

## Title: WALT WHITMAN'S ANNOTATIONS

### DESCRIPTION AND STATEMENT OF SIGNIFICANCE

In the field of scholarly digital preservation, there are few projects that gather and display an author's manuscript annotations. Annotations made by a writer in the margins of printed texts or images are crucial sources for analysis in literary, philosophical, and historical study because they are rare evidence of direct interaction between a reader and his or her influences.

Annotations also demonstrate the range of such influences, which often reach far beyond the genres in which the annotator worked. While a small number of influential scholarly studies in literature, history, and bibliography have emerged from the archival study of marginalia, such study has not penetrated the methodologies of the humanities because editions are available for only a few writers and collections of annotations are scattered, fragile, and hard to find.

Print volumes of documents featuring combinations of printed text, annotations, and images, are expensive; this cost partly explains the paucity of such editions. But even among digital archives such projects are rare. The powerful searching and display capabilities of a digital platform can make such collections far more accessible, but electronic presentation of annotated documents is difficult to effect in practice. Annotations depend on more than their linguistic content to make meaning; it makes a difference *where* on the page a comment is made. The significance of spatiality in relating layers of information pushes standards-based markup to, and perhaps beyond, its descriptive limits. Marginal annotations, in particular, open a door onto a much larger problem: despite leaps in the automated transcription of printed text, handwritten texts--which include vast numbers of key cultural historical documents--continue to be underrepresented in free online digital resources.

Under a Digital Humanities Start-Up Grant (HD-50236-07, 2007-08), we created an efficient, open-source tool for transcribing and capturing the locations of handwritten and printed words and pictures in relationship to digital images of source documents. We now apply for NEH funding to preserve and give free public electronic access to Walt Whitman's manuscript annotations. This hitherto uncollected and largely unpublished set of extraordinarily diverse and sophisticated documents shows America's most famous poet in-the-making. From classical writings to Tennyson, from Persian poets to phrenological journals, the influences on Whitman's work were manifold. For the first time, students, scholars, and casual readers will be able to explore the fertile ground of Whitman's self-education, through his reactions to the literature, history, science, theology, and art of his time. Whitman's reactions range from the caustic to the puzzled to the awestruck, and take the form of everything from simply underlining significant passages to full-length expository responses.

While electronically gathering, preserving, and making freely available these documents would alone be a tremendous step forward, we are in the position to do much more. First, the context of the *Walt Whitman Archive* gives us the power to link these annotated documents to later ones they influenced. Second, we will publish a database of Whitman's reading--a kind of virtual library of one of the world's most important literary figures. And finally, using a customized search engine and the interface we created under the start-up grant, we will offer analytical tools for users of the archive that will help researchers shed new light on Whitman's writing in the broad context of nineteenth-century literature and culture.

## TABLE OF CONTENTS

<i>Section</i>	<i>Page</i>
I. Narrative	
a) Annotations and American Literary Studies	3
b) History, Scope, and Duration of the Project	11
c) Methods and Standards	13
d) Work Plan	17
e) Staff	18
f) Evaluation and Dissemination	18
g) History of Grants	20
II. List of Participants	21
III. Budget	22
IV. Appendices	
a) Appendix A: Résumés of Principal Participants	25
b) Appendix B: Works cited and related resources	32
c) Appendix C: Holdings and costs of digitization	34
d) Appendix D: Links to demonstrations and standards	35
e) Appendix E: Letters of Commitment and Support	36

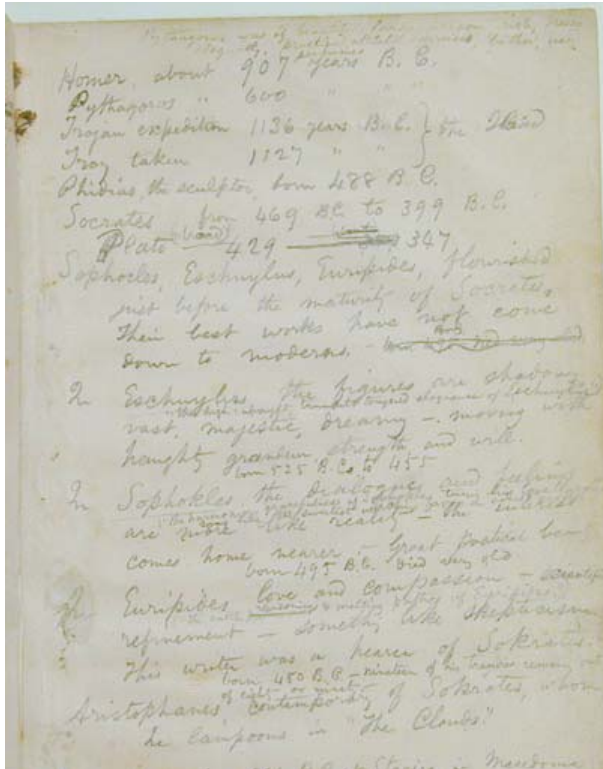
## NARRATIVE

**Annotations and American Literary Studies**

We apply for funding for a three-year project to preserve and give free electronic access to Walt Whitman's manuscript annotations, to be hosted at the *Walt Whitman Archive* (<http://www.whitmanarchive.org>). This neglected set of hundreds of documents written and drawn upon by America's most famous poet shows the process by which he came into writerly being. In his poetry, Whitman famously depicts himself as a "rough," an expression of the American land and way of life. Yet his annotations reveal that from classical rhetoric to the poetry of Tennyson, from Persian mysticism to nineteenth-century phrenological journals, the influences on Whitman's work were historically deep and culturally diverse. For the first time, any students, scholars, or casual readers with access to the World Wide Web will be able to explore the fertile ground of Whitman's self-education, through his reactions to the literature, history, science, theology, and art being discussed during his time. Whitman's responses range from the caustic to the puzzled to the awestruck, and take the form of everything from simple underlining of significant passages to full-length critical expositions. Much guesswork and close reading has been done with Whitman's work to assert its origins or its debt to the literary environment: these documents offer concrete links and will challenge a range of assumptions about Whitman and the relationship between American literature and cultural and intellectual history. Indeed, Whitman's very compositional technique derived in part from his annotational habits.

Defining any genre within Whitman's work is tricky, since the poet made a habit of hybridizing different literary forms for most of his career--breaking boundaries *was* his style. There are literally thousands of documents on which Whitman wrote, for example, no more than a simple identifying citation (as in the case of hundreds of newspaper clippings). And there are a host of documents that muse on a previous author, but that are known to be drafts of later-published work or lectures. For the purposes of this three-year project and with NEH support, we propose to identify, scan, transcribe, and make available to the public documents with significant annotations that are identifiably in Whitman's hand. Our definition of significance is pragmatic, and will doubtless evolve as we encounter new objects: if a manuscript comment is intellectual, rather than merely indexical, we will include the document. In certain cases, even indexicality is meaningful, such as when Whitman aggregates a number of newspaper clippings that are clearly related (which he does for Civil War regiments and for pieces on labor strikes and tramps). We distinguish *annotation* from *notation*. We will focus on Whitman's notes that comment on other writers' works, whether "annotations" (i.e. notes entirely in manuscript) or "marginalia" (i.e. manuscript notes that are together with, if not always in the margins of, a printed version of the original text such as a book or clipping from a periodical). We will not cover those notes or drafts that lead to Whitman's original compositions, whether poetry or prose.

