It's For Sale, So It Must Be Finished: Digital Projects in the Scholarly Publishing World

David Sewell <dsewell_at_virginia_dot_edu>, University of Virginia Press

Abstract

Since the early 1990s, theorizing in the digital humanities has often celebrated open-endedness and incompletion as inherent qualities of digital work. But a scholarly publisher undertaking preparation and sale of digital objects cannot altogether dispense with traditional notions of deadlines and completion if those publications are to enter the dual marketplaces of peer review and institutional purchase. The Electronic Imprint of the University of Virginia Press was funded in 2001 with the goal of bringing born-digital scholarly projects under the aegis of the same review and marketing system that applies to books. In this article I describe how we defined the criteria for "done-ness" in creating two very different projects, a born-digital edition of Herman Melville's *Typee* manuscript and a conversion of the letterpress *Papers of George Washington* into a digital edition. Our experience suggests that it is possible to categorize different genres of digital creations based on the extent to which intrinsic criteria for "done-ness" can be applied to them, and that decisions about completeness are always subject to extrinsic factors as well, such as budgetary constraints and the pressures created by competition and the evolution of standards.

Some History of Terms

Viewed from the perspective of someone who works for a university press, the semantics of the term "done" as applied to digital objects is rather curious. From our point of view, it's generally a good thing for a scholarly publication to be "done": review copies can be sent out, books can be shipped to distributors, and budgets perhaps even met. Traditional publication in the scholarly publishing world has always meant the implicit guarantee that a work is the end product of a rigorous process of peer review, revision, copyediting, design, and proofreading shared institutionally by author, press boards, outside scholars, and in-house staff. When a book or journal issue is "done" it is a source of pride and satisfaction for everyone concerned. [1]

The case seems to be different with digital objects. The claim that a digital project or publication is "done" may be met with suspicion. What do you mean, your Web-thing is finished? Since it's nonlinear, how do you know where it starts or ends? Won't there always be more features or links you can add? If your Web-thing is so much like an old-fashioned codex book that you can call it "done", does it really belong online in the first place? This suspicion has a history. Theoretical discussion of projects in the digital humanities has, since the 1990s, suffered from semantic slippage between two related but nonidentical pairs of contradictory terms: on the one hand, "open" versus "closed"; on the other hand, "complete" versus "incomplete" (or "unfinished" versus "done", etc.). The tendency has been to merge these two sets into a single pair, then to valorize the first pair of terms and to demonize the second. [2]

One of the more polished articles on Wikipedia these days, ironically, is on the topic "Unfinished Work"; it discusses incomplete works in various domains from literature and music through architecture to software. On the article's discussion page, the first thing we find is some amused perplexity about the label's applicability to the very source it appears in ['Unfinished Work' 2007]: [3]

It is a familiar conundrum about the nature of digital texts. Obviously, a formally defined text like a sonnet can be recognized as complete or incomplete; it's the difference between a well-wrought urn and a pot whose clay is still wet. But can a nonlinear, extensible, text ever be said to be finished? Is it by definition unfinished, or is the opposition "finished/unfinished" just plain inapplicable to open-ended texts?

[4]

These are theoretical questions I'm not in a position to answer, but I would submit that early in the 1990s the postmodern admiration of the "open-ended" at the expense of the "closed" somehow got turned into a celebration of the "unfinished" and a suspicion of the "done," and that this transmutation may have been one of the things that delayed the entrance of digital scholarship into the traditional system of peer-reviewed academic publication.

[5]

Consider these assertions from George Landow and Paul Delany's 1991 essay "Hypertext, Hypermedia and Literary Studies: The State of the Art":

[6]

> Particularly inapplicable [to hypertext] are the notions of textual "completion" and of a "finished" product. Hypertext materials are by definition open-ended, expandable, and incomplete. If one put a work conventionally considered complete, such as the *Encyclopedia Britannica*, into a hypertext format, it would immediately become "incomplete." [Landow and Delany 1991, 13]

A clever reader might object that even in print the *Encyclopedia Britannica* is always incomplete: like any reference work, it is constantly being updated and reissued. So when Landow revises this particular passage for his 1992 book *Hypertext*, he makes the claim even more radical by making a single change to the second sentence to replace the encyclopedia with a work of literature:

> Hypertextual materials, which by definition are open-ended, expandable, and incomplete call such notions into question. If one put a work conventionally considered complete, such as *Ulysses*, into a hypertext format, it would immediately become "incomplete." [Landow 1992, 59]

Landow is now claiming that even a recognizably closed, well-wrought modernist text becomes both open and unfinished when put online. And he ends his 1992 discussion of completion by citing Derrida to the effect that "a form of textuality that goes beyond print 'forces us to extend...the dominant notion of a "text" '," so that it "is henceforth no longer a finished corpus of writing, some content enclosed in a book or its margins but a differential network, a fabric of traces referring endlessly to something other than itself, to other differential traces" [Landow 1992, 59].

Julia Flanders has observed in a memorable phrase that the digital humanities have sometimes suffered from "a culture of the perpetual prototype" [Flanders 2007], and identified some plausible economic and institutional causes. To them I think we can add the theoretical conflation of the digital with *différance*. After all, what was the postmodern project if not a cult of the perpetual prototype?

[7]

Rotunda: A Scholarly Digital Imprint

My organization, the Electronic Imprint of the University of Virginia Press, was established in 2001 to test the proposition that instances of digital scholarship *can* be bounded, completed, and presented for review, sale, and academic consumption in much the way journals and monographs had been for decades. We were grant funded, with support from the University and the Mellon Foundation awarded to a proposal co-written by the Press and John Unsworth, who was then head of the Institute for Advanced Technology (IATH) at Virginia. We became fully staffed in late 2002, and two years later released our first publication, a born-digital edition of Dolley Madison's correspondence, under our new imprint name of "Rotunda." Since then we have expanded to a total of seven publications in two separate collections: nineteenth-century literature and culture, and the American Founding Era. Our main focus for the next few years will be

[8]

creating fully-featured digital versions of the papers of American presidents and other Founding Era figures that began as multivolume (and often still ongoing) print editions, joining our *Papers of George Washington Digital Edition* (*PGWDE*), which was released in February 2007.

The underlying data format of all of our Rotunda publications is XML, tagged according to the guidelines of the Text Encoding Initiative (TEI), plus accompanying digitized graphical material. Unlike many other university presses with digital projects, we outsource none of our technical work except for graphic Web design; our markup specifications, stylesheets for file transformation, and programming for Web delivery (mostly coded in XQuery using MarkLogic Server as the back-end platform) are all done in-house by several programmers and technical editors. |9|

Born-digital Rotunda publications go through the same steps that our books go through: approval by a Press committee and then the Press Board after reports from external reviewers; signing of a contract complete with royalty agreements; sharing of "review copies" (in the form of password access) with librarians and academic reviewers. Digital editions such as *PGWDE* that are based on existing print series are produced in close collaboration with the scholarly communities (historians and documentary editors) who create and use the letterpress volumes. Clearly all parties to our process of publication and sale are implicitly agreeing to bracket the theoretical issue of when or whether a digital work is ever "done" by applying a socioeconomic definition: it is "done" when the Press is prepared to offer it for purchase and customers are prepared to buy it. |10|

I turn now to two very dissimilar examples of our publications — *PGWDE* and an edition of Herman Melville's *Typee* manuscript — and discuss the decisions we made about what we could or couldn't include in the finished work; when we counted each as "done" for initial release; and to what extent we consider the published release genuinely complete or part of a work still in progress. |11|

Melville's *Typee*: A Fluid-Text Edition

Figure 2. *Typee* splash page

John Bryant's edition of a portion of Herman Melville's novel *Typee* was first envisioned and prototyped years before Rotunda came into existence. Bryant has been editing Melville's texts for two decades, and has long felt that any critical edition of a text that survives in more than a single version needs to be faithful to its evolutionary history; it should be what he calls a "fluid-text" edition. Because a fluid-text edition needs to capture a dynamic process, a computer-based format is a natural fit, and he began imagining one for Melville as early as the 1980s (personal communication).

