayles offers an astute alternative to Lev Manovich's "natural enemies" metaphor: she suggests "natural symbionts," a metaphor I plan to appropriate and use from now on.
>
> (1608)

Folsom's quick change of metaphors speaks to the difficulty of the answer to the broad question: "How will new media affect humanities studies?" It speaks to the intense desire to piece together the palm leaves in any way that might make sense even if it does not make truth.

The more responsible answer to the question of how new media will affect humanities studies is to admit technology's variable influence. While in some sectors, the digital age might ignite a complete overhaul of analytical practices and processes, in other sectors new media's influence might scarcely leave a trace. In addition, while some people, like Folsom, might actively and passionately engage with technology, others might have little use for it. Digital technologies are passive devices – computers do not actively analyze and theorize and codify; humans do. Digital technology is only one tool of many in the analytical arsenal.

Acknowledging the digital age's products (i.e computers, the Internet, digital photography, instant messaging, etc.) as tools for cultural expression underscores a better

description of the relationship between database and narrative (and, to not limit the discussion, the relationship with lyric, song, sculpture, et al.). The problem arises from presuming the devices of database and narrative are equivalent cultural structures. They are not.

The *PMLA* conversation at the center of the database/narrative debate demonstrates the errant presumption. When Jerome McGann, in his response to Folsom, offers that narrative is "as ancient a form of cultural expression as we know" (1589) he discounts the etymological appearance of the term dating it to the mid 16th century.[1] Unless McGann is suggesting that mankind had no forms of cultural expression before the Renaissance, he must be referencing not the word *narrative* but instead, narrative's concept. Since the concept of narrative can be recognized in early civilization, Folsom and McGann logically respect narrative as an older genre than database. This perception is supported by database's etymological roots, which date the term to the mid 20th century.[2] However, if dating narrative to its concept and not its verbal etymology, shouldn't database be given the same historical understanding?

Part of what confuses the discussion of the relationship between database and narrative (and the larger relationship of the written/printed word versus its digital counterpart) is a misunderstanding not of the meaning of the two words, but of their historicity. Intuitively defining narrative as an older form of cultural expression than database neglects the concept of database. A database is a collection of information – data – organized to make obtaining meaning from that information as easy as possible.

---

[1] See *Oxford English Dictionary* entry for "narrative."
[2] See *Oxford English Dictionary* entry for "database."

Using this concept of database in parallel with the story of the Cumaean Sybil demonstrates both how familiar and how "ancient" database is.

The story of the Sybil has a rhetorical moral (and is thus a favorite tool of modern rhetoricians).[3] Its lesson is to emphasize the importance of structure. The words – data – must be organized – "databased" – in order to extract meaning. The way in which the data is presented – the database's configuration (i.e. sentence structures) – allows for interpretation by software – in this case, the inner workings of the human mind.[4] Sentences, the story of the Sybil at Cumae argues, are a form of database. This same concept can be narrowed to words, which are collections of data – letters – organized to make obtaining meaning from the data – the meaning of the word – as easy as possible. The concept can also be expanded to paragraphs, where sentences become the data. Or to chapters, where paragraphs become the data. Or to books, where chapters become the data. For example, what is the bible if not a database in book form? While a sign with the words "Bible, Page 753" has no easily discernable meaning, the phrase "John 3:16" refers specifically to how a user should retrieve information from the "Bible database." The naming and numbering of the sections is a tool for interpreting this thousands-of-years-old database. Narrative, and other forms of cultural expression and analysis, function the same way – they function as tools for interpreting database.

Recognizing narrative as a tool for interpreting database is difficult if it is presented as linear narrative. But escaping the constraints of linear narrative highlights

---

[3] For a more in depth discussion of rhetoric as a result of structural formation, as well as the use of the Cumaean Sybil myth as it relates to rhetoric, see George Gopen's groundbreaking *Expectations: Teaching Writing from the Reader's Perspective*.

[4] Martha Nell-Smith, in a lecture on the importance of the human component in computing, describes how the human "software" remains the most important and complex software in any computational system, explaining: "The fanciest computational software can do nothing interesting at all, unless directed and engaged by the most important software of all – that proffered by the human touch, by, in other words, *you/us/me*" [sic] (3).

its database roots. To escape those constraints, Pamela Jennings shows linear narrative as the product of a writing-based culture. "The written culture's notion of narrative derives from the theory of dramatic progression expounded by Aristotle in his *Poetics*," Jennings explains.

