

# Melville, Revision, and Collaborative Editing: Toward a Critical Archive

John Bryant, Hofstra University

## Abstract

Although forgotten in the late nineteenth century, Herman Melville has (since the mid-twentieth century) become one of the nation's and the world's most influential writers. Rediscovered in the 1920s, his reputation and impact on modernism and postmodernism gained full recognition in the 1940s and 1950s. Supplanting Irving, Longfellow, and even Hawthorne, he has become America's representative writer: intellectually engaged, politically diverse, aesthetically daring, and economically vexed. Melville has not only survived the re-canonization of recent decades, but his work also continues to help us articulate problems of identity and culture, and it has inspired successive generations of mainstream and minority writers and readers alike. But despite Melville's enduring presence in our culture and world literature, no comprehensive electronic archive devoted to Melville has yet been launched. Designed to address this serious deficiency, the Melville Electronic Library (MEL), now in its initial planning stages, will provide scholars, critics, instructors, students and general readers with unprecedented access to Melville's standard works. In recent decades, we have also come to know various "new Melvilles" associated with less familiar materials. The heavily revised *Typee* manuscript (or any of Melville's many working-draft manuscripts) reveals the writer to be a relentless reviser; annually discovered annotated volumes from Melville's dispersed library indicate a voracious reader studying to be a poet; and Melville's recently catalogued print collection discloses him to be an avid collector of fine art. But the evidence of these new versions of Melville is lodged in unique, virtually inaccessible, heretofore irreproducible materials. If funded, MEL would provide the tools for accessing these new Melvilles: the reviser, reader, poet, collector.

In its fullest form, MEL would be a "critical archive." To be sure, like other electronic archives, it would contain digital reproductions of all relevant primary and secondary sources—manuscripts, print texts, source works, Melville's marginalia, biographical materials, a gallery of Melville's print collection, adaptations of Melville works, and a bibliography of and selections from the critical response to Melville—located in easily navigated, hyperlinked "library rooms." But the archive would also be "critical" in that editors and students would be able to interact collaboratively in building the archive, creating transcriptions and editions, engaging students in classroom projects, and exploring the interpretive consequences of the archive's content. With these ends in mind, this proposal focuses on the new Melville revealed to us in manuscript: Melville as reviser and the sequential versions ("fluid texts") of three sample works as he revised them: "The Town-Ho's Story" (*Moby-Dick*), the poem "Monody," and selections from *Billy Budd*.

A NEH Digital Start-Up grant (Level I) would provide funding to initiate the general organization of MEL and the creation of a feature called TextLab to enable collaborative "fluid text" editing. More specifically, I would be able to consult with Melville scholars to establish a general workflow schedule and agree upon assignments for handling content in the assembling of MEL. And it would also enable me to work with Hofstra's Faculty Computing department to adapt open-source software (like Batik and Subversion) that draw upon open-standards (TEI, XML, and SVG) to create an innovative application allowing individuals in an editorial or classroom group to download images of Melville manuscripts and print texts, mark the revision sites directly on each image, transcribe the texts and sites in an expansion of XML that links image and transcription, and link both image and transcription to annotated revision narratives, either to create critical editions of fluid texts or to conduct pedagogical workshops for the study of manuscripts and print texts in revision. The idea is to create a model for manuscript and revision transcription as well as collaborative editing that would be readily adapted to the works of any author.

ABSTRACT .....	1
INTRODUCTION: EDITING AND THE CRITICAL ARCHIVE.....	1
<i>Using the Textual Apparatus</i> .....	<b>Error! Bookmark not defined.</b>
<i>The Case of the Disappearing Apparatus</i> .....	<b>Error! Bookmark not defined.</b>
MEL AS CRITICAL ARCHIVE: ACCESS, DISCOURSE, AND ANXIETY .....	4
<i>TextLab and Collaborative Editing: The Example of Typee</i> .....	5
HISTORY OF PROJECT .....	8
STAFF .....	10
<i>MEL Project Staff</i> .....	10
<i>Advisory Board</i> .....	10
<i>Editorial consultants</i> .....	<b>Error! Bookmark not defined.</b>
<i>Technical Consultants</i> .....	<b>Error! Bookmark not defined.</b>
METHODS .....	10
<i>A Proposal for Starting-up</i> .....	10
<i>Software Development</i> .....	12
FINAL DISSEMINATION .....	18
WORK PLAN .....	18
BUDGET.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>

## ***Introduction: Editing and the Critical Archive***

What I call a “critical archive” advances beyond the basic notion of a digital storage place. While containing a full complement of primary and secondary texts and images, it is also a place for editing a text or set of texts, and a forum where readers may contribute, deliberate, collaborate, and respond to the transcriptions and editions they or others have constructed. By enabling scholars in a given field or instructors and students in the classroom to collaborate in editorial projects, the critical archive can demonstrate the inherently interpretive nature of editing and archive building. If structured carefully, it would familiarize users with heretofore inaccessible texts (in particular for this proposal, revision texts on manuscript) and thereby stimulate discourse that would perpetuate the archive but also revitalize textual criticism. Essential for such interaction in an online critical archive is the inclusion of software tools and applications that would enable users to compare texts and images together, transcribe versions, collate them, and construct reliable texts for their own critical ends. A Digital Start-Up Grant (Level I) would initiate the planning for a critical archive called the Melville Electronic Library (MEL) and fund a year’s work on the creation of TextLab, a tool for the collaborative editing of manuscripts.