The textual history of *Typee* is fairly complicated. The only surviving manuscript fragment covers about three chapters of the published novel. It contains a multitude of cancellations, erasures, and additions by Melville, both in ink from the time of first composition and in pencil from later stages of revision and proofreading Figure 3.

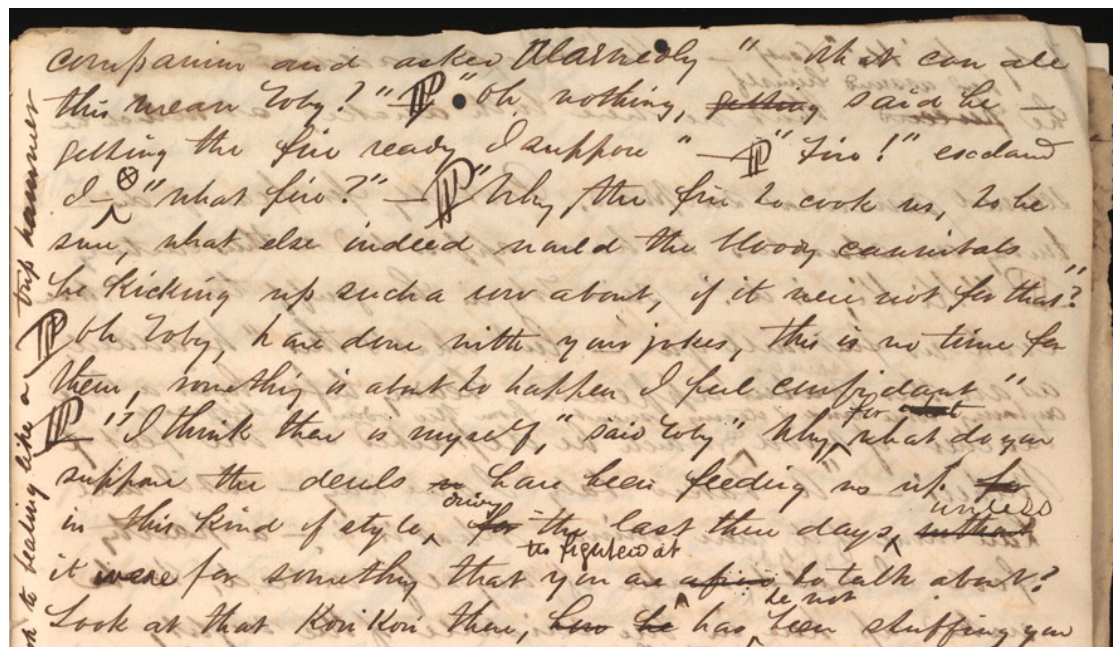**Figure 3.** Part of a *Typee* manuscript page, showing Melville's typical deletions and insertions; the circled *x* above the fourth line is keyed to a separate insertion leaf.

We know that Melville made changes in proof before the first English edition was issued, and that the first American edition contained still more changes, some requested by Melville, others made by the publisher probably without the author's assent. Bryant's goal for a digital fluid-text edition was to capture all of these stages and to allow the reader to follow the sequence of composition and the editor's narrative reconstruction of that sequence, zooming in and out to any point in manuscript time and space during the entire period from initial composition through the published editions.

Development of the Edition

Bryant was not himself a programmer or XML specialist, but he did have ideas about what a computer edition might look like, and created detailed storyboards before any actual programming work began. Although these were necessarily static, they used frame- and button-like boxes to suggest how a screen presentation might respond dynamically to reader choices (Figure 4). In 1998 he was named an IATH Visting Fellow and received technical assistance to create a first proof-of-concept prototype of the edition [Bryant 2000], which translated the storyboards into standard HTML frames (Figure 5).

Writing Typee

MANUSCRIPT   EDITIONS OF 1846   LATER EDITIONS   REVISION NARRATIVES

Transcript & Reading Text | First English | First American | American Revised | Second English | 1892 | 1968 NN | MS | Print

Revision Site: RS11ms127-128e175-176

TRANSCRIPT PASSAGE

16 horrid a custom. Nor did they omit to point out to
        challenge
17 ∧ our admiration for the natural lovliness of their
        beautiful abode
18 favored valley and the lavish abundance with which it
        paradisical
19 produced all manner of luxuriant fruits, exalting

REVISION NARRATIVE
Step 5 [print]

_____

RN11e175-176: In his fair-copy stage, HM revised his problematic opening by replacing "challenge our admiration for" with "call upon us to admire." He also dropped the ironical "their paradisical abode" for "their own abode."

1 Nor did they omit to point out to our admiration

2 Nor did they omit to [...] challenge our admiration for the natural lovliness of their favored valley

3 Nor did they omit to challenge our admiration for the natural lovliness of their [...] beautiful abode

4 Nor did they omit to challenge our admiration for the natural lovliness of their paradisical [...] abode

Nor did they omit to **call upon us to admire** the natural loveliness of their [...] **own** abode [RN11e175-176]

Figure 4. An early storyboard by John Bryant illustrating a possible interface and layout for the *Typee* edition

---

The Melville Project

Table of Contents

Leaf 1, Recto
Leaf 1, Verso

Reading Text: *The Typee Manuscript (Chapters 12 - 14)*

Leaf 1, Recto

1.
transported to the stream which flowed perhaps two
2.
hundred yards from the house. ¶ **Our**ms1
3.
appearance upon the verandah in front of the habitation
4.
drew together quite a crowd [of the natives]e1, who now
5.
stood looking on & conversing [in the most animated mannerwith one another,]e2ms2
6.
[like]e3 a group of [gossiping]e3 **village idlers gathered about the**
1.
· **tavern3 door1 of a village2**ms3 when the equipage of some ·
7.
distinguished traveler is brought round **previous to his**
1.
departure.ms4 As soon as I clasped my arms about the
8.
neck of the devoted fellow, & he jogged off with me,
9.

Revision Sequences

RSms1: On arriving up (upon) ¶ MyOur appearance upon the verandah

1. On arriving up
2. **My appearance upon the veranda**
3. **Our** appearance upon the veranda

RSe1: Our appearance … drew together quite a crowd of the natives

1. quite a crowd
2. quite a crowd [...]

Revision Narratives

RNms1: HM first considers writing something like "On arriving upon the verandah," but he cancels the idea before he completes the word "upon" [step 1] and changes to "My appearance upon the verandah" [2], which in a later proofreading phase, he emends to "Our appearance" [3], thus stressing Kory-Kory's presence as well. See RNms8 and e7.

RNe1: "Of the natives" does not appear in print, and the reason for the cut is not entirely clear given that the grammar of the resultant "crowd who" is questionable, and that removing "of the natives" breaks the parallelism with "group of gossiping idlers." HM may

---

Figure 5. A first prototype of the *Typee* edition produced in 2000 by IATH at the University of Virginia