> The *Poetics* presents a strict guideline for the drama to follow from beginning to end: the narrative increases in intensity to the climax and then gradually reaches an end parallel in tone to its beginning.
>
> (346)

But this notion of narrative is not more natural than any other. It is, as Jennings calls it, "arbitrary." Compared to other options of cultural expression, it is (was?) more convenient for the society in which it was created because it "encourages linearity and truncation of thought" (347).

The linear form of cultural expression is rooted in the traditions of a writing-based culture. But the narratives of oral cultures differ. "Unlike literature based upon the *Poetics*," says Jennings, "African oral literature may contain numerous crises or peaks tangential to the nuances of the story, reflecting the environment it is told in and the responsiveness of the audience" (347). When narrative is no longer rolled into the confines of linearity, its database roots are far less resistible. Instead of lacking an unalterable sequential structure, the variable permutations and combinations database offers open narrative to a world of more life-like cultural expressions because life is not linear. Life is cyclical. "One rhythmic cycle is completed only to begin again," Jennings reminds, "nothing is resolved" (347).

Jennings is not the only discussed reminder of the cyclical nature of life. The story of the Sybil at Cumae already revealed life's cyclical and database-like structure. Her prophetic powers were not a skill but a result of recognizing life's cycles. The

narratives the Sybil withdrew as answers scrawled on palm leaves are examples of narratives being used as tools to interpret her database of life experiences. And, as the story explains, those narratives themselves are also a form of database since they were rearranged to produce different (and often destructive) meanings.

Other forms of cultural expression are similar tools. For example, when a painter paints, the colors and objects on his canvas are a database of images. When a poet writes a poem, her verbal cues create a database of information which readers interpret. The tools of cultural expression – the painting, the poem, the linear narrative, etc. – are databases. When approached through this lens, database becomes not narrative's rival; nor is database narrative's facilitator. Narrative is a form of database, as are other types of cultural expression. Among these other types of cultural expression is the product of digital media discussed by Folsom: the digital archive. His argument should not have juxtaposed narrative and database; it should have paired narrative and digital archive since digital archive, like narrative and lyric and painting and architecture and dance, is a form of cultural expression. And like those other forms of cultural expression, digital archive will not supplant narrative, but exist alongside it. The question to be explored, as a result, is not how will digital archive supplant narrative; the question is: how does digital archive become an effective form of cultural transmission?

Section 2: From the Physical Archive to the Digital Archive

Jorge Louis Borges, in his short story "The Library of Babel," describes a universal library "composed of an indefinite and perhaps infinite number of hexagonal galleries." Each connected gallery contains a matching number of 410 page books, and, in theory, the library's shelves hold every possible combination and permutation of 410 page book, from every book ever written, to every book not yet written, and even to books composed of pure gibberish.

Borges' narrator describes the library as an infinite mystery. Men live their entire lives trying to understand and extract its secrets, but the library is vast, and the amount of knowledge it holds too much for any one man – or indeed, any number of men – to comprehensively navigate. As a result, in the mere beginning stages of their futile efforts to decode the library's answers, all its inhabitants eventually die.

However, in their thousands of years prodding and poking for the knowledge buried among the gibberish – or worse, discovering tantalizing mis-knowledge – the library's inhabitants have discovered two organizational principles on which they believe the library is built: 1) "all books, no matter how diverse they might be, are made up of the same elements: the space, the period, the comma, the twenty-two letters of the alphabet;" and 2) "in the vast Library, there are no two identical books." Using these two conditions, the inhabitants of the universal library deduce that "the Library is total and that its shelves register all the possible combinations of the twenty-odd orthographical symbols (a number which, though extremely vast, is not infinite)."

A not-so-simple calculation would reveal the extent of the Library of Babel's finitude. Somewhere in those possible combinations and permutations of 410 page books,

with 40 lines per page, and 80 letters per line, is a finite number of books – a finite amount of knowledge. And, among the Library's inhabitants, that finality caused excitement. Borges writes:

> When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness. All men felt themselves to be the masters of an intact and secret treasure. There was no personal or world problem whose eloquent solution did not exist in some hexagon. The universe was justified, the universe suddenly usurped the unlimited dimensions of hope.

The inhabitants of the library had hope because they saw, in the seeming pattern of their universe, a possible end and a possible answer to all questions.