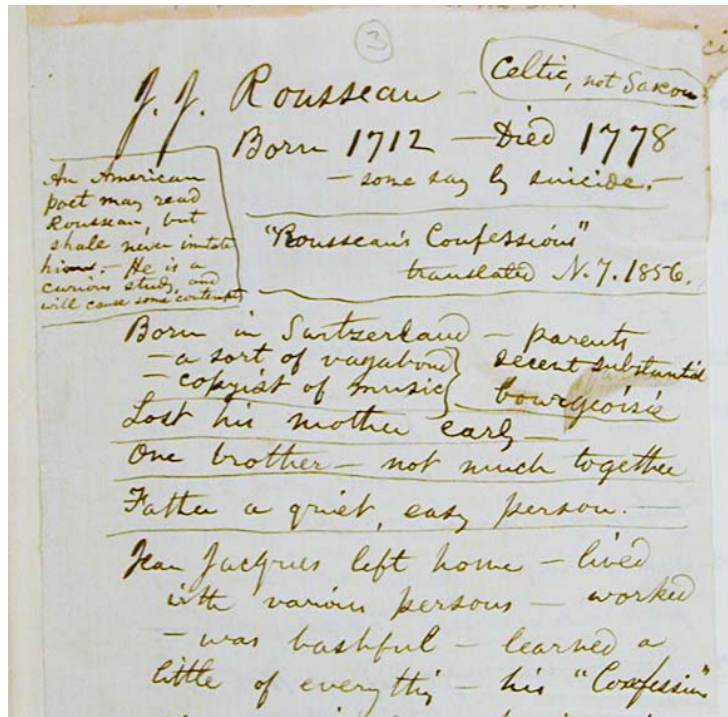
Whitman's marginalia reveal crucial links between his social context and his poetry--the origins of some of his most famous poems may be found recorded in the margins of his reading in nineteenth-century books and periodicals. Finding such connections can bring startling new interpretations to bear on Whitman's poetry, and indeed, suggest useful methods for research into other nineteenth-century authors' works. Take *Leaves of Grass*, for example, the poet's most influential work. Matt Miller has recently demonstrated that when we turn to Whitman's



annotations, we discover a different chronicity to the composition of the first edition of *Leaves of Grass* than has been assumed. Rather than composing his famous 1855 text largely from scratch beginning in a mystical inspiration starting in late 1853, the poet transformed notes he had taken earlier (on a range of texts and about American literary style) into long poetic lines. An annotation on Greek intellectuals in the collections at Duke University (left; Trent Collection, Duke Libraries; Frey III:26) offers an example. While annotations such as this may not appear to be poetry, Miller points out that not only do they often feature content that ends up in his poems, but with its hanging indentation and topical fragmentation, Whitman's annotational style "looks like his signature line" (Miller 118). We estimate that the major archives of Whitman documents contain over 1000 such documents, some many pages long, many previously uncollected and

unpublished; further exploration of Whitman's annotations will provoke scholars to rethink the chronology of the poet's work.

But more fundamentally, Miller's work suggests that Whitman's *compositional method* relied upon annotations no less than on poetic transformation. Whether writing poetry or prose, the poet turned to the notes he had taken--whether during reading, or following a conversation or performance, or from his imagination--when it came time to generate his work. Thus Whitman's annotations represent a fundamental compositional mechanism. They are the root of much of Whitman's published work and key to understanding not just the sources and the chronology, but the very form of both his poetry and prose. Indeed, it is possible to see in some annotations the layering of literary theory, content, and practice. In another document from the Duke holdings, Whitman engages with Jean-Jacques Rousseau, taking detailed notes on the French writer that will help scholars better understand Whitman's relation to continental literature and philosophy (right; Trent Collection, Duke Libraries; Frey III:14). But the poet



also makes a meta-note about how to handle figures like Rousseau: “An American poet may read Rousseau,” Whitman notes in a box at the top left, “but shall never imitate him.--He is a curious study, and will cause some contempt.”

Digitization and free access to these annotations will change dominant interpretations of Whitman's poetry, partly through revelations about dating and sources, and partly through recontextualization. Take the famous “trapper's bride” scene in the poem that became “Song of Myself.” An unusually positive representation of miscegenation for the mid-nineteenth century, it has long been known that Whitman based this scene on a painting by Alfred Jacob Miller called *The Trapper's Bride*. But a document in the Library of Congress reveals a more sustained engagement with the question of the representation in visual and poetic art of Native Americans. In the margins of an 1856 clipping on Indians in art from *The Crayon*, the poet references *The Trapper's Bride* (“by a Baltimore artist,” Whitman remembers) and then implies that his own treatment of the subject in the 1855 *Leaves* wasn't sufficient: “Has any poem yet really pourtrayed them? Would not the Indian...be a good theme for a full poem?” (LOC card 16). This comment may suggest that a much-later poem like “Osceola”--which fulfills this directive to devote an entire poem to the native theme--had its origins considerably earlier than has been supposed. And the annotations speak eloquently to the recent surge in transnational approaches to literature. Whitman's “Prayer of Columbus,” for example, parts of which are inscribed in the wall of the District of Columbia's Archives/Navy Memorial metro station, seems to have been inspired by his reading of an article in the *Irish Republic* (LOC card 720; Detroit Catalog No. 32). It has long been asserted, too, usually based on internal textual evidence, that Whitman was influenced by middle eastern spiritual writing; the Middlebury collection contains a fascinating set of annotations on Persian poetry that confirms and complicates that assertion, and that will do much to animate the story of American literature, in Wai Chee Dimock's elegant formulation, “through other continents.” Kenneth Price's *Whitman and Tradition* was a prescient study that showed Whitman's complex relationship to European literature and criticism; the methodological emphasis Price put on Whitman's annotations, however, remains rare.

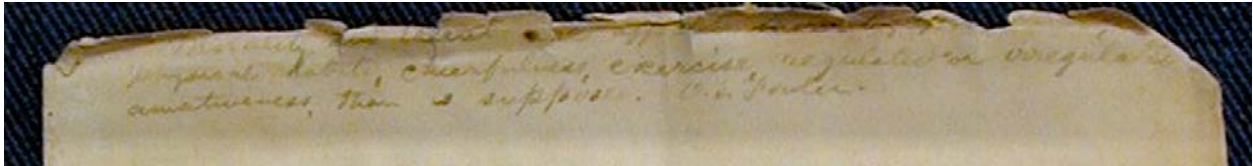
We also learn much about the poet's life from his annotations as a corpus. Having never attended college, Whitman's reading and his conversations with people, together with his visits to the theatre and opera, were his education. If these documents, then, offer insight into the self-educative possibilities of urban nineteenth-century America, they also suggest how Whitman created himself both as a prose writer and editor--and then how he transformed himself to a poet. Floyd Stovall used Whitman's marginalia to show, for example, that the poet's shift in reading from American journals in 1845-47, to British journals in 1848-9, tells us that Whitman was educating himself to become a poet in those latter years, reading literary criticism and accounts of the lives of poets (Stovall 143-152; see also Klammer 71-3).

Perhaps because Whitman's methods of writing drew upon the form and content of his annotations and marginal notes, he held on to much of this material until the end of his life. Thanks to the fastidiousness (and bibliophilia) of his literary executors and friends, a large corpus of these kinds of texts survive. Yet the scholars we have just cited here almost exhaust the list of those who have engaged Whitman's annotations in any depth, largely because the materials are held in archives all across the United States, are difficult to understand without the larger context of similar documents, and have seen only limited publication.