In late 2003 we received Bryant's "manuscript" of the edition, consisting of Microsoft Word and PageMaker files containing manuscript transcriptions flagged with hundreds of "revision sites" and for each separate revision site a "revision sequence" and a "revision narrative." We licensed from the New York Public Library the rights to reproduce their full-color photographs of the entire manuscript. Our goals at this point were (1) to convert all transcription and commentary to TEI-XML, and (2) to design an environment that could deliver combinations of text and image to realize as closely as possible the author's intentions for his edition. Our own finished rendering of the original concept [Bryant 2006] would look like this:
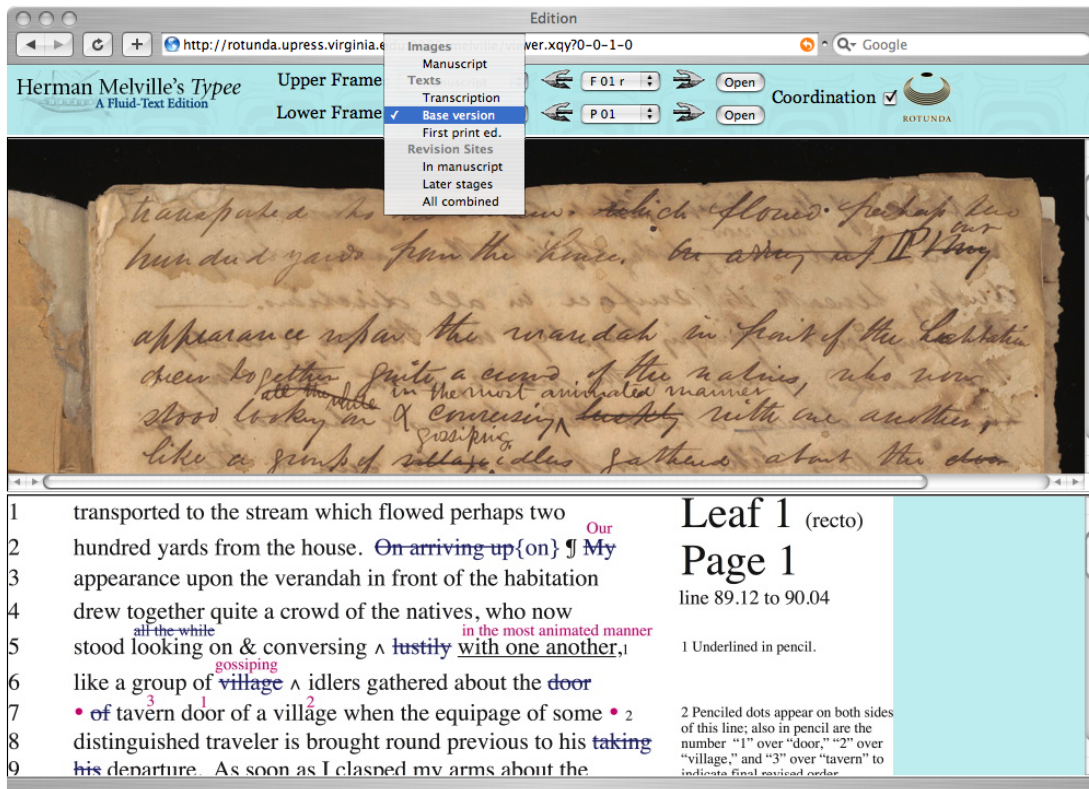
Figure 6. Default view of the Rotunda version of *Typee*, with a manuscript page image in the top frame and a bitmapped typographic transcription in the lower one (the current state of HTML made browser rendering impractical). Readers can choose to display different combinations of image and transcription state in either frame. JavaScript controls allow upper and lower frames to scroll in unison.
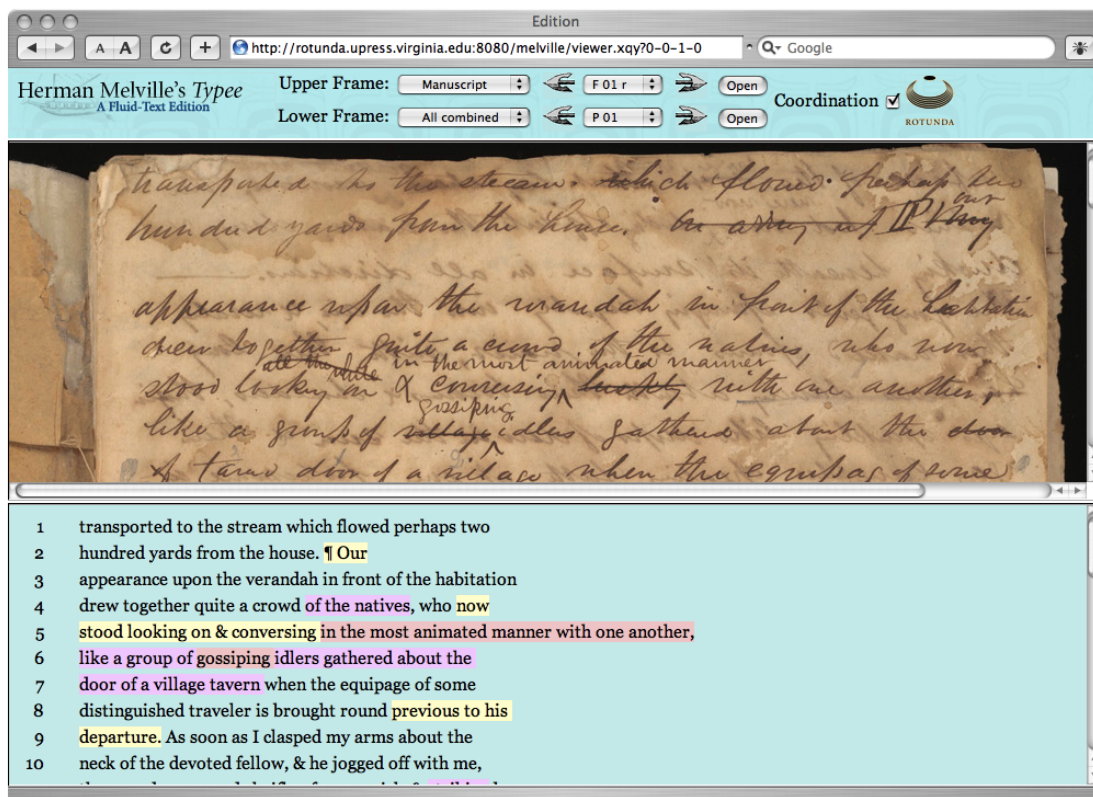
Figure 7. The lower frame now shows a textual transcription of the same lines rendered in HTML, where color-coding represents "revision sites" from different stages of compositional history.
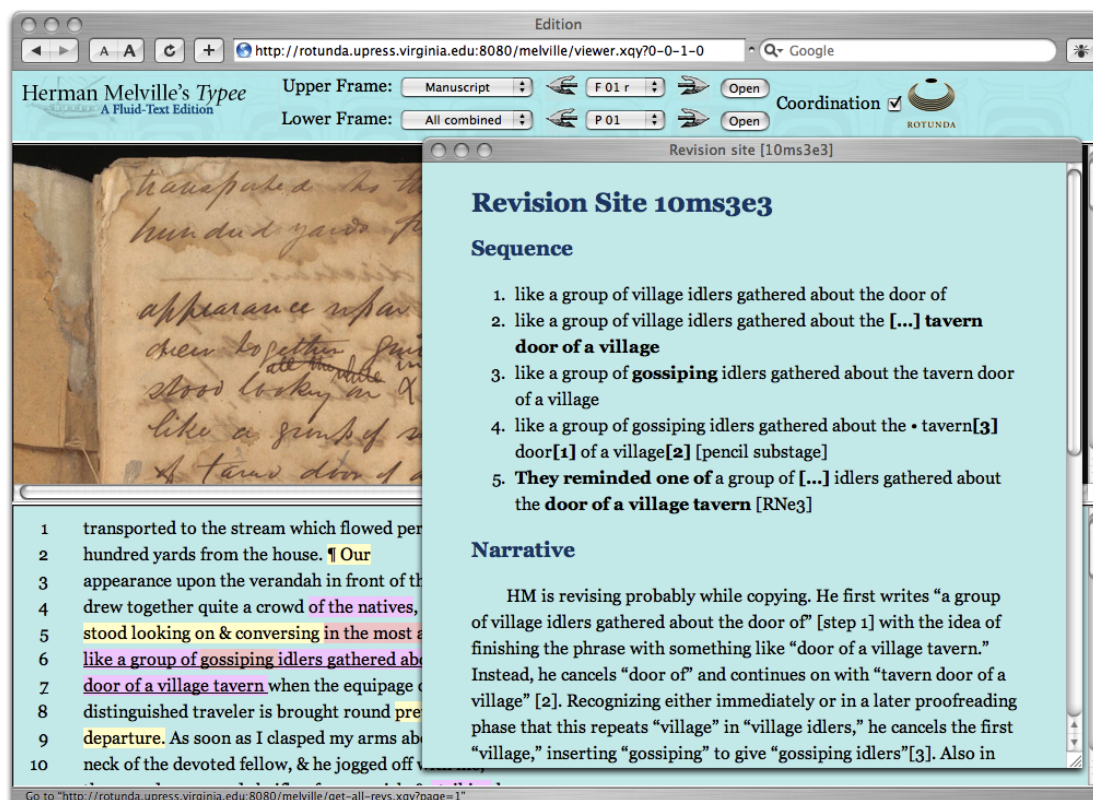

Figure 8. Clicking on any highlighted revision site brings up a window containing Bryant's reconstruction of the revision sequence and a narrative explanation of it.