One print model for a critical archive might be textual scholarship’s most comprehensive editorial genre, the “eclectic critical edition” promoted in the Greg-Bowers tradition of scholarly editing. The Northwestern-Newberry (NN) edition of *The Writings of Herman Melville* is an exemplar of this effective but controversial approach. Each volume in the series is essentially a print archive devoted to a particular Melville work. For instance, in editing *Moby-Dick*, the NN editors collated the American and British editions, studied the variants, established the American version as copy text, and emended that text in certain places either by substituting British for American wordings or by creating their own substitute wordings. Thus, the NN edition of *Moby-Dick* is “eclectic” because it combines American, British, and modern variants into one “reading text” with a massive “textual apparatus” appended to it, containing in archival form all relevant textual data.

The rationale for this editorial approach is that editing is essentially an interpretive act because editors cannot avoid making critical judgments about the words they will include in their editions. Accordingly, as a representation of the modern editors' conception of the author's final intentions, the edition's reading text itself is necessarily a kind of critical interpretation. In the interests of transparency, then, the critical edition's textual apparatus grounds the reading text in its historical context by giving readers a certain albeit coded access to the work's various historical versions. But as many have argued, the reading text of *Moby-Dick* is a version that Melville never witnessed. Indeed, in mixing versions, the "eclectic" form of critical editing muddies rather than clarifies Melville's shifting intentions, creative process, and interactions with readers. Moreover, because the edition's reading text is not directly linked (by footnoting or on-the-page annotation) to the edition's apparatus, readers become further distanced from the editors' critical discourse. As we shall see, the now-standard, eclectic critical edition approach is finally not a suitable model for exploring the broader potentials of the online critical archive.

In its defense, the Greg-Bowers approach to critical editing is designed with admirable safeguards. As a highly coded textual archive, each edition's textual apparatus serves as a repository of all textual variants so that readers with differing conceptions of intentionality can explore the apparatus and in effect reconstruct, let's say, the two historical editions of *Moby-Dick*—the American and British—or create their own critical edition of that masterwork. However, my own experience in navigating the apparatus to create the Longman Critical Edition of *Moby-Dick* (2006) proved inhibiting. I used the NN textual apparatus to generate a reliable text of the American version of *Moby-Dick*. That is, I set out to "de-eclecticize" the NN edition's clear reading text in order to create a base version that simulates Melville's original text in the first American edition. Of course, I also made several direct inspections of both the American and British edition to double-check my work. The process, which took five years, confirmed the reliability of the NN collations but also affirmed my belief that the NN apparatus, indeed any textual apparatus, is highly cumbersome. But is it not the lot of textual scholars to confront arduous work so that critics, students, and general readers need not bother with textual variants?

Why would anyone want to pursue textual variants? In fact, a wealth of information exists in the variants of *Moby-Dick* (and all of Melville's works). Some, of course, are merely typos, but most are revisions Melville made or a British editor imposed, and many of the latter are sizable expurgations of sexual, political, and religious content. If readers want to know more about how Melville revised himself, how British editors censored him, how Melville's culture received his work, and even how modern editors have altered Melville's text, they need access to these variants, on the reading page, so that they may study them, while reading, to consider who made them, how, and why. These variants are the physical evidence of how Melville and his editors negotiated one of world literature's greatest novels into existence. However, in the NN critical edition of *Moby-Dick*, that evidence is segregated from the edition's reading text, which has no intra-textual referencing to the apparatus to alert readers, while they are reading, to the existence of textual variants. The reading experience gives the impression, to those otherwise unaware, that the NN *Moby-Dick* is a fixed and stable text when, in fact, *Moby-Dick*, and all of Melville's writing, and the writing of all writers, is a variant and fluid text continually modified by authors, editors, and scholars. And because commercial publishers reprint the clear reading text only, and drop (or severely reduce and further codify beyond readability) the textual apparatus, most readers of *Moby-Dick* today have little or no access to the NN edition's validating "archive" of variants or any evidence that the text itself is a modern, conceptualized construct.

To be sure, Greg-Bowers notion of a critical edition as a print archive anticipates the idea of an online "critical archive." And from our nascent online perspective, we can readily see how the spirit if not limitations of Greg-Bowers might be more fully achieved in a critical archive like the Melville Electronic Library. MEL would be designed to give users unprecedented access to all historical and modernized versions of *Moby-Dick* but also a better means of navigating the revisions inherent in those versions. In doing so, the critical archive allows us to fulfill the general Greg-Bowers goal of underscoring the necessarily interpretive nature of critical editing. If the genre of the "eclectic edition" has faded, today's editorial goal of analyzing variants to clarify the stages of an author's fluid text,

revision practices, and shifting (not fixed) intentions will surely be more fully realized through online critical archives such as the proposed MEL.

### ***MEL as Critical Archive: Access, Discourse, and Anxiety***

If MEL is to become an effective critical archive, it must confront the problematic division of textual scholars and literary critics into separate spheres. The tradition of textual scholarship has been to establish texts so that critics will have reliable texts to interpret. But an unintended consequence has been the false assumption, among scholars and critics alike, that editing is not interpretive and that the interpretation of texts is independent of editing. By failing to acknowledge the interpretive dimensions of scholarly editing, critics also lose sight of the material indeterminacy and instability of texts manifest in any editorial project, which ironically would amplify their own critical endeavors.