This isn't an uncommon literary historical situation. In the broader field of literary studies, there are few projects that gather and display an author's marginalia. Annotations made by writers in the margins of printed texts or images are crucial sources for analysis in literary,

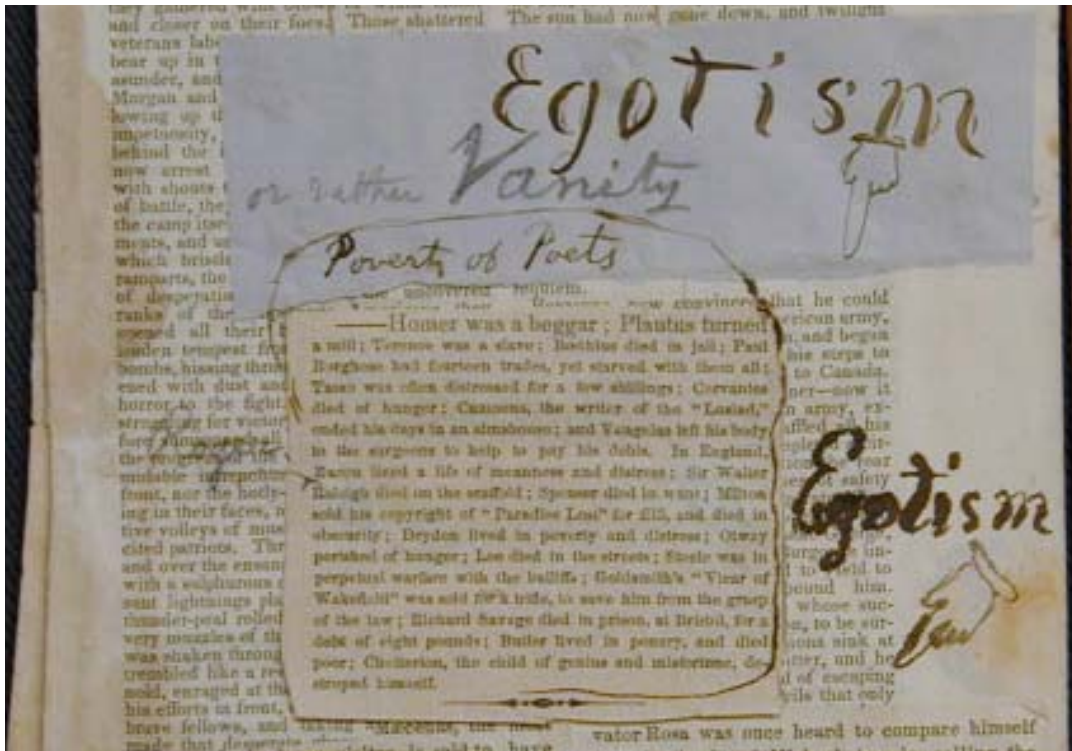
philosophical, and historical study because they are rare evidence of direct interaction between a reader and his or her influences. Marginalia also demonstrate the range of such influences, which often reach far beyond the genres in which the annotator worked. In a famous marginal note, Whitman wrote that “all kinds of light reading, novels, newspapers, gossip, etc, serve as manure for the few *great productions*.” While a small number of influential scholarly studies in literature, history, and bibliography have emerged from the archival study of marginalia, such study has not penetrated the methodologies of the humanities or social sciences because collections of annotations are available for only a few writers, such as Herman Melville and Samuel Taylor Coleridge, and statesmen such as John Adams.

Print editions, even simple transcriptions, of documents that combine printed text, marginalia, and images, are expensive; this cost may partly explain the paucity of published annotations collections. But even among digital archives, initiatives to preserve annotations are still rare. The powerful searching tools of a digital platform would make such collections far more accessible, while the availability of both images and transcriptions of texts would solve economic and ergonomic challenges faced by print editions. Digital imaging and transcription of annotations has much to offer from a preservation standpoint too; marginalia, by their nature,

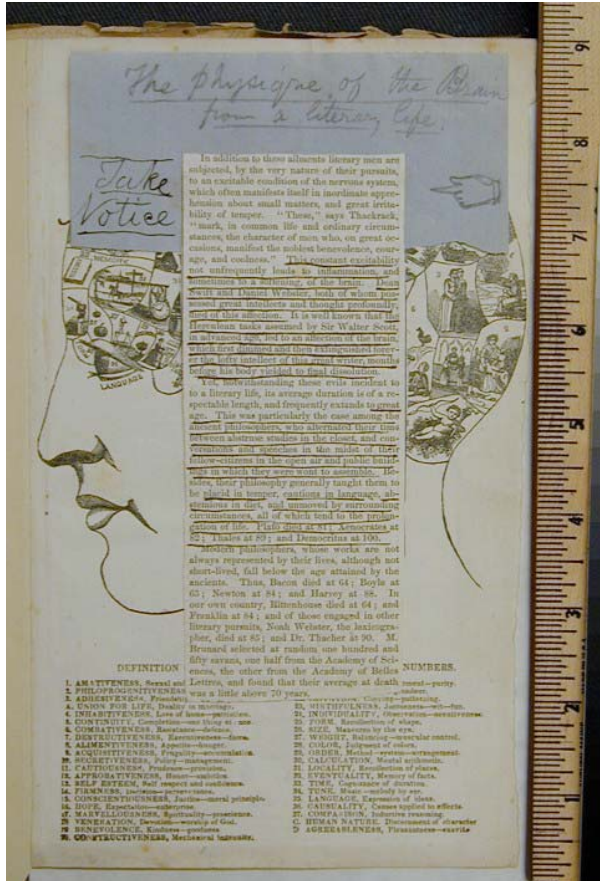


reside at the unstable edges of pieces of paper, and are prone to fragmentation. A notation among the Whitman documents at Duke University offers a telling example, crying out for digital imaging, transcription, and publication--yet no there is no collected publication of Whitman's extensive body of marginalia (above; Trent Collection, Duke Libraries; Frey p. 66 item 24). With the exception of the marginalia section at the Library of Congress, notes and annotations are combined, meaning that the researcher has to look through the entirety of these collections (no small task) to find the existing marginalia. Some of it can only be found in unlikely and undercatalogued places; in oversize folders, correspondence, ephemera, or collectors' specially bound volumes of Whitman objects or nineteenth-century literary ephemera. A focused effort to generate a digitized collection would begin to provide an index that would reduce both research time and unnecessary contact with fragile documents.

Digital presentation of annotated documents has much to offer, but in practice it is difficult to effect; the approach we will take to offering an interface for these documents reflects several important generic features of annotations and the possibilities of the digital platform for giving access to them. Annotations, after all, are a form of text markup, themselves a re-presentation or re-framing of an original text. They depend on more than their linguistic content to make meaning; it makes a difference *where* on the page a comment is made. The significance of spatiality in relating layers of information pushes markup based on the Text Encoding Initiative (TEI) standards to, and perhaps beyond, its descriptive limits. Take, for example, one of Whitman's annotations on a collection of periodical pages he cut out and assembled on top of an 1847 article on George Washington (Trent Collection, Duke University, Frey III:7):



Represented here in about a three-inch-square space are challenges to several levels of scholarly digital editing--analytical, historical, technical, and bibliographical. First, if we think of annotations as a kind of old-fashioned “tagging,” this document exemplifies the difficulties of capturing metadata with more metadata--the theme of Whitman’s markup here is the relationship between poets and their predecessors or between poets and their own reputations. Whitman has manipulated the physical properties of the elements of the collage in order to relate texts without necessarily unifying them. So the precise relationships of one paste-on to others, and to the commentary written on them or between them, are essential to how the document makes meaning. TEI-based XML makes it simple to mark sections that have been underlined or otherwise grouped by a reader on a single piece of text. But the section here labeled “Poverty of Poets” by Whitman embraces three different pieces of paper, and presumably includes the marginal note “Roger” he has added to indicate which Bacon the text refers to--a note written in the margin relative to the annotated text, but *not* in a margin relative to the whole document. That he has written over the underlying document in making that note suggests that we might treat the underlying article as insignificant, as “background” relative to encoding the “Poverty of Poets” section that covers it. But can we be absolutely certain that Whitman did not intend to let the words “uncovered” and “requiem” peek through here? And even if Occam’s razor suggests Whitman’s razor merely cut too close by chance, revealing those words, it should in principle be possible to allow them to appear in multiple hierarchies in our markup in order to encourage richer or more robust interpretations.