Once we had our basic page display working, all that remained was to code a search page and add the editorial introductions before declaring the edition "done" and releasing it in March 2006.

Three months later we added an enhancement, our major one to date, an XML-based version of the entire first British edition of the novel, which the University of Virginia Library digitized for us from a copy in their holdings. We created for it a display interface combining a transcription of the text with images (Figure 9).
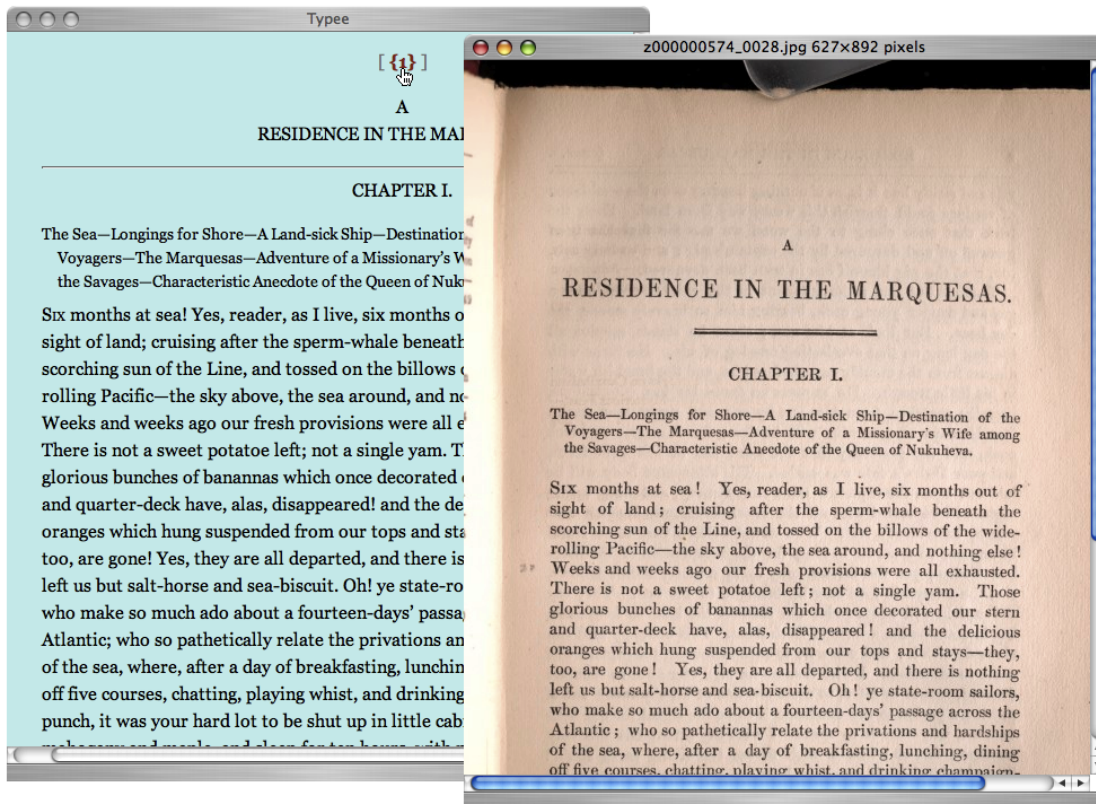

Figure 9. The added first British edition includes an XML-based transcription of the text plus optionally viewable page images from the printed text.

Is Our *Typee* Done?

Neither Bryant nor the Press conceived of the *Typee* edition as an open-ended project. The editor's work was done once he had finished all the manuscript transcription, identification of revision sites, exposition of revision sequences and narratives, and the introductory editorial essay. Our work was done once we had translated the editor's vision into a fully functional edition that coordinated photographic facsimiles with several transcription formats, and that hyperlinked all "revision sites" with their editorial expansions. The March 2006 edition was lacking one intended feature, the first British edition, owing to extrinsic factors (our library's digitization schedule). Once that was added, *Typee* was for practical purposes stable and complete.

Nevertheless, we were aware of the potential for improvements and enhancements, some more immediately practicable than others:

- We could generate RDF metadata files in the format used by the Collex tool created by Jerry McGann and his NINES team. In July 2007 we did this, so that the base view of each manuscript page exists as an indexed object in Collex, along with the editorial introduction and the publication home page.
- The full-text search needs improvements to return hits on supplied text and to properly handle word tokens containing XML tags (for example: `savage<add>ry</add>`). The first item is on our to-do list; the second is on hold until our MarkLogic software adds the ability to ignore selected elements for the purpose of word

tokenization.[1]

More radically still, it is conceivable that all of our underlying XML markup and presentation might be entirely revised if John Bryant were to incorporate the proposals for temporal encoding in genetic digital editions that Elena Pierazzo has advanced [Pierazzo 2007], as both her tagging strategy and theoretical approach vary significantly from his own. A revision of that magnitude would be analogous to issuing a second edition of a book that differs markedly from the original because it has responded to new evidence and/or arguments. All scholarly and scientific publications are potentially imperfect and thus "incomplete" to the extent that later work can call them into question, but it would be an equivocation to say that they are therefore always unfinished in a formal sense.

20

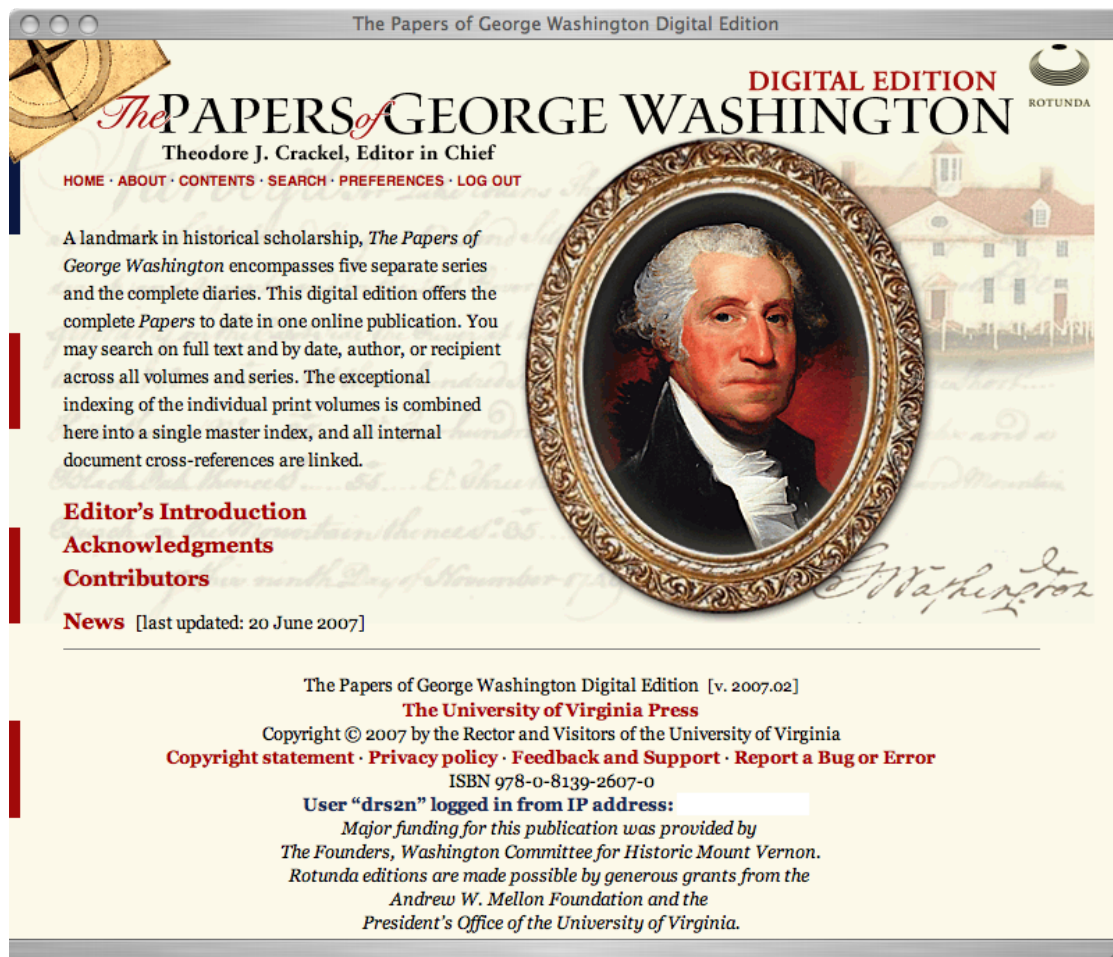## The *Papers of George Washington Digital Edition*