One problem then is how to construct a critical archive that breaks down the barriers between textual scholarship and other fields of critical discourses. One sure way to de-mystify the editorial process is to provide fuller access to the very discourse that editors perform. If carefully structured, an online critical archive can exceed the print critical edition in two ways. Obviously, it can provide a fuller gallery of editorial materials—the variants but also full transcriptions and images of the variant books—but, more importantly, it can also permit users to perform editorial acts themselves. By performing such acts, they will necessarily confront variants and therefore certain inevitable problems of interpretation: what is the meaning of this variant or that? Who or what caused the variant? What rhetorical strategy is inscribed in an apparent trajectory of revision at one revision site, or in several? What combined social and cultural forces impinge upon such revision sequences? By engaging directly in such problems, users of a critical archive must also enact various forms of literary and historical interpretation. Rather than being a static apparatus, the online critical archive is, in effect, a forum for humanistic debate, one that allows discourse on textuality, creative process, revision, book history, literary history, and cultural analysis to unfold. In

such an environment, users become editors, and the misconceptions about textual criticism and the unwarranted bifurcation of scholars and critics is more likely to dissolve.

Of course, it is all well and good to argue that a critical archive is a “forum,” and that it can facilitate new scholarship and new interpretation. But the “liberation ideology” inherent in these arguments carries with it certain anxieties. After all, we expect a high degree of rigor from textual scholarship, and we rightly wonder what would prevent the creation and distribution of flatly erroneous texts online, if untrained users in a critical archive take on editorial roles?

The response to these legitimate concerns is not to limit access to (or fetishize) editorial practice but rather to extend it and manage it through repeated rounds of monitored discourse. Just as any critical edition in book form is the product of a team of scholars and students interacting on the mechanics and interpretations of textual criticism governed by project directors, so too must a critical archive provide procedures that ensure fluidity of discourse, clarity of interpretation, opportunity for disagreement, and yet mutually agreed upon and routinely enforced standards of readiness for the online display of texts. The real concern, it seems to me, is not that “bad texts” will proliferate, but rather how might we develop software and protocols to facilitate yet manage the discourse of textual criticism so that access to images of manuscripts and other materials is increased, yet under-examined editings and transcriptions of print texts or manuscripts are not mounted online. The practical problem, then, is how to structure MEL to meet these ends.

### **TextLab, Collaborative Editing, and The Example of *Typee***

As a critical archive, MEL would be divided into eight content “rooms” (see Fig. 1), and TextLab, an innovative feature for collaborative editing, equipped with technology and protocols designed to insure rigorous scholarship, reliable content, and fuller discourse on Melville’s fluid texts. TextLab will enable users to work separately and together to transcribe revised manuscripts, link them directly to manuscript images, and track their independent and collaborative editorial work. To clarify the need for such a tool,

let me offer one textual problem from Melville's *Typee* manuscript that illustrates how MEL's TextLab might address the problems of access, discourse, and interpretation.

In Ch. 13 of *Typee*, Melville's first novel, published in 1846, the protagonist Tommo mentions his abhorrence of cannibalism. His island hosts, the Typees, disavow such behavior, attributing the practice to rival tribes. To underscore their benevolence, the Typees point to "the natural loveliness of their own abode," as if a natural environment is inherently lovely, as if nature dictates character, as if benevolent people cannot turn cannibal. But the working manuscript of this passage reveals that Melville, and probably his editors, worried over what eventually became "their own abode" (see Fig. 2).

Originally, in manuscript, Melville had his Typees call their abode a "favored valley" (Fig. 3). But after crossing out those words, he substituted for them "beautiful abode" in the space above "favored valley." Still later, he deleted "beautiful" and, in the space below the already deleted "favored valley," he inscribed "paradisical," to give a final manuscript reading of "their paradisical abode." After completing his manuscript, Melville fashioned a fair copy (now unlocated), and in it, he apparently deleted "paradisical" and substituted for it simply "own," to give the final reading found in print, "their own abode." Alternatively, Melville may not have revised "paradisical"; instead, an editor or even Melville's brother Gansevoort might have made the change while copyediting or proofing Melville's text.

The revision sequence involving *favored valley* then *beautiful*, *paradisical*, and *own abode* shows how a writer's wording can oscillate in and out of different cultural, aesthetic, and religious ideologies during the creative and revisionary processes. These oscillations also provide a pathway for mapping out a revision strategy revealing a shift in Melville's rhetorical intentions, which in turn suggests how the writer may or may not be negotiating with British editors (and hence a British readership) over deeply structured anxieties involving, oddly enough, cannibalism and urban versus pre-industrial life. Of course, anyone reading *Typee* in a single fixed print version has access only to his final wording "own abode," and no hint that this innocuous visible phrase is a textual hotspot concealing several otherwise "invisible texts of revision." Such a textual and interpretive goldmine begs for electronic editing.

In 2006, through the University of Virginia Press's Rotunda series, I launched "Herman Melville's *Typee*: A Fluid-Text Edition" (Fig. 4), which enables the user to navigate Melville's manuscript in various ways. Briefly, the site allows users to compare any two images or versions of the manuscript (Fig. 5). For our purposes, we might compare a manuscript page to my transcription of it (Fig. 6) or to a "base version" of the final reading of that manuscript (Fig. 7). I also use the base version as a textual terrain for mapping all 1100 revision sites (Fig. 8) of the *Typee* manuscript. Each highlighted site on the base version is linked to a "Revision Sequence" that enumerates the steps Melville took to revise his text at the given site (Fig. 9). And attached to each sequence is a "Revision Narrative" that tells the story of how Melville or his editors revised, step by step (Fig. 10).