However, the possibility of an answer does not equate, automatically, to that answer. What explorers of the Library discovered was that, even if the Library of Babel was finite, the compendium of knowledge was so large that the probability of finding what you were looking for still computed as zero. Knowing – or assuming – the scholars had a comprehensive library did not lead to the knowledge they sought, nor its resulting satisfaction. The result was, instead, the opposite. Borges explains that, "As was natural, this inordinate hope was followed by an excessive depression. The certitude that some shelf in some hexagon held precious books and that these precious books were inaccessible, seemed almost intolerable."

The story of the Library of Babel offers a parallel to the present digital moment – a parallel that should be explored before chasing the tantalizing knowledge offered by expansive and seemingly infinite digital archives. The first component of this parallel to explore is that of the relationship between physical libraries and databases. While the comparative in Section 1 describing verbal constructions as databases stopped at the level of the book and the Bible, it can be expanded to libraries. A library is an organized

collection of data – in this case, books – arranged so as to make retrieval of that data as easy as possible: A library is a type of database.

Classifying (the concept of) libraries as a form of (the concept of) database helps sidestep the first temptation of digital archive – the temptation of comprehensiveness. This tantalizing possibility – the possibility of having all the information if one simply knows the right place to look – is the same problematic temptation of scholars in the Library of Babel. Having a comprehensive archive did not provide answers. Instead, it made the search for answers more frustrating.

This frustration from comprehensiveness should serve as a warning for those creating archives. However, because of emerging digital technologies, a powerful new tool makes the dream of a comprehensive archive with all the world's knowledge seem – while perhaps not plausible – at least more possible, and thus, more tempting. That technology is digital archives. While the size of a library to hold all the world's knowledge in book form may have taken a building as big as the planet on which it was gathering information, micro digital technologies have created the illusion – and temptation – of unlimited storage.

In a sense, the seeming limitless storage space of a digital media database offers an unfair comparison to the physical, library form of a database. While a researcher might look at the walls of a library and say, "These walls can only physically hold a limited number of books," the same researcher can look at a portable hard drive and say, "I can fit the contents of every book in this library on this hard drive. And I can fit the contents of every book in that other library into a second hard drive. And all the

knowledge in both of these libraries can fit on my desk – perhaps, someday soon, it can even fit in my pocket."

This ability to store all the data in a library in boxes that might take up less space than a single book shelf produces a natural response. The same person who condenses a library into a hard drive (or series of hard drives) will, in the natural progression of compression theory, wonder if he can't consolidate the entirety of every library into a single area, placing all the thin-spread knowledge of the library world into one, convenient location. But how useful is the consolidation of knowledge? For example, though endless, the Library of Babel was still compact – at least in a sense that walking from one room of books to another took but seconds. The problem in the Library was not accessing more books – data – but instead, the problem was traversing all the data to find the needed information. Any goal of comprehensiveness, as a result, should be paralleled to the problem faced by the inhabitants of the Library of Babel. If the seemingly limitless available amount of storage (or, more precisely, the extreme compactness of digital storage space) allows for the consolidation of knowledge into one easily traversed location, does that consolidation of knowledge merely let users move the Library of Babel around themselves as they remain stationary? If so, are users any more likely to discover answers?

In addition to the practical use problems of a comprehensive archive, Jacques Derrida, in his seminal archival theory discussion *Archive Fever*, makes an unavoidable case for the inability of an archive to achieve comprehensiveness. Derrida explains that:

> The archive, as printing, writing, prosthesis, or hypomnesic technique in general is not only the place for stocking and for conserving an archivable content *of the past* which would exist in any case, such as, without the archive, one still believes it was or will have been. No, the technical structure of the *archiving* archive also

> determines the structure of the *archivable* content even in its very coming into existence and in its relationship to the future. The archivization produces as much as it records the event. [sic]
>
> (16-17)

Derrida expresses what must be as true for a digital archive as a physical archive. The process of archiving itself produces archivable substance on the content being archived. How, as a result, can an archive – whether digital or otherwise – ever be comprehensive?

These warnings of the inherent problems of comprehensiveness go unheeded by the editors of the *Walt Whitman Archive.* The brief blurb introducing the *Archive* presents the archive's intention of comprehensiveness. The editors, Ken Price and Ed Folsom, write:

> The *Archive* sets out to incorporate as much of [Whitman's vast work] as possible, drawing on the resources of libraries and collections from around the United States and around the World.
>
> (http://whitmanarchive.org/about/index.html)

While the phrase "as much as possible" is not the same phrase as "everything," the implication is undeniable. Surely, if given the physical, theoretical, and financial opportunities, Price and Folsom would prefer a comprehensive *Walt Whitman Archive*. Other printed statements confirm this desire. Folsom even writes in the aforementioned *PMLA* article that:

> Our goal when we began this project in 1996 was to make all of Whitman's work freely available online: poems, essays, letters, journals, jottings, and images, along with biographies, interviews, reviews, and criticism of Whitman. We plan to keep growing and altering the site as new materials are discovered and as we find the time and energy to follow other root systems into the unknown.
>
> (1573)

A visit to the *Whitman Archive* reveals the scope of "all of Whitman's work" that Folsom and Price have already brought online. It is a vast amount. The *Archive* features the main American editions of *Leaves of Grass*, foreign editions, images of hundreds of

manuscripts, personal letters, and a number of the poet's pictures. The A*rchive* includes outside criticism, reviews, biographies, and even helpful teaching materials. But the *Whitman Archive* is not comprehensive, and considering all the materials to be included, the editors would not likely claim otherwise. Despite all the manuscripts digitally available, the archive is surely missing hundreds more – perhaps even ones not yet known, or ones destroyed. Despite all the criticism included, the archive is noticeably missing my freshman year English paper on *Leaves of Grass* and the thousands of freshman English papers like mine. And despite all the published editions included on the Whitman Archive, the 1876 "Centennial Edition" is not present. Whitman made no changes between the 1871-72 edition and the Centennial Edition four years later, but it was an edition printed by the poet. Why is it not included? Who decides when the *Walt Whitman Archive* achieves its goals of having all Whitman-related works freely available online?

That very question can be (hypothetically) asked to two humanists with a stake in the answer. The first is Derrida, who, as already noted, was so concerned with the nature of archiving that he composed a series of lectures/essays entitled *Archive Fever*. In those lectures, Derrida explains that, "Archivable meaning is also and in advance codetermined by the structure that archives" (18). For Derrida, the decision of what to include in the archive is not a conscious decision of the archivist on a per-component basis, but instead, the decision is one made in advance as a result of the archive's structure.

In addition, and not mentioned by Derrida, is an argument of practicality. While a basic human desire for comprehensiveness would encourage every archive to be just that – comprehensive – the structure Derrida refers to is not as much a structure of what an

archivist might want to include, but what can reasonably be included considering the archive's (and life's) physical limitations. For example, an archivist specializing in Shakespeare, in an ideal world with no constraints of time and space and money, might want a copy of every version of every Shakespeare work ever printed, but such a goal is not practical. Instead, the astute archivist spends his time (and resources) as best he can in order to obtain the most complete archive available. Whether that archive becomes the largest archive of *Othello* texts, or the largest archive of Shakespeare pre-1600 texts, or the largest archive of all Shakespeare texts is more a question of resources than intent.

Digital archives, in relation to Derrida's pronouncement, are a way of overcoming some – but certainly not all – of an archive's physical limitations. They can minimize physical limitations of space and access, making it possible to archive anything and everything that might seem relevant. They can, through digital replication, even minimize the limitation of having only one copy of a document or component. As a result, the *Walt Whitman Archive*, along with many other academic and non-academic online repositories from Amazon.com to Google, is tempted into believing it can achieve comprehensiveness, but it still cannot. Posing the same question asked of Derrida to the second invested humanist explains why. If asked, "Who decides when the *Walt Whitman Archive* achieves its goals of having all Whitman-related works freely available online?" what would Walt Whitman himself say?

No one can technically ask Walt Whitman, but his poetry already provides a sort of answer. One example appears in Whitman's poem "To A Stranger." In it, the poet writes:

Passing stranger! You do not know how longingly I look upon
you,
You must be he I was seeking, or she I was seeking, (it comes
to me as of a dream,)
I have somewhere surely lived a life of joy with you,
All is recall'd as we flit by each other, fluid, affectionate,
chaste, matured,
You grew up with me, were a boy with me or a girl with me,
I ate with you and slept with you, your body has become not
yours only nor left my body mine only,
You give me the pleasure of your eyes, face, flesh, as we pass,
you take of my beard, breast, hands, in return,
I am not to speak to you, I am to think of you when I sit alone
or wake at night alone,
I am to wait, I do not doubt I am to meet you again,
I am to see to it that I do not lose you.

The theme Whitman creates in this short poem is one often repeated throughout his poetry. The title, "To A Stranger," indicates his audience, and to that passing stranger he explains that, whether they realize it or not, they have had and will continue to have a lasting impact on one another's lives. They are connected, by friends, relatives, events, actions, and all other sorts of interactions that ripple, chaotically, through the progressions of life and time. As a result, even though Walt Whitman might never speak to the passing stranger, that stranger's life is connected to his. And if an archive is to achieve the goal of having all Whitman-related content, Whitman himself might wonder, "Where is information about every man, woman, and child I ever passed along the street?"