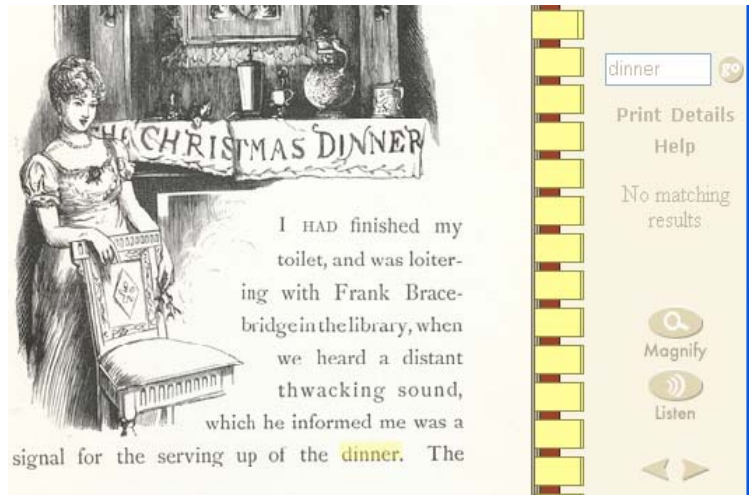
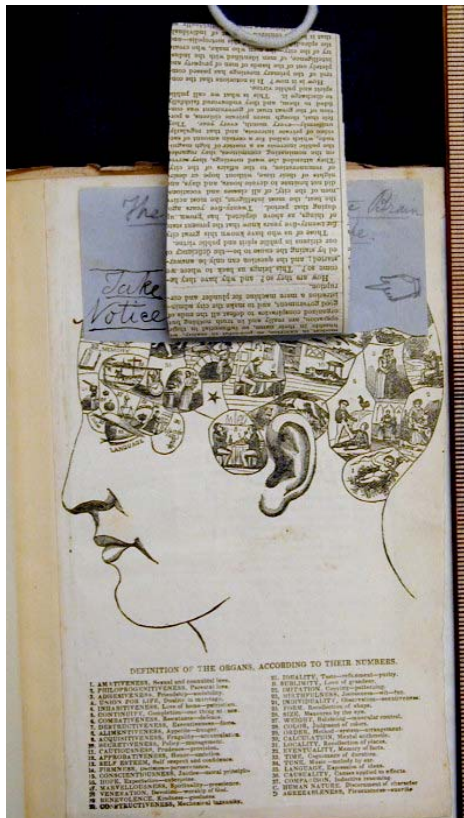


Annotations are made on images as well as printed texts; take, for example, Whitman's notes on a phrenological chart (left; Trent Collection, Duke Libraries; Frey p. 66 item 24). The issues represented in the previous annotation are complicated here not only by the presence of an image in the hierarchy of meaning, but by the fact that the front clipping is pasted on only at its top: Whitman has positioned it so that it can be lifted and the emblematic map of the human brain underneath accessed (below left).

Under a Digital Humanities Start-Up grant, we created an encoding approach for multimedia documents that tackles the basic difficulty of capturing visual relations among content elements (see Methods section for detailed description). We did this in part because of the intellectual significance of textual spatiality, but in part to match user expectations for internet interfaces. While users have become accustomed to the mixing of texts and images in search engines such as Google Earth and Google Books, the only such

interfaces available in textual studies are based on scans of print editions, not manuscripts.

Note below, for example, that in an Open Library word search, terms in print are returned by a search, but not those in images, the word "dinner" in this case (<http://www.openlibrary.org>). It seems unlikely that transcription of manuscript documents will be automated in the near future; our interface makes it easy for encoders to locate a particular word on a page and encode that



word's location richly for machine analysis. Our software and encoding practices allow for tagging, displaying, and searching documents that mix print, manuscript, and visual images. We have adapted standards for encoding coordinates in XML transcriptions so that archival search engines can, at a user's option, visually display--overlaid on top of the document image--results of user searches for manuscript words and phrases together with printed ones and, if desired, for simple image content (such as "chair" or "brain" in the images above) with metadata containing the search terms.

There are a number of benefits to suiting the interface design to the peculiar morphologies of annotations. Having Whitman's marginalia available with both a visual-interactive search capability and transcriptions will make this collection of benefit for students learning paleography. It will also facilitate feedback between editors and audiences: if a user contests the transcription of a word, it will be simple to locate, reference, and correct. A standard coordinate system also makes it easy to reference the site of the contested transcription when opening such questions up to a wider scholarly audience in online fora. Our design has been and will continue to be open source and hence adaptable by other projects. But it has also been built with an eye towards a time when readers have more control over stylesheet displays and when archives offer users more options with respect to searching and analytical interfaces. And finally, if the coordinate system we hope to create is obviously designed to relate text-based searches to image-based results, it could also be used to aggregate linguistic data *based on* coordinate locations. A great range of analytical queries would be possible: to single out all manuscript comments located within particular or relative coordinates on a page, for example; or all pages with a certain rough percentage of page space covered with manuscript. With marginalia in particular and perhaps in other kinds of heterogeneous media documents, such visually biased searches would produce informative new patterns.

Annotations give virtually irrefutable evidence of a writer having read a text. For historians and literary critics, knowing whether or not a writer read a particular work is of tremendous value, but often difficult to ascertain given the scattering of documents and the heterogeneity of archival finding aids. With this in mind, we will use the metadata from the headers of our XML-encoded transcriptions to build a searchable database of Whitman's reading. Whitman's library was a constantly shifting entity during his life, as the poet shared books, periodicals, and letters with his friends and they in turn frequently gave or loaned him texts. In collaboration with Vanessa Steinrötter at the University of Nebraska-Lincoln, and using the basic structure of the Reading Experience Database (RED, <http://www.open.ac.uk/Arts/RED/>), we are creating automated procedures to harvest header information from the annotations XML files to populate a database with MARC information about these objects. The database will also feature a user interface that will allow approved users to add new documents based on other sources, and a "proof" field in which to describe the source of the assertion that Whitman read a particular text. Our database will be compatible with RED's, and we will give that project access to our data; the database and the header-harvesting procedures will be fully functional by the time NEH funding would begin.

Our documentary test-bed during the start-up grant period was drawn from Duke University's collection of Whitman's marginal annotations. The most substantial holdings of Whitman's annotations are found at the Library of Congress, Duke University, the Harry Ransom Humanities Research Center at the University of Texas at Austin (HRC), the New York Public Library, Bryn Mawr College, Middlebury College, and the Beinecke Library. (In communication with archivists at Ohio Wesleyan and the Bancroft Library we determined that

they do not hold documents fitting our definition of Whitman annotations.) Here we will briefly state the nature and known state of each of these collections; see Appendix C for detail on holdings and costs. Because tracing the provenance of some of the texts on which Whitman made annotations can be difficult, particularly with nineteenth-century periodicals, and because versos have been important in dating documents, we will include scans of both rectos and versos of each object carrying an annotation.

The known previous publications of annotations are in *The Complete Writings* (volumes 6 and 7), edited by Whitman's literary executors (a republication of Richard Maurice Bucke's *Notes and Fragments* of 1899); Edward Grier's *Notebooks and Unpublished Prose Manuscripts* (1961-1984; volumes 5 and 6); and in Joel Myerson's two-volume edition titled *The Walt Whitman Archive* (1993). The first features transcriptions of many Whitman annotations, but no facsimiles and no information about the location of documents. In many cases, too, these notes exclude marginalia; where marginalia are transcribed, the source text is sometimes named but never transcribed. Grier's work expands on but overlaps with Bucke's volume, but is likewise incomplete; it contains, however, detailed headnotes that will be the basis of our metadata in many cases for our digitizations of the documents included in his volume. Myerson's edition contains facsimiles from the Duke and Ransom Center collections, but it focuses on the poetry manuscripts and notes, not on annotations--only three documents that we will be preserving digitally here appear in his edition (2:654-55). Several articles and bibliographies, including a list of newspaper and periodical clippings owned by Whitman from the *Complete Writings* (7:63-97) have guided us in our development of the list of documents to be digitized (see Appendix B for a list of these secondary sources).