However, as versatile as Rotunda's *Typee* manuscript site is, it does not allow users to "click" directly from a manuscript image to the transcription; nor does it include TextLab. That is, it represents my own editorial initiative, and though I invite others to offer alternative readings and editing, the site provides no way for others to interact, agree, disagree, or collaborate. The creation of TextLab would take fluid-text editing to a new dimension. It will feature a version of XML (tentatively called AtlasXML) developed at Hofstra that leverages the SVG standard to allow precise specification of locations within any manuscript page. It will also feature an enhanced subset of the TEI standard, which will link to AtlasXML and therefore to the specifically marked regions within the manuscript image. TextLab will reside within a publicly-available Subversion repository, which will allow users to work collaboratively to transcribe revised manuscripts and to maintain detailed metadata tracking their collaborative editorial work.

Working with TextLab, users would download the image of a manuscript leaf, use a mouse to mark a revision site directly on the image, and then transcribe the text of the leaf, while encoding its marked revision locations. (Such locations can vary in size and content from a half-inscribed letter or a marginal doodle to multiple words or lines.) Users will also be able to annotate the transcription, providing a Revision Sequence that accounts for the revisions occurring at those sites, and constructing a linked Revision Narrative explaining the sequencing. Because TextLab resides in a Subversion

repository, multiple users in collaboration can track their editing in order to discuss the textual regions marked on the manuscript image, as well as their transcriptions, revision sequences, and narratives before and after mounting any content online.

If fully realized, TextLab would be an excellent tool for students and teachers to examine the practical and interpretive dimensions of Melville's writing process. With it, users could inspect the "paradisical abode" revision site in *Typee*, generate plausible variant transcriptions, consider the cultural and rhetorical strategies implicit at that site, and forge (together with other Melville scholars) an explanatory narrative, or alternatives to it. In this way, TextLab would be a powerful means of implementing the discourse forum so vital to an effective critical archive.

But TextLab's importance goes beyond the study of Melville. For centuries, manuscript study in general has been seriously impeded because manuscripts are famously inaccessible and difficult to read; scholarship has been limited to a small handful of experts, and publication of their findings has been hampered by book production costs; thus, critics and students rarely get a glimpse at the "invisible" content of revision embedded in a manuscript leaf. However, by giving critics and students easy protocols for the editorial process, we begin to familiarize the otherwise alienated discourse of manuscript study and revision, and thereby open up new and vibrant areas of interpretation for our citizens. As a model for collaborative editing, TextLab can be used in the transcription, analysis, editing, and reading of the manuscripts and print texts of any writer, not just Melville.

## ***History of Project***

The idea of creating the Melville Electronic Library began with John Bryant and Haskell Springer in the late 1990s. At that time, they consulted on the feasibility of MEL and fluid-text editing with John Unsworth, then director of Virginia's Institute for Advancement of Technology in the Humanities (IATH), IATH's Daniel Pitti, and Virginia's Jerome McGann. As IATH Associates, Bryant and Springer wrote two NEH grant proposals for Collaborative Research to initiate MEL, in 1999 and 2000; both were

unsuccessful. Bryant then turned to developing *Herman Melville's Typee: A Fluid-Text Edition* (2006), with Rotunda editor David Sewell and programmer Timothy Finney. At the same time, Bryant and Springer collaborated on the equally innovative fluid-text print edition of *Moby-Dick* (2007) for the Longman Critical Editions series. Both efforts improve upon Bryant's earlier editorial strategies for giving readers fuller access and deeper understanding of revision and creative process. Springer, now in retirement, continues his connection to the MEL project as an editorial associate.

As Americanist co-editor of NINES (Networked Infrastructure for Nineteenth-century Electronic Scholarship), Bryant began consulting with McGann, Nowviskie, NINES associate director Laura Mandel (Miami of Ohio), and Julia Flanders (Brown) on strategizing specific goals and work plans for MEL, delving into technical requirements and training, and developing initial drafts of the present proposal. In 2006-07, he established a working relationship with Hofstra's director of Faculty Computing Services (HFCS), Judith Tabron, who will be setting aside her time and that of two staff members (Alex Smiros and Adeel Raja), as well as server space for MEL. Bryant and the staff have also coordinated with systems engineer Marshall Flax, who has assisted the team in formulating a project plan that utilizes current open-source technology and open-standards to create TextLab. And in April, 2007, Bryant and Leviathan's associate editor Wyn Kelley, met with Houghton's manuscript curator Leslie Morris and researcher Dennis Marnon to arrange for the digitization of the Melville manuscripts and print items that will serve as model images and texts for the proof of concept version proposed in the present grant.

In 2007-08, the year before the proposed grant year, Bryant, Tabron, Smiros, and Raja will familiarize themselves with TEI and XML as well as an XML editor (probably oXygen). In January, 2008, Bryant and one staff member will attend the XML and TEI workshop scheduled at SUNY-Buffalo; he will also attend the NINES workshop that summer at the Miami University of Ohio. With Hofstra's Faculty Computing staff and selected Hofstra graduate students, Bryant will begin encoding some of the featured Melville texts. At the same, Bryant will apply for a full year's sabbatical for 2008-09, to coincide with the proposed NEH grant year. The sabbatical will give Bryant the entire year to conduct the proposed project for implementing MEL; grant money will not be necessary for any released time.