Figure 10. *PGWDE* splash page in July 2007

The *Papers of George Washington Digital Edition* [Crackel 2007] is a very different project, one initiated in 2004 by the Press in collaboration with the editorial staff of the Papers of George Washington (also based at the University of Virginia), and partially funded by a grant from Mount Vernon. Our mission was to produce an online version of the fifty-two volumes then in print of our letterpress *Papers of George Washington* [Jackson 1976], the authoritative scholarly edition of the documentary legacy of the first president. Owing to the size and complexity of the letterpress edition, [2] its adaptation to a fully-featured online format offered us as many design and programming challenges as a born-digital project like *Typee*. We needed to establish an appropriate XML schema and encoding specifications, decide on what structural and semantic tagging to do and what metadata to add, figure out how much regularization of inconsistencies in the letterpress edition we could accomplish, and design a Web environment for display, navigation, and searching of the edition usable by advanced scholars and beginning students alike.

21

For the editorial staffs of The Papers of George Washington and UVa Press, the criteria for regarding a letterpress volume as complete have been well established since the project began in the 1970s:

- all known documents from the period covered by the volume are included or referenced
- all document transcriptions are complete and have been checked for accuracy against manuscript facsimiles
- all possible identifications of persons have been made and included in the endnotes
- all other annotation and editorial introductions are written
- the manuscript has been copyedited
- page proofs have been checked and used to produce a back-of-the-book index.

But for us, including the full content of the print edition would not be sufficient. We could consider *PGWDE* "done" only when we had reliably translated textual and scholarly conventions into an online format that offered as much (to use the inevitable marketing phrase) added value as possible beyond simply being able to access the publication without visiting a library.

Goals for the *PGWDE*

Determining where we could add value to the print edition required a preliminary analysis of what makes a scholarly edition valuable in the first place. In such an edition, the basic textual unit is the single document, always accompanied at a minimum by bibliographical information and usually by editorial annotation, and sometimes by translations, enclosed documents, or other ancillary material. (Diaries and journals are a special case: depending on how chronologically structured they are, the basic textual unit may be the single-day entry, the single-month entry, or a longer narrative.) Besides the original text and editorial material, documents contain metadata, cross-references, abbreviations and other special features that are represented using a variety of editorial and typographical conventions, as highlighted in the facsimile of the original letterpress version of a letter from William Livingston to George Washington (shown below). Beyond the document level, most volumes contain scholarly apparatuses (lists of abbreviations and bibliographic expansions of short-title references), editorial and historical introductions, and a detailed index of all proper names and hundreds of topic categories. The translation of all of these print conventions into their TEI-XML equivalents is what must undergird a digital edition. (Our final XML encoding of the Livingston letter may be seen in the appendix.)

Figure 11. A page from the Revolutionary War Series of the print edition of *The Papers of George Washington*. Many, but by no means all, of the features in a documentary edition that need to be captured in the underlying XML representation are typographically distinct; determining the optimal mixture of automated and human tagging is one of the challenges of converting a legacy edition.

Our initial goals for the digital edition were:

1. to provide document-by-document display (or, for diaries, month-by-month or day-by-day, as appropriate) closely resembling that of the letterpress source;
2. to offer a wide variety of means for navigating into the documents: through full-text search; through a hyperlinked consolidated index based on the back-of-the-book print indexes; via tables of contents similar to those in the print edition; and by chronology (in order to collect all documents and diary entries for a given date, for example);
3. to use as much tagged information as possible for display, linking, and refinement of searching;
4. to create a genuinely new edition incorporating corrections to the print edition submitted by the Papers of George Washington staff, along with consolidated and regularized lists of names and titles that had varied from volume to volume in the letterpress edition.

Work on *PGWDE* began in fall of 2004; a beta version for public display was ready by October 2006; and we formally released a published version for sale in February 2007. Screen captures of the online version of the Livingston letter illustrate how we realized some of our goals (Figure 12, Figure 13). Compasses are used to navigate the four hierarchies identifd in goal 2. A "breadcrumb trail" allows quick navigation up to any higher node of the current tree. Hyperlinks or mouseovers provide dynamic equivalents of their print counterparts in ways that are familiar to Web users: endnote superscripts connect to their notes via bidirectional linking; abbreviations and short-title references (indicated