Here at UT-Austin, the Ransom Center has developed a rich collection of Whitman materials; none of the annotations documents have been digitized. At Duke University, 100 digital images have been created out of roughly 480 (see History, Scope, and Duration of the Project). Middlebury College holds a small connection of richly annotated periodical clippings; none have been digitally imaged. These three collections feature predominantly notes that Whitman made on his reading, both on the original texts and on separate sheets, and booklets Whitman made by gluing together clippings from books and magazines and annotating them. The New York Public Library and the Beinecke Library also have small collections of loose notes and clippings on which Whitman has commented; these libraries have digitized their collections in large part, so we anticipate merely defraying the cost of obtaining copies of 600 dpi TIFFs of each side of each annotated object. Bryn Mawr's special collections holds a set of books once owned by Whitman in which the poet has made annotations; under a previous *Whitman Archive* initiative 30 pages of annotations were imaged, and we propose to image the rest. The Library of Congress's Rare Books division also holds a number of Whitman-annotated and owned volumes, and the manuscripts division holds a substantial number of annotations in both the Harned and the Feinberg collections. These we propose to photograph digitally ourselves (travel costs are listed in the budget; \$1500 requested of NEH and \$1500 contributed from the project director's UT-Austin research account). While many of the objects from the manuscripts division were microfilmed, there are serious shortcomings to the microfilm. Some faults are familiar: the annotations are often illegible and sometimes do not even make it into the frame of the photograph, and the images are black-and-white, making it difficult to use the color of paper for chronological estimation (and often, when the imaged document is not on white paper, the annotations are invisible). The microfilm set is expensive and not generally available

even at research libraries. Finally, the microfilm does not contain images of all of the items in the collections, nor does it always feature both rectos and versos of them.

The documents held in these collections are extremely diverse; as the illustrations above suggest, Whitman wrote marginalia on, cut out, pasted together, commented on, and rearranged for his own use a host of images and texts in the course of generating his unique poetic vision. The well-developed standards of the *Whitman Archive*, a free archival site at which the project director is an editor, will ensure that only open-source, standards-based technologies will be used to create the project's software and encoding schemes (for editorial standards and encoding guidelines see the links provided in appendix D). The *Whitman Archive* offers extensive experience with scholarly digital preservation, a record of successful federally funded efforts, and, most importantly, dozens of ways in which Whitman's annotated documents will "plug in" to the rich resources already available in other sections of the *Archive*. But from the broader standpoint of literary-studies-based archival projects, we hope that ours will serve as inspiration, model, and resource for others. The striking range of shapes that Whitman's annotational practice took has challenged us to develop an approach and a set of software tools that are adaptable, and we are confident that the historical and literary-historical research discoveries to be made using these documents will encourage other preservation initiatives. Marginalia will hold interdisciplinary and non-academic interest, diversifying the reading population for American literature, and will expand a teaching resource that already receives tens of thousands of visits weekly.

### **History, Scope, and Duration of the Project**

An edition of Whitman's marginalia was first proposed at the annual *Whitman Archive* group meeting in July 2005. In 2005-2006, initial discussions of the scope and nature of Whitman's marginal annotations led the project director to look at other marginalia projects and online interfaces for them for guidance. Few of these coming to light at that time, a brainstorming session and extended email conversation were held in summer 2006 at Duke University with participants from Duke Library's digital projects group, professor Timothy Lenoir's digital humanities group, and humanities and social science scholars with experience working with marginalia and other complex multimedia documents.

The conclusion of this working group was that the *Archive's* extant stylesheets were incapable of representing marginal annotations, and that current open-source solutions to both the capture and the display of search results with documents of this mixed nature were also unavailable. In response to this, we applied for and were awarded a Digital Humanities Start-Up Grant to create the software described above and the encoding scheme that we will use, in slightly modified form, going forward. This grant also supported the digital imaging of 100 annotations documents from Duke's Special Collections. The software has been installed at the *Whitman Archive* and is now also being used at *Melville's Marginalia Online*; continued refinement of the software will be conducted in conjunction with these entities and ongoing similar work at the Maryland Institute for Technology in the Humanities (i.e., TILE) and elsewhere. Our work has been presented at Duke University, at the Society for Textual Scholarship annual meeting (2007), and at the University of Texas at Austin; a white paper describing the project's results may be found at the Office of Digital Humanities web site (<http://www.neh.gov/ODH/Default.aspx?tabid=111&id=8>).

In spring 2010, we made contact with all of the major repositories holding documents that fall under our definition, and have obtained the necessary letters of commitment or promises

of access. The total number of images to be generated we estimate at around 2300; given the difficulty of categorizing some of these documents and the heterogeneity with which archives indicate them in finding aids, we know that we will find more documents than we have been able to document in conversation with scholars (including Kenneth Price, Ed Folsom, Matt Miller, Michael Winship, and Ted Genoways); research in secondary literature on Whitman's marginal annotations (see the bibliography in Appendix B); and with archivists (see specific estimates for each major collection in Appendix C).

These documents will be made available in facsimile, so that users of the transcriptions will have access to a clear image of the documents on which transcriptions are based. Given that no one claims copyright on the content of Whitman's manuscripts, we are free to publish transcriptions of them. Reproduction of manuscript images made by the repositories does require their permission; all of the holders from whom we need them have granted us the required permissions (and often generously reduced their rates; see Appendix E).

*The Walt Whitman Archive* sets out to make Whitman's vast work, for the first time, easily and conveniently accessible to scholars, students, and general readers. We have been providing free access to carefully preserved Whitman material for the last fifteen years, and we have projected another fifteen years of work. The archive's long-term goal is to encode and to provide digital images of all the documents in Whitman's vast oeuvre, including manuscripts, letters, notebooks, daybooks, and published work. Since 1995 the *Archive* has received generous support from several universities, from four different federal agencies, and from two private foundations. During this time we have garnered positive publicity in the *Chronicle of Higher Education*, the *Washington Post*, the *American Scholar*, *PMLA*, and other publications. The 2007 edition of *American Literary Scholarship* says that the *Whitman Archive* "may be the most important editorial undertaking in the history of Whitman studies." Our work in creating an integrated finding guide to Whitman's dispersed poetry manuscripts has been honored with the C. F. W. Coker Award from the Society of American Archivists.

Whitman's marginal annotations will serve as a flashpoint for researchers using the *Whitman Archive*. The archive already contains an enormous amount of work, carefully indexed using tags that identify individual works and unique identifiers for each document. As Whitman's annotations are added to the archive, they will be tagged with new identifiers and cross-referenced to other materials already on the archive. Source information for the original texts on which Whitman has commented will be added to headers where known and researched where hitherto unknown. For example, in merely surveying the known collections, we have already identified some source texts using searches through online newspaper and periodical databases. As each annotation document goes into the archive, then, it will become linked to the vast sources already resident at the *Archive*: the editions of *Leaves of Grass*; Whitman's poetry manuscripts; his correspondence; over 160 Whitman poems appearing in forty-eight different periodicals from the late 1830s to 1892; Horace Traubel's *With Walt Whitman in Camden*; and, eventually, Whitman's notebooks and prose works. Our design will also accommodate downloading of our XML-encoded files and transcriptions in multiple formats, protected by a Creative Commons license. Our intent is to structure our preservation of the annotations such that they will interface easily both with resources within the archive and beyond it.

From an infrastructure standpoint, we are lucky to have the resources of both the University of Texas at Austin and the *Walt Whitman Archive* both for the development and the maintenance of the annotations preservation project. The *Whitman Archive* stores archival-quality CDs containing all images it hosts, and has both mirror sites and a server backup system

for all online material, interfaces, databases, and software. It will be the long-term host for the project in its finished form; the University of Nebraska has made a commitment to sustaining the archive, which with an NEH Challenge Grant is now supported by its own endowment. At UT-Austin, the Digital Writing and Research Laboratory has committed server space and time, versioning software, and backup and maintenance for our development efforts here at Texas. The project director will provide a high-quality (Canon 5D) digital SLR camera and backup system for use in off-site imaging; the Department of English at Texas has provided laptops, software, and minor technical support necessary for the project.

### **Methods and Standards**

This project draws on a range of technical precedents. The University of Virginia's Institute for Advanced Technology in the Humanities developed INote, a tool for annotating images in quadrants, components of which have been useful in our conceptualization of relating text and images with greater precision. For marking up (and for displaying using XSLT) the often complex, even parallel hierarchies of annotations, the *Whitman Archive's* poetry manuscripts project gives us a head start, while projects like the Gabriel Harvey marginalia edition at the University of London and the ARCHWAY project have developed useful strategies for XML markup of layered text and annotations. We have also been in productive conversation with Doug Reside at the Maryland Institute of Technology in the Humanities about the Text Image Linking Environment (TILE) project there, which allows some automation in producing standards-compliant XML for linking image to text and images to images. The procedures, standards, and technologies described below have been chosen both with best practices and protocols for scholarly preservation in mind and with an eye to the rapidly changing state of open-source tools for this kind of work. While we are reasonably certain of the technological configuration for the project, then, we have tried to indicate aspects of that configuration that may shift over the intervening months in response to the performance or general uptake of cutting-edge open source technologies for document preservation and access.