## **Staff**

### MEL PROJECT STAFF

**John Bryant**, Project Director, English Dept., Hofstra University; **Judith Tabron**, Project Manager, Hofstra Faculty Computing Services; **Alex Smiros**, Project Assistant, HFCS; **Adeel Raja**, Project Assistant, HFCS; and **Marshall Flax**, Programmer, Tullett Liberty.

As Project Director, John Bryant conceived and planned MEL; he will be on sabbatical leave for the entire 2008-2009 academic year and will be able to give 100% of his work time to the project. As Project Manager, Judith Tabron will devote 50 hours of her time as director of Hofstra's Faculty Computing Services to the year-long project. Project Assistants, Alex Smiros and Adeel Raja (both instructional technologists in Hofstra's FCS) will devote a total of 150 hours to web-design and other duties. Marshall Flax, a Systems Engineer for a New York area firm, will be the primary programmer for the project, devoting 100 hours to the project throughout the grant year.

**ADVISORY BOARD Melville Consultants:** Mary K. Bercaw Edwards (Connecticut), Dennis Berthold (Texas A&M), Patrick Bryant (Ind. Scholar), Wyn Kelley (MIT), Dennis Marnon (Houghton Library), Peter Norberg (St. Joseph), Steven Olsen-Smith (Boise State), Haskell Springer (Kansas), Christopher Sten (George Washington), Robert K. Wallace (Northern Kentucky), and John Wenke (Salisbury State).

**Editorial Consultants:** Joel Myerson (South Carolina), Ken Price (Nebraska), Stephen Railton (Virginia), Martha Nell Smith (Maryland), G. Thomas Tanselle (Guggenheim, emeritus), and Marta Werner, (D'Youville). **Technical Consultants:** Laura Mandell (Miami of Ohio), Jerome J. McGann (Virginia), Bethany Nowviskie (Virginia), and David Pitti (Virginia).

## **Methods**

### **A Proposal for Starting-up**

In general, the goal of MEL is to create a critical archive that gives open access to Melville texts and images, and which, in its most innovative proposed feature (TextLab), provides adaptations of open

source software tools to facilitate collaborative editing and discourse on manuscripts and the process of revision. We recognize that however reasonable, innovative, and fetching an idea MEL may be, it is a long-term venture, and cannot come into existence without external funding and the commitment of various people with various skills at Hofstra and elsewhere. While MEL will not “happen” overnight, it is ready to be started.

The proposed Level I Start-Up grant will help fund two endeavors: the structuring among Melville scholars of the workload involved in gathering the content of MEL, and the more intense process of designing and creating TextLab as model feature for MEL. In particular it will provide 1) travel funds to bring Melville scholars to Hofstra for a planning and scheduling meeting to organize and delegate responsibilities in the future development of MEL so that the project director may prepare a grant proposal to fund the entire MEL project, and 2) funding for a Java programmer (Marshall Flax) who will develop a proof of concept version of the collaborative fluid-text editing feature of MEL’s TextLab.

Already organized are a group of Melville scholars and editorial consultants who will in future years provide content for MEL and a group of skilled and committed HCFS technical associates who will assist in building TextLab. The proof of concept version site will use digital images of three Melville works: “The Town-Ho’s Story” (Ch. 54 of *Moby-Dick*), which exists in variant magazine, American, British, and NN edition versions; “Monody,” a poem on two attached manuscript leaves composed over several years; and eight manuscript leaves from *Billy Budd* that exhibit both prose and poetry. Harvard’s Houghton Library, which owns these and almost all of Melville’s late manuscripts, has agreed to supply high resolution digital images of these sample works for storage in Hofstra’s server. These three sets of manuscripts and print texts have been chosen because they represent the kinds of texts (prose, poem, prose and poem) that exhibit most of the structuring concerns inherent in Melville’s oeuvre.

As part of the organizational phase of the project, Bryant will conduct (on Hofstra’s server) a listserv for Melville scholars to set up preliminaries, in advance of the on-site meeting at Hofstra conducted in the conference rooms of HCFS and Hofstra’s English department. Participants will receive

stipends to cover part of their travel expenses. Our plans for the technological development of TextLab during the proposed grant year are described in the next section.

### **TextLab Development**

Project funding will be used to leverage recent developments in web technology to build a discrete segment of TextLab as a model for transcribing the revision text of manuscript (and print) pages in a collaborative editorial environment.

The technical challenge is to develop a tool for producing fluid and live annotation directly on images of heavily revised manuscript pages (what we call “locations of interest”) and to link them to scholarly transcriptions that can also be annotated. (Since the annotation of printed pages offers no further challenges, the proposed software can be readily applied to print texts as well.) Textual scholars may use this live annotation tool to identify and discuss sites of revision, undeciphered words, or graphic marginalia within manuscripts. The annotations are fluid in that they can be added to or changed over the years without having to rebuild from scratch the image’s display technology or data. The proposed TextLab tool will build upon open-standards such as the XML standards developed by the Text Encoding Initiative (TEI) and Scalable Vector Graphics (SVG), which can use that XML set to present information layered with images of manuscript or book pages. It will also use the open-source technology of Subversion (SVN) for its version control system. In addition, a simple community system such as a blog or a wiki will be attached to the display and annotation system to allow communities of scholars and students to discuss proposed revisions and interpretations of the text. The result will be a powerful tool for textual scholarship and the study of revision in manuscripts and print text that is easy to use, extensible, reproducible, and pervasive.