As though recording the lives of those who had merely passed Whitman would not prove challenge enough for any archive, Whitman believes he is also influenced by those he's never met, those from past generations, and those from future generations. He expresses this belief in the sixth section of "Song of Myself" where he asks:

> What do you think has become of the young and old men?
> And what do you think has become of the women and children?
>
> They are alive and well somewhere,
> The smallest sprout shows there is really no death,
> And if ever there was it led forward life, and does not wait at
> The end to arrest it,
> And ceas'd the moment life appear'd.
>
> All goes onward and outward, nothing collapses,
> And to die is different from what any one supposed, and
> luckier.
>
> (123-130)

For Whitman, every person – past, present, and future – is connected to, influenced by, or has an influence on every other person. Thus, if asked what a comprehensive archive might include, Whitman would surely expect the inter-connectedness of mankind to necessitate the inclusion of everything, making all the seemingly infinite space and accessibility of the digital archive appear ill-suited to the task of comprehensivity.

Perhaps, in the very comprehensiveness and inclusiveness of his poetry that suggests universal interconnectedness, Walt Whitman is providing the only viable example of a truly comprehensive archive – the universe itself.

While the *Walt Whitman Archive* continues its admirable attempt at comprehensiveness – however contradictory to the beliefs of its central subject matter that attempt may be – not all archives make the same editorial decision. The editors of the *Dickinson Electronic Archives*, for example, take a different approach. That approach, however, was not without its own seduction by the idea of digital comprehensiveness.

In her introduction to *Emily Dickinson's Correspondence*, Martha Nell Smith, editor and founder of the *Dickinson Electronic Archives*, admits the original goals of the

Archives, explaining: "Immensely exciting about the original goal of the *Dickinson Electronic Archives* scholarly edition was to collect – by deep linking and markup – this diaspora of surviving Dickinson documents, gesturing toward the promise of completeness." Smith, unlike Folsom and Price, recognized the futility of such a goal, describing it instead as "the romance of comprehensivity."

The result of the DEA's departure from attempted comprehensiveness is a digital archive that dramatically deviates for the *Whitman Archive* model. Whereas the *Walt Whitman Archive* noticeably and immediately directs its users to the vast stores of collected Whitmania held on its servers, the *Dickinson Electronic Archives* focuses on pieces of Emily Dickinson's source documents, analyzing them in order to build and share the knowledge those documents inform. The imperfect metaphor I will appropriate for this comparison is that of a puzzle, with the central figure of the author being the broken picture. The *Walt Whitman Archive* attempts to provide all the pieces of the puzzle but leaves the adjoining of those pieces entirely to the user. The *Dickinson Electronic Archives* provides fewer pieces, but offers suggestions for ways in which those pieces might be joined.

Instead of arguing which current approach to digital archive is better (the decision is surely rooted more in personal preference and personal intentions than anything else), the more important lesson is recognizing how neither approach creates comprehensiveness. The *Whitman Archive*'s approach can have an expansive collection of pieces to the Whitman puzzle without ever having them all, and the DEA's approach can offer an expansive collection of ways to piece together the Dickinson puzzle without offering them all. Neither is or ever will be comprehensive.

Perhaps as importantly as recognizing the inability to achieve comprehensiveness, Smith and her original attempt to create comprehensiveness with the DEA provides a valuable lesson about why the quest for comprehensiveness is problematic "That promise [of comprehensiveness]," Smith warns:

> Temporarily obscured what are in fact vitally important critical achievements that can be facilitated in the digital realm, achievements that do not depend on the romance of comprehensivity. That mind-blowing capacity for gathering together that which had been scattered can distract one from posing questions about the archival logics of the physical and virtual archives and about the archival practices both informed by and informing those logics.
>
> (*EDC*, introduction)

Smith's statement will provide the foundation for all further discussion of digital archives. The creators of digital archives cannot be consumed by the "mind-blowing capacity for gathering together that which had been scattered." Instead, the creators should pose questions about the distinct natures of virtual archives, and from the resultant answers, they must decipher the most responsible and useful practices of the archival construct as related to the user's needs.