#### *Original Document Handling*

With the exception of the items at the Library of Congress, all document handling and imaging will be done--or has already been completed--by conservators at the institution possessing the originals. The Library of Congress documents will be handled according to the Library policies and photographed once, using a digital SLR camera in a stand and light environment approved by LOC staff. Several documents in the Trent Collection at Duke University have already seen conservation repair in order to prepare them for imaging.

#### *Storage and Backup of Master Image Files*

All images that are received from library digitization services or collected by members of our team by digital photography will be converted into lossless TIFF (Tagged Image File Format) files at the maximum possible resolution. These master images will be in color whenever possible, and in grayscale otherwise. As the image files are collected, they will be added to a central repository and renamed according to the *Whitman Archive* schema, indicating the source, timestamp, and a unique identifier. In the development stages of the project this repository will be hosted on FreeBSD servers in the Digital Writing and Research Lab (DWRL) at the University of Texas. All data associated with the project, including these master image files, will be backed up regularly according to the standard backup protocols of the DWRL. Copies of the

master images will also be kept on our development machines and periodically written to DVDs for an additional measure of safety. In order to ensure the integrity of these copies, we will compute the SHA-1 checksums of all images as they are added to the system, and these checksums will be stored in a Mercurial distributed version control repository that will also contain all of the source code associated with the project. As the collection is publicly deployed, these master images will be copied to the *Walt Whitman Archive* GNU/Linux web servers at the University of Nebraska-Lincoln; however, copies will also remain on the DWRL servers.

We are already using this system for storing and tracking master images in other contexts, and we have successfully deployed other projects from the DWRL servers to the *Whitman Archive* servers, so we do not anticipate technical problems at the level of file storage, backup, or network transfer.

### *Web Applications*

The project will involve the development of two distinct web applications: a private application for use in the preparation of the images, and a public interface to the document collection that will be hosted on the *Whitman Archive* servers and will allow users to search and browse the collection. In order for the marginal annotations to be searchable, they must be transcribed and stored in a database that links the transcribed text to regions in the facsimile images. The private web application that will enable this transcription and annotation will be hosted on the DWRL servers and will be used only by members of the project team for document preparation and maintenance.

We will take as a starting point for this application the prototype annotation tool developed by Matt Cohen with support from an NEH Digital Humanities Start-Up Grant in 2007-2008 and described above (see also link in Appendix D). In its current form this web-based tool provides an Adobe Flash interface to high-resolution JPEG images that allows a user (with proper privileges) to pan, zoom, select rectilinear regions, and add or edit transcriptions associated with those regions. These transcriptions are stored as XML text in a relational database on the server. The model layer of the software is implemented in Django, a Python web framework, and the Django data model works with a wide range of relational databases; we have successfully used the application with MySQL and SQLite, and will probably use PostgreSQL in the future. The current XML schema is a simplified version of TEI, and is designed to be converted easily into the specific form of TEI required by the user via an XSLT stylesheet.

While this prototype software has been successfully used by teams at the University of Texas and Boise State University, we intend to make several major changes to the way that it operates. Most importantly, we will follow the lead of projects such as the Text Image Linking Environment (TILE) in separating the image presentation layer from the annotation and transcription layer. All of the image navigation functionality will be handled by an image server such as aDORe Djatoka, an open source JPEG 2000 server developed at the Los Alamos National Laboratory, or IIPImage, another popular open source image server. This includes tiling, deep zooming, and the dynamic generation of images at specified resolutions. These functions are required by both the private and public web applications, and by separating this layer from the other elements of the interface we will be able to use the same underlying solution for both systems. In other words, both our markup specialists and users of the resource will be using the same platform to view the images, metadata, and transcriptions--but approved users will have the ability to add and modify the database's contents.

While both Djatoka and IIPImage are open source and are available under similar licenses, Djatoka depends on a proprietary JPEG 2000 library licensed by Kakadu Software. This fact has not prevented many online archives from adopting Djatoka, however, and we will evaluate both solutions to determine which is most appropriate for our situation. In either case the master images will be the TIFF files described in the previous section; if we use JPEG 2000 it will be only as a service format, to provide better performance at the level of the image server. The choice of image server will not affect the experience of the end user, who will be able to download images that do not require a proprietary library for viewing.

We will also consider using the aDORe Archive platform, either in conjunction with aDORe Djatoka or separately, in order to provide OAI-PMH-based access to the collection (Open Archives Initiative Protocol for Metadata Harvesting). Our current plan, however, is to follow projects like TILE in choosing methods of exposing metadata that are more lightweight than aDORe Archive. This will almost certainly be our approach in the development stages of the project.

In addition to abstracting out the image navigation functionality, we will also make several other changes to the prototype annotation tool. In particular, we will rewrite the schema for the text-image linking data in order to ensure that we can capture the full range of phenomena that is of interest in our collection. We will also implement tools for converting the data into a variety of output formats, including TEI P5 with either the facsimile module or SVG (Scalable Vector Graphics) as the method for handling the transcription-facsimile linking. Management of the image preparation workflow will be facilitated by the integration of the annotation tool and the Trac open source issue tracking system.

We will implement the public web application using a JVM (Java Virtual Machine) web framework such as Cocoon, Wicket, or Lift. This restriction is necessary because the application will run on the *Whitman Archive* servers, and the rest of the *Whitman Archive* is served from a Java web container. The public interface to the document collection will necessarily be more interactive than the majority of the content currently provided by the *Whitman Archive*. Given the size and organization of the collection, for example, users will need to be able to create sophisticated search queries and browse large sets of images. Neither kind of functionality is currently available through the Archive's web interface. We will select a framework after discussing the candidates in greater detail with the server administrators and the consultant, Brett Barney, at the University of Nebraska-Lincoln. Our goal will be to enable faceted searching of encoded regions of the texts as well as of metadata. Users, for example, will be able to search by the usual categories--date ranges, authors, publishers--but also for text features such as strikeouts, underlining, and Whitman's omnipresent "pointy hand" (or manicule).

### *Automation*

Given the size of our document collection, manual transcription of the marginalia alone will be an enormous undertaking; it would be virtually impossible to transcribe the base texts as well. Making the base texts searchable, however, greatly increases the usefulness of the collection. A user might wish to find all of Whitman's marginal notes near passages of text that mention the Quaker preacher Elias Hicks, or the political group known as the Hunkers, or Dante Gabriel Rossetti, for example. In order to make this kind of targeted search possible, the first automation component that we will implement will be a system that matches the base texts in our images with public domain nineteenth-century texts from Google Books and other sources (including the Networked Infrastructure for Nineteenth-Century Electronic Scholarship; the Making of America

project; the National Digital Newspaper Program; and the Our Americas Archive Partnership, for example). This will often be possible with even a very low-quality OCR pass over our images. Given a page of a story from Volume 26 of *Graham's Magazine*, for example, only a handful of words would need to be correctly recognized in order for our system to link the page with high confidence to the version of the text that is available from Google Books. The coverage of nineteenth-century texts in Google Books is improving every day, and these texts are available without restriction, so our system will be able to identify a large percentage of the books and periodicals in Whitman's library—and more importantly to link them to searchable, high-quality text—with minimal effort.

The second component of our automation toolkit will be a system for performing our own high-quality OCR to create searchable base texts for the pages that are not recognized by the first component as being present in our public domain database. This is a larger undertaking, but we have used Google's open source Tesseract OCR software successfully on similar texts in previous projects, and we have developed a pipeline for efficient batch preprocessing that will easily scale to ten of thousands of documents.