Digital scholars in the 1990s attempting to solve the problem of linking specific words in a manuscript image to an encoded transcription considered the idea of adding “overlays” to the digital images involving both picture and text, but they were not able to develop a method that provided

sufficient granularity and precision to identify small, irregular regions of revision. But the recent development of SVG technology helps solve some of these problems.

One problem is evident in the difference between print and web technologies. Since the advent of digital laser printers, PostScript has been used to describe the visual information of both text and graphics so that everyday computer users can scale them up or down and print out clean, smooth copies. As a language for vector graphics, PostScript can be sized or placed mathematically, but it was designed specifically for printing. In contrast, web technologies tend to rely on raster graphics, which assign colors to specific pixels on the screen. Raster graphics such as photographs may contain rich information but are not smoothly sizable, or mathematically controllable, in the same way that vector graphics are. SVG technology includes image languages that provide information on the screen that is as mathematically controllable as PostScript is for print.

Proposed originally as a competitor to the Precision Graphics Mark-up Language (PGML) standard, SVGs describe vector graphics mathematically. They are structured as XML data, which makes them flexible and means that they can contain raster images, text, and vector graphics, all combined, layered if necessary, and presented with sufficient control that annotations can be located specifically on a raster image, even down to the pixel. For our purposes, an SVG document would combine a photograph of a manuscript leaf on one layer with a second overlaid image containing user-drawn “bounding boxes” surrounding revision sites, and these boxes can be linked to machine-readable text on yet a third layer. The visual representation of all this information on all these layers can be determined as precisely as it can in print technology.

Using SVGs, we can embed high-resolution photographs of manuscript pages, then use XML to add simple annotations, including a bounding box of any shape or size to indicate a revision site within the manuscript as well as a text explaining the site. We then give that revision site a unique identifier, and add machine-readable text as an overlay to the underlying raster graphic. SVGs can be rendered in such web browsers as Firefox and Opera, and are seamlessly integrated into XHTML documents. This XML DTD, used for the naming of SVG-defined sites within the manuscript (and tentatively called

“AtlasXML”), will allow for uniform and consistent reference to manuscript locations, especially by TEI-XML documents. Because much of the information is contained in XML, SVGs can be easily edited for the web and the update is immediate and live. It need not be exported in any fixed form over and over again.

More specifically, we will generate, in addition to any necessary descriptive and administrative metadata, two types of XML documents containing information about a manuscript. The first type will be the AtlasXML, which acts to assign names to locations (hence the name “Atlas”) within the reference manuscript image. The AtlasXML document will also contain “location names” for words that have not yet been assigned a specific location within the image. These are easily encoded placeholders for the words within the manuscript that are not yet the object of textual study (typically because there is no perceived revision or undeciphered text at that site). The second type will be a TEI-XML transcription of the manuscript page. This document will make frequent reference to the AtlasXML document; indeed, it is possible for a single element within TEI-XML to refer to multiple AtlasXML locations (e.g. in the case of multi-line revisions). For this reason, AtlasXML documents are typically only appended-to, while the TEI-XML documents are far more revisable.

Here is how we propose to develop AtlasXML. We will 1) create a minimal DTD leveraging the SVG web standard (<http://www.w3.org/TR/SVG11/>) to define location identifiers of regions within the initial source image of a manuscript page (AtlasXML); 2) extend the TEI document format (<http://www.tei-c.org/P5/>) so that TEI-XML documents can refer to these AtlasXML locations; and 3) create a tool to fill in the data of AtlasXML documents.

TextLab will also allow any number of people to work on XML documents from different locations. To facilitate collaborative editing and the tracking of multi-sourced editorial emendations, we propose to create a “version control system” using industry-standard Subversion (SVN) repository software and a tool for editing this XML information, which would make it possible for a community of scholars or students in a classroom to interact with the text without having to go through graphic artists and without having to learn the art of preparing raster graphics for the web.

The project's SVN repository will not only include all original and scaled images, application and tool source code and scripts, XSLT scripts, and image metadata, but will also include all versions of all AtlasXML and TEI-XML documents. The branching-version model within Subversion allows for each user to maintain a "thread" of his/her modifications to any file and to merge any of those acceptable changes with changes made by others; it also allows for an editorial board to merge proposed changes into a "consensus" branch of changes. Integration with the SVN repository will begin with the integration of an open-source SVN client library (perhaps <http://svnkit.com/>) within the TextLab application, so that "File | Save" would be replaced by "Commit To Current repository branch" and "File | Save As" by "Commit To a New repository branch." Integration will be continued by support within TextLab for the display of differences between versions of documents within the repository and for the merging of changes from one branch into another.

During the grant year, Hofstra's technology staff will build upon Batik, an open-source Java-based toolkit (and an active project within the Apache Foundation) for manipulating SVG documents, to create an application that will display SVGs containing images of the sample manuscripts and their associated annotation data. If time permits, we will also use XSLT to convert XML to HTML for display in web browsers. If it is determined that it is not possible to create a web-browser compatible SVG display application for this purpose, a Java runtime application may be created instead, which will run locally and serve as the SVG-handling application described herein. This would contain HTML links back to the community system.