The user's needs are the final organizing component in digital archive concepting and construction. As with any tool of cultural expression, an archive exists for an audience, and an archive is successful when it adequately addressing that audience's needs. However, no set of rules dictate what person or entity is or should be responsible for the creation and maintenance of digital archives. Should libraries manage them for the benefit of patrons? Should scholars manage them for the benefit of researchers? Should universities manage them for students? Should foundations manage them? Should governments manage them? Should publishers?

To this point in their history, digital archives have been managed primarily by literary scholars for academic research. Academic Whitman enthusiasts, for example, manage the *Walt Whitman Archive.* Academic Dickinson enthusiasts manage the *Dickinson Electronic Archives*. The *Rossetti Archive* is managed by, as expected, academic Rossetti enthusiasts. The result of this trend does more to define what literary archives should be than the fact that digital archives can, technically be managed by anyone.

For example, digital archives could be managed by libraries, or publishers, or technology offices. Digital archives might even be managed by people entirely unaffiliated with the university. Such a model is common on the Internet. A Brittney Spears fan website or a Washington Redskins fan website might be completely unaffiliated with the entities they discus. Why couldn't a digital Milton archive be managed not by a university-affiliated Milton scholar, but instead a person who simply enjoys reading John Milton?

Though anyone or any entity can, theoretically, manage a digital archive, the responsibility has been taken by university-affiliated literary scholars. This distinction of ownership is important because, more than any other archive facilitating component, it defines the ways in which digital archives are and should be structured. Literary scholars are not libraries, and as such, are not required to do the work of libraries; as a result, a digital archive should not be a comprehensive repository of research materials that might actively assist, but does not actively create knowledge. Literary scholars are not publishers; they do not need to distribute the writings of others, and they do not need to develop an archive that is profit-driven. Literary scholars are not information technology

specialists; while they might rely on technologies of the digital medium, their purpose is not the development of new technologies. And literary scholars are not casual fans; while they often are enthusiasts of their topics, their purposes extend beyond that of the traditional fan whose goal is to know more about his passion.

The literary scholar's goal is the production of knowledge. Libraries don't produce knowledge, they compile knowledge. Publishers don't produce knowledge, they distribute knowledge. Technology specialists don't produce knowledge, they distribute knowledge in a digital medium. And fans don't produce knowledge, they acquire knowledge.

Using these distinctions of purpose between the literary scholar and those of other entities that might have taken responsibility for digital archives, digital humanities scholars can define the purpose of a digital archive run by literary academics. Components of the functions other managers might perform will exist. An archive cannot produce knowledge without first compiling knowledge. For an archive to be beneficial it must distribute knowledge. Since the nature of the medium is technological, the unique distribution medium will require some technology innovation. And because those doing the work of the digital archive are enthusiasts of the materials being archived, they will want a place where they can acquire new knowledge about their subject. As a result, the digital archive will incorporate and require aspects of a library, of a publisher, of a technology specialist, and of a fan. But the ultimate purpose for the literary scholar is a digital archive that provides new knowledge and becomes a tool of cultural expression. Achieving this goal is not the product of ignoring the other functions a digital archive might have, nor is it a product of ignoring the functions for which other

managers might be more responsible. Instead, the most successful digital archive managed by the literary scholar is an archive that synthesizes the components other entities are normally responsible for to produce a space in which new knowledge can be created and shared.

Inter-Section: Digitizing the Library of Babel

The universe (which others call the Library) has changed a great deal over the last 60 years. According to the older men of the Library, it used to be composed solely of an indefinite and perhaps infinite number of hexagonal galleries, with vast air shafts between, surrounded by very low railings. The galleries connected to one another, either through narrow hallways or stairwells, and the arrangement of each gallery repeated the previous gallery invariably, with twenty shelves, five long shelves per side (except two), with each shelf containing thirty-five books of uniform format. Each book was four hundred and ten pages; each page, of forty lines, each line, of some eighty letters which were black in color.

As the many generations of men passed through the library, they gradually began to discern a sort of order to the seemingly infinite chaos. First, they discovered that the total number of orthographical symbols in each book was twenty-five. Second, they believed that the Library contained all possible variants, combinations, and permutations of the twenty-five orthographical symbols on 410 pages, meaning the Library contained all possible books. While some books – in all probability, most books – would appear as absolute gibberish, somewhere in those countless books and repeating hexagons were books with definitive answers and statements and explanations. Some men even believe on some shelf in some hexagon there exists a book which is the formula and perfect compendium of all the rest. But, to my knowledge, it has yet to be found.

The library is vast – some even argue it is infinite and interminable – meaning the probability of finding any one book, even that catalogue of catalogues, can be computed as zero. Still, that improbability has not stopped the men of the Library from searching.