by dotted underlining) are expanded when the user mouses over the abbreviated text; and cross-references to other documents in *PGWDE* are active links.
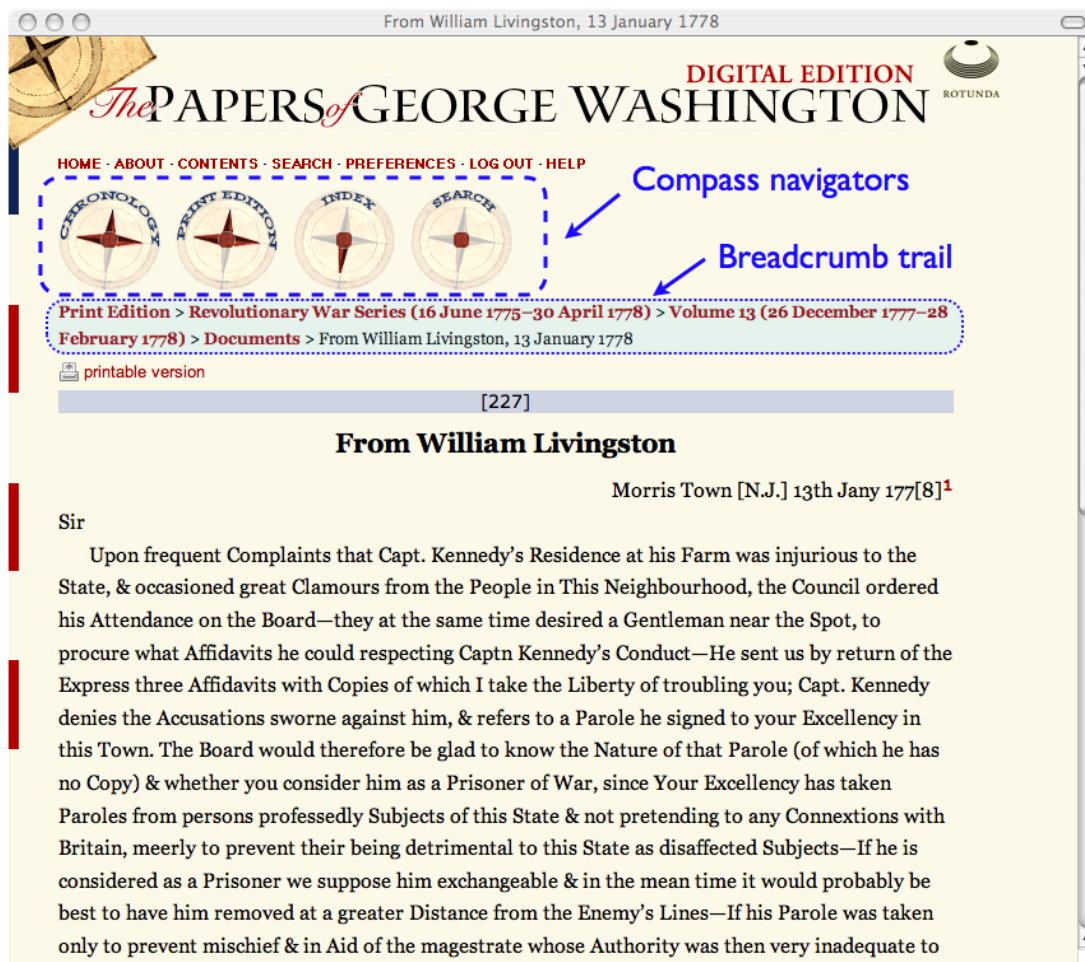
Figure 12. Top of the Livingston letter in a browser view.
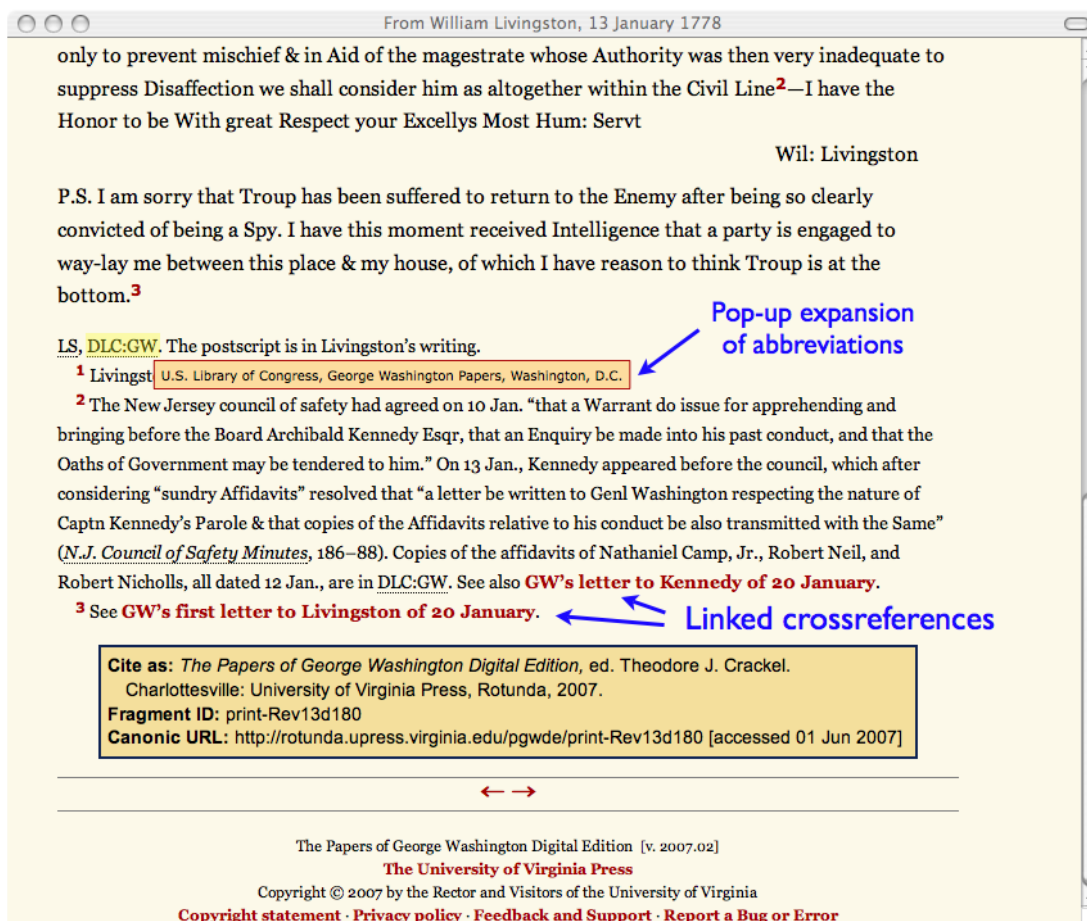
Figure 13. Bottom of the Livingston letter.

Along with the document navigation and display, we programmed a search page that combines full-text search with optional filtering based on author, recipient, and/or date range; and we added an online version of the consolidated index that resembles a back-of-the-book index except that document titles and dates replace page numbers and are, of course, hyperlinked. | 26

We had scheduled official release of *PGWDE* for President's Day — February 19, 2007. By a month or so ahead of this deadline, we realized that every last cleanup task could not be completed by that date. Online publication meant we could do a triage: fix first the things that affect the most documents, or that are most obvious to the average reader; fix afterwards problems or errors limited to single documents, or ones that would be noticed only by a specialist (for example, an incorrect birth year for a minor historical figure). Corrections of bad links and minor formatting glitches continued for about a month after the February 19 release. Corrections to errors in transcriptions or annotations, as identified by Papers of George Washington staff, have been ongoing. So, too, have further regularization and consolidation: since first publication, PGW staff have provided us with fully normalized lists of names of all document authors and recipients that we have used to update the document metadata, and with a corrected and up-to-date list of repository abbreviations and expansions based on MARC codes that we have used to globally update the XML volume files. | 27

Planned Enhancements

So is *PGWDE* done? Yes and no. It is a stable release version with some remaining imperfections, but there's a lot more we plan to do with it, even apart from adding content as we digitize new volumes after they appear in print. We've recently met with the PGW staff to agree on a list of priorities for enhancement. The tasks fall into three general categories: | 28

1. Optimization of existing features. Examples: improving search speed and index retrieval; rewriting the search parser to make it more Google-like and to include more boolean operators.
2. New features. Add an "advanced search" page that will allow users to search by document features or language, for example. We'll add full-text searching on the index. Farther down the road may be "keeping up with the Joneses" enhancements, like enabling the user to save a personal workspace containing bookmarked documents and search result sets as the Works of Jonathan Edwards Online at Yale has done.
3. Features required by aggregating *PGWDE* into the larger Founding Era collection. Over the next year we will be adding editions of the *Adams Family Papers*, the *Papers of Thomas Jefferson*, and *The Documentary History of the Ratification of the Constitution*. In order to build an extensible framework, we have already begun a thorough rewrite of our delivery code and encoding specifications so that our publication system will scale gracefully as more publications are added.

But it is impossible for us to project all of the enhancements that may become desirable or possible in the future. We are in a position not too different, really, from that of the editors who began planning American documentary historical editions beginning in the 1950s. They of course knew that their completed volumes would eventually need to be supplemented and corrected as new documents and historical information emerged. What they couldn't envision was a time, our own, when scholarship that exists only in print is increasingly seen as *ipso facto* lacking an essential quality. Likewise, we can really only guess at what new features advancing technology may make possible. For instance, we have often wished we had time and funding to add rich XML tagging for personal and place names (beyond the author/recipient identifications in existing metadata), and assumed that this would require a major commitment of human labor. But it's not impossible that advances in automated named entity recognition will enable us in a not-too-distant future to pipe *PGWDE* through a program that will reliably recognize and tag those names and link them to, say, genealogical databases or a GIS-based interface like Google Earth.

So we don't really expect ever to be able to say more than that *PGWDE* is done — for now.

29

A Few Conclusions: Generalizing from the Rotunda Experience

> Ask any developer out there if a program is ever finished and they'll tell you, "No, of course not, I still need to...". But, ask any developer out there if the program is *almost* finished and, assuming that the development cycle has progressed far enough along, their answer will invariably be, "Yes, all I have to do is...". They may even quantify it: "80% complete". Ask them a couple of days, weeks, months (depending on the magnitude of the project) later and you will get a similar response, but with a different percentage, say 90%. And so forth…but never 100%. ("Never Finished; or Zeno's Paradox as an Analog to Software Development")

30

It is entirely possible to define "done-ness" for computer software in such a way that no instantiation of a software project can ever satisfy the definition. For example, suppose we stipulate that: "A program is complete and can be distributed when it (1) satisfies all initial design requirements, (2) is known to run 100% bug-free under all potential conditions of use, and (3) incorporates state-of-the-art programming techniques and tools at the time of distribution to offer the user an optimal mix of powerful features and ease of use." The second and third criteria entail that none but the simplest programs could ever count as complete. To claim "my program is bug-free because I have found no bugs in it" is *argumentum ad ignorantiam*; that you haven't found any doesn't mean they aren't there. And criterion 3 turns done-ness into a Red Queen's race, since the state of the art is constantly advancing, and at release time most complex software projects are already "belated" relative to the cutting edge of technology. To the extent that digital publications count as software projects, they too would fail ever to count as finished under such a definition.