The final automation component that we will implement will perform marginalia box recognition. While offline handwriting recognition is tremendously difficult, simply identifying regions of handwritten text is much easier. In our preliminary experiments we have been able to train Tesseract to locate bounding boxes around handwritten words with reasonable accuracy. We will be able to use this approach in conjunction with basic machine learning techniques to create an initial list of annotation-requiring regions for our team of human annotators.

#### *TEI Markup Levels*

We will use the current *Whitman Archive* encoding guidelines to encode a version of our transcriptions in TEI-compliant XML markup. This includes standard MARC metadata about each object, as well as the assigning of a unique identifier within the Whitman Archive (the "WorkID") to each document. The WorkID is the key referential linking mechanism within the *Archive*; once assigned, the transcriptions will be linkable to any other XML document in the *Archive*. Under the Digital Humanities Start-Up Grant, we created a small number of extensions to the current *Whitman Archive* DTD (Document Type Definition) and honed several of the uses of attributes to suit the peculiar relationship between space and textual content of these documents (described in the white paper linked from Appendix D). While the manuscript encoding captures a range of textual features such as strikeouts, insertions, erasures, and so forth, much of the textual content of the annotations documents is printed text—the original text on which Whitman made comments. The prototype annotation tool developed under the Start-Up Grant uses elements from these extensions but does not currently support the full *Whitman Archive* DTD. Subsequent versions of the tool will allow more general annotation. As described above, much of the capture of printed text will be automated, and markup of printed text will be minimal (indicating source information, heads, paragraphs, block quotes, and illustrations) and often automated. Brett Barney will continue to consult on XML development to ensure best markup practices, TEI compliance, and alignment and integration with *Whitman Archive* markup and interface design standards.

#### *Documentation*

We will use project management tools such as Mercurial and Trac to create a workflow that is largely self-documenting. All revisions to all files associated with the project will be tracked in a

version control repository, and project tasks and milestones will be managed in an issue tracker. This project data will be backed up in a timely and secure manner. We will also provide XML schemas for our custom document formats, and will release all software that we develop under the Apache License, Version 2.0. As stated above, all images and TEI-encoded files will be freely downloadable by users.

### **Work Plan**

This work plan is based on our experience digitizing a dozen of the more difficult documents from Duke's collection under the NEH Start-Up Grant, tracking the amount of time for transcription, encoding, editing, and process tracking. We have also spent the summer of 2010 developing a workflow that combines OCR recognition of printed text-blocks with human transcription of handwritten annotation. We now have a workflow, version control system, encoding guidelines, and transcription interface in place for the treatment of annotations, increasing processing efficiency. That said, these are time-intensive documents to encode, even under the pragmatic approach of the *Whitman Archive*. Each page of text takes from three to six hours to process, from transcription to final proofing. We estimate a workload equivalent to 800 pages of printed text with handwritten annotations; at an average of four hours per page, the transcription and markup labor alone will occupy 80 weeks' worth of work at 40 hours per week.

#### *Year One (5/1/2011-4/30/2012)*

Ongoing: Begin transcription and encoding of extant scans (Gray, Brown, Cohen)  
 May: Presentation of project at symposium, "The Digital and the Human(ities)"  
 May-June: Travel to LoC for imaging (Gray)  
 May-January: Complete scanning of all documents (performed by respective archives)  
 May-June: Reading database interface testing (Brown, Barney)  
 July-October: Automated upload of extant records, automated extraction algorithm developed (Brown)  
 November-February: Finish tweaking of user and transcription interface (Brown)  
 July; December-January: Evaluation (board; *Whitman Archive* annual meeting)

#### *Year Two (5/1/2012-4/30/2013)*

Ongoing: Continue transcription and encoding of scans (Brown, Gray)  
 Ongoing: Proofing of transcriptions (Gray, Cohen)  
 March, May, June/July: Conference presentations (Society for Textual Scholarship [STS]; American Literature Association [ALA]; Digital Humanities [DH])  
 August-October: Tweak integration with *Archive* resources (Barney)  
 May-August: Begin search and interface implementation; testing by *Whitman Archive* staff and advisory boards of both annotations project and larger *WWA*, and by collaborating libraries and archives (Brown, Barney)  
 November-April: Develop download formats, CC license stamping, and other interoperability mechanisms  
 July; December-January: Evaluation (board; *Whitman Archive* annual meeting)

#### *Year Three (5/1/2013-4/30/2014)*

May-August: Finish transcription and encoding of scans (Brown, Gray, Cohen)  
 March: Finish proofing of transcriptions (Gray, Cohen)

May-September: Complete adjustments to interface and search; publish interface at *Whitman Archive* (Brown, Barney)

January-April: Develop article for refereed journal

March, May, June: Conference presentations (STS; ALA; DH)

March-April: Final evaluation, including project white paper

## Staff

Project director Matt Cohen is a contributing editor at the *Walt Whitman Archive* and Associate Professor of English at the University of Texas at Austin. He has worked at the *Archive* since 1995. In 2010 he will complete a major editorial project, the digitization of the nine-volume biography of Whitman, Horace Traubel's *With Walt Whitman in Camden* (<http://whitmanarchive.unl.edu/disciples/index.html>). In February 2007, with Rachel Price, he published a digital edition of the first full-length translation of Whitman into Spanish, A. A. Vasseur's *Poemas: Walt Whitman* (<http://whitmanarchive.unl.edu/published/index.html>). In 2005, he published an edition of letters by the author of *Tarzan*, titled *Brother Men: The Correspondence of Edgar Rice Burroughs and Herbert T. Weston* (Duke University Press). He is the author of *The Networked Wilderness: Communicating in Early New England* (University of Minnesota Press, 2010) and many journal articles. In 2010-2011 he will serve as co-director (with Lars Hinrichs) of the Texas Institute for Literary and Textual Studies, whose theme for the year is "The Digital and the Human(ities)." In the budget, Cohen's salary and fringe (25% of time for three years of the project) are contributed by UT-Austin.

Travis Brown, Assistant Project Director and Developer, is a graduate student in the Department of English at UT-Austin. He was the lead developer for the NEH Start-Up Grant funded project eComma (<http://ecomma.cwrl.utexas.edu/e392k/>); he joined the *Whitman Archive* staff in 2009 as manager for projects at the University of Texas. In the budget, Brown's salary, tuition, and fringe (5 semesters and three summers) are contributed by UT-Austin.

Nicole Gray, Assistant Project Director and Coordinator, is a graduate student in the Department of English at UT-Austin. Her research is on nineteenth-century American literature and emphasizes the complex relationship between bibliographic methodologies and race. She joined the *Whitman Archive* staff in spring 2010, researching the holdings of Whitman annotations. She also has experience in management in the private sector that will be valuable in her role as project coordinator. In the budget, Gray's salary, tuition, and fringe are included (five semesters and three summers) as NEH-requested funds.

Brett Barney, consultant, has been Research Assistant Professor at the Center for Digital Research in the Humanities/Electronic Text Center at UNL since 2003 and is a specialist in literature. He is a contributing editor at the *Whitman Archive* and largely responsible for the XML architecture of the entire project. Barney helped create the unified finding aid of Whitman's manuscripts, a path breaking, IMLS-funded project that unified EAD records of Whitman's poetry manuscripts from disparate repositories and made digital editions of manuscripts available via the unified aid. In the budget, Barney's time is included (\$500/day, three days per year of the grant) as NEH-requested funds.

## Evaluation and Dissemination

The primary means of dissemination will be free public online access at the *Walt Whitman Archive*. The Archive sees over 20,000 visits weekly during the academic year, from all over the globe. As documents are made available, the results of the project will periodically be

disseminated on the *Whitman Archive*'s email list, on the TEI email list, and at other scholarly editing conferences such as the Society for Textual Scholarship, the Modern Language Association, the Digital Humanities, and the American Literature Association annual meetings, both to raise awareness about the resource and to solicit feedback on its functionality. The project will be previewed and promoted to audiences at "The Digital and the Human(ities)," a national symposium on the digital humanities to be hosted at UT-Austin in May 2011. Evaluation of results and workflows will be pursued during the annual two-and-a-half day meeting (customarily in July) with the larger *Whitman Archive* staff and consultants at the University of Nebraska at Lincoln. All project participants and implementers of our software, such as the Melville's Marginalia Online project, will be asked to reflect in writing on the progress of the project as a whole and to suggest directions for future development.