Someone once proposed a regressive method: To locate book A, consult first book B which indicates A's position; to locate book B, consult first a book C, and so on to infinity. In searching adventures such as these, many men have squandered and wasted their years.

More than half a century ago, one of the wanderers of the Library developed a new method of searching for books – a method many in the Library believed would not only revolutionize the knowledge search process, but would also ensure that we would find the books we had previously sought in vain. He called the process "digitization." Instead of wandering the hexagons and paging through the books one leaf at a time, he created a machine he called a "scanner." This scanner could make a digital image copy of a page and store it on a device he called a "drive" which took up no more room than a person's hand. Each drive could hold millions of pages – equaling thousands of books – and thus, instead of distributing the knowledge of the books through vast and untravelable amounts of hexagons, the man argued every book could be digitized and stored together in one compact and central place. Then, without ever leaving a single hexagon, people could access the digitally stored page images through a device he called a "computer." The computer has changed shapes over the years, mostly getting smaller. The first computers took up entire hexagons, but today's computers are not much larger than the books they are replacing and have a glowing sheet of glass beneath one of their covers capable of displaying varying images.

At first, the elders of the Library did not care for the man's computer invention, and they did their best to outlaw the process of digitization, but as is the case for all men

who resist societal progress, those men eventually died and were replaced by younger generations who saw the potential of centralizing information.

As the years progressed, so too did the popularity of the computer and digital archiving. Thousands of men were sent throughout the Library to scan every book on every shelf in every room. Some of the most dedicated men could scan an entire room in a week (if they worked through most nights), and in the decades since the task began thousands upon thousands of rooms have been digitized. The process is still not complete today, but the best estimates believe we have digitized nearly a third of the Library, though how you can digitize any portion of a possibly infinite repository remains a source of discussion and disagreement for many.

Regardless of the percentage of the Library now digitized, the current electronic archive is vast, and other men – those not appointed to the task of scanning – are searching the books in their digital formats in our continued attempt to probe and uncover the Library's secrets and revelations. Some men, such as myself, rely on computers only part of the time. They still enjoy the occasional freedom of wandering hexagons and manually paging through books, though they increasingly spend more time scanning the Library with computers. Other men, however, neglect the physical library completely, opting instead to spend those hours, those days, those weeks, those years, those lifetimes once reserved for shuffling through the Library's many hexagons to instead stare at computer screens, continuously rotating through page image after page image.

Thousands of hexagons of books have been displaced or destroyed – after digitization, of course – to make room for terminal hexagons (as hexagons with computer terminals are known). These terminal hexagons have their four walls of books replaced

by computer shelves, six computers per wall and totaling twenty-four computers per room, as well as a table centrally located in the hexagon with a main computer attended by a technology professional who both oversees the work of the others in the hexagon as well as solves any computer errors that might arise during the process of searching. There are many computer errors and the technology professional is often a hexagon's busiest person.

Instead of shuffling through their Library universe, those men who live and work in the terminal hexagons allow their universe to shuffle, more conveniently they argue, around them. Some men have never even left the hexagons in which they were born, all but their most basic arm and hand muscles atrophied by lack of use. These men, men who have viewed in a decade as many books as some men view in a lifetime, are highly revered in our society for the vast amounts of knowledge they've encountered. One, the eldest grandson of the computer's inventor, is said to have encountered a book with 12 coherent pages in a row that, once translated from the Anglo-Germanic dialect of its original wording, was determined to contain a detailed description of the humming bird's reproductive cycle. And while no one from the great terminal hexagons has ever encountered a humming bird, assuming one should eventually fly in, perhaps in some distant lifetime, they will surely understand the little creature's reproductive processes far better than I.

Despite these small successes of the newly digital archive, a growing number of the Library's inhabitants have expressed concerns. They complain, since not everyone has access to the computers, any books displaced or destroyed by the many thousands of terminal hexagons are not accessible to those without computers. But the builders of

terminal hexagons counter this problem by constructing more of these hexagons, arguing that soon every inhabitant of the Library will have computer access, and thus, there will no longer be a need for physical book objects.

This insistence on giving all inhabitants of the Library computer access has its complaints as well. Some Library travelers have argued, perhaps correctly, that digitizing the Library does not make the Library more accessible, nor does it make the Library's users any more likely to find the answers for which they are searching. Instead of users moving through the Library, they point out how the digitization process merely forces the Library to move around its users, but the probability of finding any specific book in the same vastness is still as unlikely as ever. Again, the proponents of the digital library have responded. Less than two decades ago they introduced a "search" function. This function allows users of computers to scan all the digitized text for certain words and phrases, making it possible to find in the Library every instance of a specific orthographic symbol cluster.