The adjective "simplest" in the preceding paragraph hints at a way around the paradox. If my goal is to write a "hello world" program in Perl and I respond with the one-line program " `print "Hello World!\n";` " I can confidently say I'm done. If my task is to write a new operating system, it's another matter. As a rule, the more *complex* a task is, the less susceptible it is of being judged finished by any set of formal criteria. Contrast these two assignments:

1. Create a crop circle in the shape of a simple circle with a diameter of 40 meters.
2. Create a crop circle representing the coastline of Great Britain at 1/10000 scale.

The first project is done once you've made a single circuit while tromping down wheat at the end of a 20m tether. The second project is done when you've created as accurate a representation as your time and skill allow. Tracing a coastline is a problem in fractal geometry for which completeness will always be relative. In a sense, it is a formal property of project 2 that it is done only when you decide it is done. To put it another way, intrinsic criteria are used in both cases to determine when the project qualifies as finished, but as project 2 is formally undecidable (embodying the Turing halting problem that Matt Kirschenbaum mentions in his introduction), extrinsic criteria are also required to make the determination.

The digital publications that we have worked on in Rotunda have tended to resemble the fractal project more than the simple circle. With *PGWDE*, for example, the "coastline" we needed to reproduce was, like that of Britain, a pre-existing and well-defined object, the fifty-two volumes of the print edition. To have omitted a volume would have been as clear a sign of incompleteness as leaving Cornwall out of the crop circle. But decisions about the richness of our feature set were very much a matter of "how far to trace", and in the end were dictated by our available time and skill (and budget). If our experience is representative, deciding when to call a digital project "done" usually requires a process of negotiation between intrinsic criteria and external factors.

Intrinsic Criteria

Intrinsic criteria are formalist: they assume that the completeness of an object derives from its inner properties alone, without reference to any social or other external context. In the following table there is a continuum from objects like a monograph (or a lyric poem) that can be judged to possess organic unity, to ones like a collaborative virtual world that cannot. It is no accident that the latter are the ones felt to be characteristically "digital".

Category	Object has definable boundaries?	Object has satisfied its design goals?	Print world example	Digital example	Is it "done"?
1	yes	yes	monograph, journal article	monograph-like object, online article	yes
2	yes	no	preprint, "rough cuts"	beta or 0.9 release	not yet
3	no	yes (for current stage)	encyclopedia; any work issued in discrete series	same as for print world	yes (for now)
4	no	no	? ?	open-ended wiki, collaborative blog or social space, virtual world, etc.	no (by definition)

Table 1. "Done" as a function of intrinsic criteria

Category 1 objects are the most familiar in scholarly publishing, hence the most fully integrated into the tenure-and-review system. Bryant's *Typee* essentially falls into this category. Probably owing to the influence of online publishing, Category 2 objects are becoming more familiar: online preprints are accredited scholarly communications in a growing number of disciplines, and cutting-edge book publishers like O'Reilly with their Rough Cuts series of early-access PDF are adopting a "versioned" model of publication.

Category 3 objects are also familiar from the print world, where they represent the one kind of open-endedness that does not upset traditional notions about scholarly authority. The *Oxford English Dictionary* is a good example. It has been supplemented, transformed, and extended many times since the first fascicles were issued in 1884. Yet each discrete stage of publication was accepted by the academic community as authoritative for its moment. It is no accident that this category has translated easily to digital format: the only essential difference between the print and online *OED* is that the latter is updated far more often.

Category 4 is the one to which the term "done" seems the most inapplicable. Its characteristic objects are more like processes than products, and it is difficult to think of genuine analogues in the print world outside the realm of experimental literature of the Oulipo variety. A publication like *Wikipedia* can perhaps be seen as a special case of Category 3 in which discrete stages succeed each other with extreme rapidity, but a virtual world like *Second Life* exists in such constant motion that it requires something akin to calculus for adequate description. Unsurprisingly, Category 4 is the form of digital creation least amenable to naturalization in the academic reward system or the scholarly publishing marketplace.

Extrinsic Factors

For a scholarly publisher, intrinsic criteria of done-ness are important but are often trumped by extrinsic factors. The judgement that a book manuscript is done and ready for press requires an agreement among author, acquiring editor, external reviewers, and the manuscript editorial and production departments that is based largely on its formal content. But completely extrinsic factors such as the desire to include the book in a particular season's list will often lead a press to veto an author's wish to continue tinkering with a manuscript. Similarly, an author may not consider a monograph on Chinese art formally complete without the inclusion of several dozen full-page color reproductions on glossy inserts, but a publisher may omit them for the wholly extrinsic reason that the profit-and-loss sheet doesn't budget for them. Once a book is in print, decisions about its subsequent "done-ness" (i.e., whether to reprint, revise, issue in paperback, etc.) are based almost entirely on economic factors. In the case of digital publications, I will suggest, extrinsic factors become important at an earlier stage and are proportionately more important at every stage of composition and publication.

The following list of extrinsic factors is not meant to be exhaustive; they are the ones that have been most prominent in Rotunda's experience.

> **Economic constraints** Two maxims apply: *(1) if a digital publication doesn't sell, it's "done"; (2) if the projected cost of upgrade exceeds projected revenue, it's "done"* . (For freely distributed projects, substitute "when no more grant funding or volunteer time is available, it's done.")

> **Competition** Maxim: *when your competition is adding features to its product, they can render your finished product "incomplete."* In the print world, this phenomenon is familiar in textbook and reference publishing. In the digital world, it is absolutely pervasive. No online publication, free or for sale, can afford long-term stasis when the peer publications it is compared with are adding bells and whistles (a list that would include, as of 2008, things like Ajax-powered form fields, tag clusters, user reviews and personal workspaces, page previews on mouseover, selectable themes or "skins" . . .) In the prestige economy as in the market economy, keeping up with the Joneses is not optional.

> **Standards evolution** Maxim: *even absent competition, the evolution of standards can make a finished project "incomplete."* This is primarily a matter of adhering to best practices, though not entirely free from the keeping-up-with-the-Joneses factor. Certainly if your academic discipline adopts a new format for metadata, or your institution adds a requirement that Web publications meet accessibility guidelines, your projects need to be revised for conformity. In other cases, it may be a matter of pride to demonstrate that a project has upgraded to the latest standard, for example by converting archival XML from TEI P4 to P5 compliance, or by following the very latest W3C recommendation for XHTML or CSS.

> **Aggregation** Maxim: *a stand-alone publication will probably become "incomplete" when it is aggregated*

with other material. In Rotunda's experience, it is inevitable that the user interface and back-end coding one develops for a single digital project will need to be substantially revised once a second project is added and meant to interoperate with the first. (As a case in point, it would require major effort to get our first publication, the *Dolley Madison Digital Edition*, seamlessly integrated with *PGWDE*, as the back-end programming and underlying XML data formats of the two publications are quite different.)

Technological change Maxim: *new technology **will** make your publication "incomplete"* . This goes almost without saying. The evolution of hardware, operating systems, programming languages, and Web standards will eventually make any online publication obsolete. Failure to migrate a digital object periodically as technical conditions require is the analogue of allowing a published book to go out of print. (In fact it's worse: it's like printing the book on high-acid-content paper with ink that fades on exposure to light, and *then* letting it go out of print.)

A Necessary Synthesis

Whether you are a publisher or the editor of an open-access publication, allowing extrinsic factors to influence your decision about whether a digital project is done is in no way an admission of defeat or an abdication of responsibility. It is, in fact, the only way to avoid the form of Zeno's paradox whimsically propounded in the epigraph to this section. The progress of knowledge in the arts and sciences is continuous, but in order for it to happen at all, scholarly discourse must be distributed in the form of discrete objects that can be shared, read or viewed, responded to, assimilated, quoted, disputed, and revised. In the marketplace of ideas, it's less important how you decide when your piece is done than that you *do* decide, label it and put it on display, and prepare to haggle with others over its value.