The project will be Internet-based (since all users will use a common commercially-available SVN hosting service such as <http://cvsdude.com/> or <http://www.sourceforge.net/>), but it will be only trivially web-based. While some static information (e.g. project home-page, application download link, immutable image files, and project documentation) may be presented on a web page for convenience or efficiency, in all cases the definitive copy of all artifacts will be held within the SVN repository. In some cases, as with source code, this procedure is industry-standard practice. But in the case of alternative transcriptions of a single manuscript image, the resulting "tree" of revisions and the user activity of

“merging” from one branch to another will itself be a self-documenting community enterprise of dialogue and improvement. Ordinary version-control-repository functions, such as the ability to distinguish any two versions or to associate any change to a specific user and time, have tremendous power when applied to TEI-XML documents of variant readings of a given manuscript. On the technical side, since all data persistence would be within the Subversion server, this monolithic Java application could even run within a Java WebStart (<http://www.jcp.org/en/jsr/detail?id=56>) sandbox. In addition to leveraging the SVG standard for the AtlasXML format, we will also use SVG as the basis of a Graphical User Interface (GUI) that would allow for the population and maintenance of AtlasXML documents. We plan on using (and contributing back when useful) Apache’s open-source Batik project (<http://xmlgraphics.apache.org/batik>), which provides SVG manipulation, rendering, and interaction libraries.

To store annotations consistently so that the SVG renderer can be programmed to display them appropriately, the project team proposes to add on to existing TEI standards. As previously stated, the team has already begun to investigate current TEI standards, and in advance and during the grant year we hope to define a provisional extension. As both TEI and SVG standards are broad-based community initiatives, we would expect further refinement of both would be forthcoming from their respective communities if our proof of concept works as anticipated. Informing these communities of this work and soliciting their input and help is a planned part of the project’s execution. In addition, the staff will consult with the Apache / Batik open-source development community in the development of the SVG display system. The project has the following intermediate goals:

- Rent commercial webserver to house the project.
- Continue TEI-XML training for project participants.
- Accept delivery of high-resolution photographs provided by the Houghton Library.
- Define descriptive and administrative metadata for these digital images and define a workflow model for maintaining all created documents within a revision-control system.
- Define additional TEI-type markup language as is necessary to complete the proof of concept.
- Generate machine-readable transcriptions of the texts of the manuscript pages.

- Write Batik-based SVG handler application (AtlasXML).
- Define Atlas-XML format including the kinds of data it stores. Atlas-XML files will serve as a bridge between scanned images and TEI documents by containing the coordinates of each named revision site and word location. This will allow for a consistent nomenclature for manuscript locations across all TEI documents and textual scholarship.
- Embed raster graphics of pages in AtlasXML-based SVG documents.
- Associate machine-readable words with raster-graphic representation of words
- Label revision sites with unique identifiers (readable both as AtlasXML and HTML anchors) and TEI-type markup indicating revision action at that site.

Throughout the grant year, a website will serve both as a test site for all parts of the project and as the team's primary mode of communication with each other and with the rest of the world. Not only will we be soliciting input from participants in worldwide projects like TEI and Batik, we also want those participants to be able to see the project's progress and share information. The website will serve this purpose by displaying materials in progress as well as providing e-mail or any other necessary communication among participants. Following this proof of concept project, we will seek funding to continue the project, both to acquire and markup images of the rest of Melville's manuscripts at the Houghton Library and to refine our XML DTD and the SVG-handling application that we will have initiated (AtlasXML).

Throughout the grant year and after, the project specifications and code we create will be freely available for comment, addition, or modification. HFCS will also provide access to a commercial-type webserver to sustain the project during the grant year and beyond. Maintaining its information and user accounts will be rolled into regular research computing activities for the department. The skills needed to complete the project are available in the project team, including a scholar who has produced an electronic edition of the *Typee* manuscript as well as a distinguished publishing record on Melville, an experienced Java programmer who has built systems utilized commercially in live, real-time systems for over ten years, two instructional technologists with graphic arts backgrounds and additional experience

programming for the web, and a project manager with ten years' experience managing academic computing development projects.

## ***Final Dissemination***

To survive, a critical archive such as the Melville Electronic Library must have the long-term maintenance support of its home institution, broad appeal in the academic community and general public, and interactivity with other research environments. Because Melville continues to be a popular and useful writer, nationally and internationally, we have no doubt that MEL will attract numerous users, and Hofstra has the capacity, academic standing in Melville studies, and staff to maintain the site in perpetuity. In creating TextLab and its SVG and collaborative editing features, we hope that as a critical archive, MEL will also become a model for other archives and digital editorial projects.

MEL will also be designed for optimal interoperability. Eventually, we will represent the site's text and image objects in Resource Description Framework (RDF) so that it can become a part of the NINES research environment and thereby serve a large, growing, and critically acclaimed research and teaching community. Considered a premier model for online research, NINES is a consortium of individual sites dedicated to Romantic, Victorian, and American texts of the long nineteenth century (1775-1920). Individual affiliated sites include *The Blake Archive*, *Rossetti Archive*, *The Poetess*, and *Whitman Archive*. Currently, NINES also offers several powerful software programs, in particular Juxta and Collex. Versions of Melville works, submitted in plain text, can be collated with Juxta. And with Collex, Melville texts and images may be placed beside objects from Blake or Whitman.