An advisory board for the Whitman's annotations project has been established (see List of Participants). Heather Jackson, Steven Olsen-Smith, and William Sherman are leading textual scholars who possess theoretical, bibliographic, and methodological expertise on marginal annotations in a range of historical periods and locations. Michael Winship is a renowned bibliographer and historian of nineteenth-century American literature; and Terry Catapano is a sought-after consultant on the technological infrastructure--metadata, project architecture, interface and search design--of electronic humanities archival projects. This board will provide ongoing guidance and annual evaluation of the work described in the Work Plan. Reports will be delivered to the board by project staff in December of each funded year, and board members will be consulted regularly as the work goes forward.

### **History of Grants**

This project received a Digital Humanities Start-Up Grant (HD-50236-07, 2007-08) in 2007-08 (through Duke University). Travis Brown's continued work on the project, including towards the development of this proposal, was funded by the Department of English at the University of Texas at Austin.

This project is also affiliated with the *Walt Whitman Archive*, which is based at the University of Nebraska-Lincoln. During the last thirteen years, the *Whitman Archive* has received a series of grants from federal agencies, including the U.S. Department of Education (1997-2000), the National Endowment for the Humanities (Collaborative Research 2000-2003; Preservation and Access 2003-2005; We the People Challenge Grant 2005-2009), and the Institute for Museum and Library Services (2002-2004 and 2005-2007), and the National Historical Publications and Records Commission (2008-2009 and 2010-2011).

## LIST OF PARTICIPANTS

### *Project staff:*

Barney, Brett. XML Consultant, Research Assistant Professor, University of Nebraska-Lincoln Libraries

Brown, Travis. Assistant Project Director and Developer, University of Texas at Austin

Cohen, Matt. Project Director, Associate Professor of English, University of Texas at Austin

Gray, Nicole. Assistant Project Director and Coordinator, University of Texas at Austin

### *Advisory Board:*

Catapano, Terence. Special Collections Analyst/Librarian, Libraries Digital Program, Columbia University

Jackson, Heather J. Professor of English, University of Toronto

Olsen-Smith, Steven. Associate Professor of English, Boise State University

Sherman, William H. Professor of English, University of York

Winship, Michael. Professor of English, University of Texas

### *Walt Whitman Archive Directors and Advisory Board Members:*

Ayers, Edward L. University of Richmond

Flanders, Julia. Brown University

Folsom, Ed, co-director. University of Iowa

Gross, Robert A. University of Connecticut

Grünzweig, Walter. University of Dortmund

McGann, Jerome. University of Virginia

Myerson, Joel. University of South Carolina

Pitti, Daniel. University of Virginia

Price, Kenneth M., co-director. University of Nebraska-Lincoln

Reynolds, David S. City University of New York

Smith, Martha Nell. University of Maryland

Unsworth, John. University of Illinois, Urbana-Champaign

APPENDIX B: *Works cited and related sources*

- Coleridge, Samuel Taylor. *The Collected Works of Samuel Taylor Coleridge*. Ed. George Whalley, et al. 16 vols. Princeton: Princeton University Press, 1969-present. Vol 12, parts 1-6.
- Dimock, Wai Chee. *Through Other Continents: American Literature across Deep Time*. Princeton: Princeton University Press, 2008.
- Finkel, William L. "Walt Whitman's Manuscript Notes on Oratory." *American Literature* 22 (March 1950): 29-53.
- Frey, Ellen Frances. *Catalogue of the Whitman Collection in the Duke University Library*. Durham, NC: Duke University Library, 1945.
- Furness, Clifton. *Walt Whitman's Workshop*. Cambridge, MA: Harvard University Press, 1928.
- Glicksberg, Charles. *Walt Whitman and the Civil War*. Philadelphia: University of Pennsylvania Press, 1933.
- Grier, Edward, ed. *Notebooks and Unpublished Prose Manuscripts*. 6 vols. *Collected Writings of Walt Whitman*. Ed. Gay Wilson Allen and Sculley Bradley. New York: New York University Press, 1961-1984.
- Hendrick, George. "Whitman's Copy of the Bhagavad-Gita." *Walt Whitman Review* 5 (March 1959), 12-14.
- Klammer, Martin. *Whitman, Slavery, and the Emergence of Leaves of Grass*. University Park: Pennsylvania State University Press, 1995.
- Jackson, H.J. "John Adams's Marginalia, Then and Now." Unpublished conference paper. 2009. Web. <[http://www.adamsjefferson.com/papers/adams\\_marginalia\\_jackson.pdf](http://www.adamsjefferson.com/papers/adams_marginalia_jackson.pdf)>
- \_\_\_\_\_. *Marginalia: Readers Writing in Books*. New Haven: Yale University Press, 2001.
- \_\_\_\_\_. *Romantic Readers: The Evidence of Marginalia*. New Haven: Yale University Press, 2005.
- LeMaster, J.R. and Donald D. Kummings, eds. *Walt Whitman: An Encyclopedia*. New York: Garland Publishing, 1998.
- Lipking, Lawrence. "The Marginal Gloss." *Critical Inquiry* 3.4. Summer 1977. 609-655.
- Melville's Marginalia Online. Ed. Steven Olsen-Smith. Web. <<http://www.boisestate.edu/melville/>>

- Miller, Matt. "Composing the First *Leaves of Grass*: How Whitman Used His Early Notebooks." *Book History* 10 (2007): 103-129.
- Myerson, Joel, ed. *The Walt Whitman Archive*. 2 vols. New York: Garland, 1993.
- Price, Kenneth M. *Whitman and Tradition: The Poet in His Century*. New Haven: Yale University Press, 1990.
- Reynolds, David. *Walt Whitman's America: A Cultural Biography*. New York: Vintage, 1996.
- Sherman, William. *Used Books: Marking Readers in Renaissance England*. Philadelphia: University of Pennsylvania Press, 2007.
- Stovall, Floyd. *The Foreground of Leaves of Grass*. Charlottesville: University of Virginia Press, 1974.
- \_\_\_\_\_. "Whitman's Knowledge of Shakespeare." *Studies in Philology* 49.4 (Oct 1952): 643-669.
- The Thomas Biggs Harned Collection of the Papers of Walt Whitman. Michael McElderry, prep. Rev. ed. Washington, D.C.: Library of Congress, Manuscript Division, 2002.  
<<http://hdl.loc.gov/loc/mss/eadmss.ms002007>>
- Unpublished finding aid, Charles E. Feinberg Whitman collection. Washington D.C.: Library of Congress, 1986.
- Van Egmond, Peter. "Bryn Mawr College Library Holdings of Whitman Books." *Walt Whitman Review* 20 (June 1974): 41-50.
- Westerfield, Hargis. "Walt Whitman's Reading." PhD Diss. University of Indiana. 1949.
- Whitman, Walt. *The Complete Writings of Walt Whitman*. Richard Maurice Bucke, et al, eds. 10 vols. New York & London: G.P. Putnam's Sons, 1902.

APPENDIX D: *Links to demonstrations and standards*

Annotations markup and display tool:

<http://camden.cwrl.utexas.edu:8000/markup>

*Whitman Archive* manuscript encoding guidelines:

[http://whitmanarchive.org/mediawiki/index.php/Whitman\\_Encoding\\_Guidelines](http://whitmanarchive.org/mediawiki/index.php/Whitman_Encoding_Guidelines)

*Whitman Archive* editorial standards description:

<http://whitmanarchive.org/about/editorial.html>

*Whitman Archive* conditions of use:

<http://whitmanarchive.org/about/conditions.html>

NEH Start-Up Grant white paper, "Interface Development for Static Multimedia Documents":

<http://www.neh.gov/ODH/Default.aspx?tabid=111&id=8>