Appendix: XML markup of a sample Washington letter

(Formatting has been applied for convenience in reading; carriage returns are not introduced within mixed-content elements in the original files.)

```xml
<div1 xml:id="Rev13d180" type="doc">
   <FGEA:mapData id="GEWN-03-13-02-0189">
      <bibl>
         <title>From William Livingston, 13 January 1778</title>
         <author>Livingston, William</author>
         <name type="recip">GW</name>
         <date when="1778-01-13"/>
      </bibl>
      <FGEA:Author>Livingston, William</FGEA:Author>
      <FGEA:Recipient>GW</FGEA:Recipient>
      <FGEA:mapDates>
         <FGEA:searchRange from="1778-01-13" to="1778-01-13"/>
         <FGEA:dayRange from="1778-01-13" to="1778-01-13"/>
      </FGEA:mapDates>
      <FGEA:pageRange from="Rev13p227" to="Rev13p227"/>
   </FGEA:mapData>
   <pb n="Rev13p227"/>
   <head>From William Livingston</head>
   <div2 type="docbody">
      <opener>
         <salute>Sir</salute>
         <dateline>Morris Town [N.J.] <date when="1778-01-13">13th Jany
177[8]<ptr n="1"
                  target="Rev13d180n1"/></date></dateline>
      </opener>
      <p>Upon frequent Complaints that Capt. Kennedy's Residence at his Farm
 was injurious to the
         State, &amp;amp; occasioned great Clamours from the People in This
 Neighbourhood, the
```

Council ordered his Attendance on the Board — they at the same time desired a Gentleman
                near the Spot, to procure what Affidavits he could respecting Captn Kennedy's Conduct —
                He sent us by return of the Express three Affidavits with Copies of which I take the
                Liberty of troubling you; Capt. Kennedy denies the Accusations sworne against him,
                &amp; refers to a Parole he signed to your Excellency in this Town. The Board would
                therefore be glad to know the Nature of that Parole (of which he has no Copy) &amp;
                whether you consider him as a Prisoner of War, since Your Excellency has taken Paroles
                from persons professedly Subjects of this State &amp; not pretending to any
                Connextions with Britain, meerly to prevent their being detrimental to this State as
                disaffected Subjects — If he is considered as a Prisoner we suppose him exchangeable
                &amp; in the mean time it would probably be best to have him removed at a greater
                Distance from the Enemy's Lines — If his Parole was taken only to prevent mischief
                &amp; in Aid of the magestrate whose Authority was then very inadequate to suppress
                Disaffection we shall consider him as altogether within the Civil Line<ptr n="2"
                target="Rev13d180n2"/> — I have the Honor to be With great Respect your Excellys
                Most Hum: Servt</p>
        <closer>
                <signed>Wil: Livingston</signed>
        </closer>
        <ps>
                <p>P.S. I am sorry that Troup has been suffered to return to the Enemy after being so
                clearly convicted of being a Spy. I have this moment received Intelligence that a
                party is engaged to way-lay me between this place &amp; my house, of which I
                have reason to think Troup is at the bottom.<ptr n="3"
target="Rev13d180n3"/></p>
        </ps>
    </div2>
    <div2 type="docback">
        <note type="source" xml:id="Rev13d180sn">
            <p>
                <bibl n="docSource">
                    <rs type="dType">
                        <abbr>LS</abbr>
                    </rs>
                    <rs type="dWhere">
                        <ref target="GWPrep37" type="repository">DLC:GW</ref>
                    </rs>
                </bibl> The postscript is in Livingston's writing.</p>
        </note>
        <note n="1" type="fn" xml:id="Rev13d180n1">
            <p>Livingston wrote '1777.'</p>
        </note>
        <note n="2" type="fn" xml:id="Rev13d180n2">
            <p>The New Jersey council of safety had agreed on 10 Jan. 'that a Warrant do issue for
                apprehending and bringing before the Board Archibald Kennedy Esqr, that an Enquiry

```
                          be made into his past conduct, and that the Oaths of Government
          may be tendered to
                          him.' On 13 Jan., Kennedy appeared before the council, which
          after considering
                          'sundry Affidavits' resolved that 'a letter be written to Genl
          Washington respecting
                          the nature of Captn Kennedy's Parole &amp;amp; that copies of
          the Affidavits
                          relative to his conduct be also transmitted with the Same' (<ref
          target="PGWst1204"
                          type="short-title">
                          <hi rend="italic">N.J. Council of Safety Minutes</hi></ref>,
          186—88). Copies of
                          the affidavits of Nathaniel Camp, Jr., Robert Neil, and Robert
          Nicholls, all dated
                          12 Jan., are in <ref target="GWPrep37"
          type="repository">DLC:GW</ref>. See also <ref
                          type="document" target="Rev13d236">GW's letter to Kennedy of
          20
                          January</ref>.</p>
              </note>
              <note n="3" type="fn" xml:id="Rev13d180n3">
                  <p>See <ref target="Rev13d242" type="document">GW's first letter to
          Livingston of 20
                          January</ref>.</p>
              </note>
          </div2>
      </div1>
```

Notes

[1] In fact we could get around this software limitation by adding TEI tagging to provide the regularized word tokens for indexing, and adjusting our processing code accordingly. The decision not to do so was based on a cost-benefit analysis, feeling that the amount of time it would require outweighed the added utility in the context of this particular publication.

[2] The size and scope of the project can be illustrated with a few statistics:

- 52 printed volumes (originally published between 1976 and 2004) totalling ca. 30,000 printed pages, rekeyboarded and tagged as TEI-XML
- ca. 18,500 separate documents from 29,000 original sources (much of Washington's official correspondence survives in multiple drafts and copies)
- entries from Washington's diaries covering 560 months, with 6800 separate daily entries
- 2930 unique document authors and 1900 unique recipients
- corpus of ca. 11.5 million words (documents + annotation), with ca. 91,000 different word tokens, all indexed for full-text searching
- separate back-of-the-book indexes from each volume merged into a consolidated index containing over 35,000 main entries and nearly 100,000 subentries, with a total of over 400,000 separate page references

Works Cited

'Unfinished Work' 2007 Wikipedia. *Unfinished Work (discussion page)*. Wikipedia: the Free Encyclopedia, 2007. http://en.wikipedia.org/wiki/Talk:Unfinished_work.

Bryant 2000 Bryant, John, Haskell Springer and John Unsworth, eds. *The Melville Electronic Library*. The Institute for Advanced Technology in the Humanities, 2000. http://rotunda.upress.virginia.edu/melville.

Bryant 2006 Bryant, John, ed. *Herman Melville's 'Typee': A Fluid-Text Edition.* University of Virginia Press, 2006. http://www3.iath.virginia.edu/melville/.

Crackel 2007 Crackel, Theodore J., ed. *The Papers of George Washington Digital Edition*. University of Virginia Press, 2007. http://rotunda.upress.virginia.edu/founders/GEWN.html.

Flanders 2007 Flanders, Julia. "Panel presentation during session Coalition of Digital Humanities Centers". Presented at *Digital Humanities 2007* (2007).

Jackson 1976 Jackson, Donald, W.W. Abbot, Dorothy Twohig, Philander D. Chase, Theodore J. Crackel and Edward G. Lengel, eds. *The Papers of George Washington*. Charlottesville, VA: University of Virginia Press, 1976-.

Landow 1992 Landow, G. P. *Hypertext: The Convergence of Contemporary Critical Theory and Technology*. Baltimore: The John Hopkins University Press, 1992.

Landow and Delany 1991 Landow, George P., and Paul Delany. "Hypertext, Hypermedia and Literary Studies: The State of the Art". In George P. Landow and Paul Delany, eds., *Hypermedia and Literary Studies*. Cambridge, MA: MIT Press, 1991. pp. 3-50.

Pierazzo 2007 Pierazzo, Elena. "The Encoding of Time in Manuscripts Transcription: Toward Genetic Digital Editions". Presented at *Digital Humanities 2007*. *Digital Humanities 2007 Conference Abstracts* (2007).