## ***Work Plan***

<b>Month / Year</b>	<b>Task</b>
MARCH 2008	<ul style="list-style-type: none"> <li>• Acquire manuscript reference images (JB and FCS)</li> <li>• Define metadata to be stored for each reference image and each scaled image (entire team)</li> <li>• Propose a high-level directory structure for the SVN repository</li> </ul>

	<p>into which all artifacts would be placed. (Flax)</p> <ul style="list-style-type: none"> <li>• Choose a commercial SVN hosting service. (Tabron)</li> <li>• Sign-off on SVN directory structure (entire team).</li> <li>• Configure a development environment using Apache standards and the Eclipse IDE. (Flax)</li> </ul>
MARCH & April	<ul style="list-style-type: none"> <li>• Script and automate the conversion of images to usable resolutions and the creation of metadata files for the converted files (FCS)</li> <li>• Configure SVN repository (Flax and FCS staff)</li> <li>• Investigate the availability of a web-platform for threaded discussion (Wiki or Blog) that uses SVN as a backing store and determine web hosting service (Tabron)</li> <li>• Choose an open-source SVN client library. Modify Batik's Squiggle application to use the SVN library to read and write. (Flax)</li> <li>• Define branching model to be used within SVN for collaboration. (JB)</li> </ul>
MARCH thru June	<ul style="list-style-type: none"> <li>• Extend TEI-XML to represent revision sequences. Validate by the test encoding of small sequences into an off-the-shelf XML editor such as oXygen. (JB)</li> </ul>
APRIL	<ul style="list-style-type: none"> <li>• Process manuscript reference files as they arrive. (JB and FCS)</li> <li>• Choose an open-source "graphical difference library" and integrate into Squiggle. (Flax)</li> <li>• Define the DTD for AtlasXML, leveraging SVG. (Flax)</li> </ul>
APRIL & May	<ul style="list-style-type: none"> <li>• Define a naming scheme for elements within AtlasXML. (Flax)</li> <li>• Create a small utility program to generate AtlasXML frameworks given a plain-text transcription of a page, an image of the page, and the metadata for that image. (i.e., create AtlasXML with specific locations not yet populated) (Flax)</li> </ul>
APRIL thru June	<ul style="list-style-type: none"> <li>• Populate metadata for all reference images as they arrive. (JB)</li> <li>• Document and create training materials for the branch-and-merge collaborative model. (FCS staff)</li> </ul>
MAY & June	<ul style="list-style-type: none"> <li>• Modify Squiggle to amend AtlasXML by adding SVG elements as users specify exact locations on the page using a "Paint"-like GUI. (Flax)</li> <li>• Begin planning for meeting of MEL editorial and technical associates in October</li> </ul>
JUNE	<ul style="list-style-type: none"> <li>• Decide upon an open-source library for XML entry (Flax)</li> <li>• Extend TEI-XML to refer to AtlasXML (Flax)</li> </ul>
JUNE & July	<ul style="list-style-type: none"> <li>• Integrate XML editing of TEI-XML into Squiggle (Flax)</li> </ul>

JUNE thru August	<ul style="list-style-type: none"> <li>• Populate AtlasXML with locations of interest on manuscript reference images. (JB)</li> <li>• Specify transformation from TEI-XML to HTML for display purposes. (JB)</li> <li>• Layout content of project home page (JB)</li> </ul>
JUNE thru September	<ul style="list-style-type: none"> <li>• Implement TEI-XML to HTML transformation in XSLT (FCS staff)</li> </ul>
JULY	<ul style="list-style-type: none"> <li>• Link TEI-XML and AtlasXML within Squiggle. (Flax)</li> <li>• Decide upon an open-source XSLT library. (Flax)</li> </ul>
JULY & August	<ul style="list-style-type: none"> <li>• Integrate XSLT library into Squiggle for the rendering of transformed TEI-XML. (Flax)</li> </ul>
JULY thru September	<ul style="list-style-type: none"> <li>• Design project home page (FCS staff)</li> </ul>
AUGUST & September	<ul style="list-style-type: none"> <li>• Create batch tool to generate XHTML+SVG suitable for web display, based on AtlasXML, TEI-XML, and XSLT transforms. (Flax)</li> <li>• Determine which branch of transcription should be displayed on homepage (JB)</li> </ul>
SEPTEMBER & October	<ul style="list-style-type: none"> <li>• Use batch tool to populate home page with highly linked and linkable XHTML+SVG documents. (FCS)</li> </ul>
SEPTEMBER & November	<ul style="list-style-type: none"> <li>• Modify code to accommodate project requests. (Flax)</li> </ul>
OCTOBER	<ul style="list-style-type: none"> <li>• Hold MEL editorial and technical meeting</li> </ul>
OCTOBER thru January 2009	<ul style="list-style-type: none"> <li>• Bulk of manuscript revision site transcription (JB)</li> <li>• Launch MEL's TextLab prototype</li> </ul>
NOVEMBER	<ul style="list-style-type: none"> <li>• Begin writing grant proposals for future funding of MEL and the digitizing Melville manuscript holdings at the Houghton</li> </ul>
JANUARY	<ul style="list-style-type: none"> <li>• Refine MEL's TextLab prototype</li> <li>• Complete Grant proposals</li> </ul>
JANUARY & February	<ul style="list-style-type: none"> <li>• Handoff code base to Hofstra Faculty Computing staff (Flax, FCS)</li> <li>• Refine MEL's TextLab prototype</li> </ul>