Some have hailed the invention of the search function as one of the most important developments in human history, perhaps only trailing the computer itself. But the more pragmatic men among the library have noted the search function's limitations. For example, if the Library is indeed as expansive as men believe it to be, even a computer searching one million times faster than a man would still require any number of lifetimes to complete its search, and the man reviewing those search results may never complete his task. But, as I begin to feel my eyes strain more every day I stare at the soft glare of a computer monitor, I find myself wondering how useful a search function is if one can never know exactly what he should search for.

Not too many years ago, certainly no more than a decade has passed since, while on one of my increasingly less-frequent trips away from computers and the digital archive, I was many days walk from the nearest terminal hexagon when I came upon a young man laboring over a cluster of what looked to be broken computer parts scattered about the floor. I assumed he was one of the rumored technology professionals to have gone insane as a result of the constant computer errors he'd been forced to fix, and I would not have disturbed him, preferring, instead, to slip quietly into another nearby hexagon, but for a small piece of the detritus I accidentally stepped on. To my immense relief, when the man turned upon hearing the crunch beneath my foot, he did not look the least bit insane and, to the contrary, insisted I be the first to test his new invention.

As he hovered over the device, appearing to make some final adjustments, he explained that, while digital reproductions of texts were powerful tools for compacting and disseminating the written content of books, their fault was that they did only that and nothing more. "What if," he reasoned, "the content of books is more than just the words on the page?" He presented a catalog of other variables about each book: the margins, the bindings, the page thickness, even the smell. His point, I'm sure, was that while digital copies recreated the physical image of each page of a book, they were not detailed enough to accurately and completely digitally reproduce every component of the physical object, and perhaps the true answers we were seeking inside the library could be found not only in a book's letters and words, but in those other physical characteristics.

This man was not, I should explain, the first to put forth such theories. The elders of the Library still tell stories of a legendary man who read books by holding them to light, arguing that the translucence of a page influenced the meaning of the text. Using

this method he discovered a book he claimed predicted the day of the Library's destruction. The day he predicted is now nearly three centuries past, but the lesson of his story is a repeated reminder that the answers to searches might be visible in places we do not commonly look. Even an old tale traditionally told when putting the children of the Library to sleep speaks of a man who found a book that, while he could not read the language, he was positive was a cook book because each page smelled of a different kind of pastry.

While in my youth I admit to ascribing to theories of the importance of the physical artifact while reading books, analyzing such components was too time-consuming, and the digital archive limited the extent to which a man might scrutinize the physical object. So I, like most others when searching the Library, learned to concentrate on the text of a book and ignore all other aspects. This man I had met, however, refused to overlook the physical components delivering the text, deciding instead to develop a way to digitally replicate them.

He presented to me his new invention; it was a marvelous object. It had all the letters from a book, but also reproduced seemingly every element of the physical book object. The detail on every page was exquisite. Each margin of the digital page equated perfectly to that of its corresponding page in the book. The fading or darkening of fonts was recreated gracefully and without error. The three dimensional presentation of the pages could trick even the keenest of eyes, and the mechanism to turn from one digital page to the next was so realistic I felt as though I was turning a page in an actual book. Even the smell of the digital object he created had the scent of a dusty old Library book, untouched for hundreds, perhaps even thousands of years. I marveled at the device as he

explained it. He told me the amount of digital storage required to create such an exacting copy could have stored thousands of books in what he described as "the old way" – the way in which books are presently digitally preserved. But the extra storage was essential to properly recreate all of a book's physical properties.

My amazement at his invention must have encouraged the man, and he was eager to begin digitizing the next book so that he might have an entire hexagon digitized before he shared his invention with the rest of the Library's inhabitants. I excused myself, not wanting to disturb him any longer, and walked toward the hall leading me to another hexagon. As I left the room, I watched the man. He closed the cover on his device and fitted it onto the shelf. Again, I was amazed at its exacting reproduction – its dimensions allowed it to fit perfectly in the space on the shelf where the book it was reproducing once rested. As he wedged into the slot his digital reproduction, the man withdrew the next book on the shelf, and began work on its digitization while I, in a new hexagon, opened one of the Library's long-untouched books and wondered when would come the day in which all books in the Library would be perfectly and exactly digitized.