



2013

Is Reliable-Social-Scholarly-Editing an Oxymoron

Peter Shillingsburg

Peter L Shillingsburg, peter.shillingsburg@gmail.com

Follow this and additional works at: https://ecommons.luc.edu/ctsdh_pubs



Part of the [Arts and Humanities Commons](#)

Recommended Citation

Shillingsburg, Peter, "Is Reliable-Social-Scholarly-Editing an Oxymoron" (2013). *Center for Textual Studies and Digital Humanities Publications*. 1.

https://ecommons.luc.edu/ctsdh_pubs/1

This Article is brought to you for free and open access by the Centers at Loyola eCommons. It has been accepted for inclusion in Center for Textual Studies and Digital Humanities Publications by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](#).

Peter Shillingsburg
Senior Professor
Loyola University Chicago
Submitted to the conference proceedings 8/24/13

Aug 24, 2013

"Is Reliable-Social-Scholarly-Editing an Oxymoron?"

The answer to this supercilious question depends on what one means by "reliable" what one means by "social" and what one means by "scholarly editing"--it might also depend on what one means by oxymoron.

Social could mean opening the development of the scholarly edition to one's partner, one's trained work team, a membership of vetted colleagues, or the public. Three is a crowd, as everyone knows; so if three people work together it is crowd-sourced. Although in common usage, the phrase "three is a crowd" suggests that a crowds are unwanted, crowd-sourced digital projects have much to be said for them. Crowd sourcing digital projects involves the humble admission that one person cannot do all tasks equally well and that many heads will have information and arguments that are not available to a lone head. The combination of humility and nobility ensconced (pun intended) in crowd sourcing may be blinding us, however, to its dark sides, by which I mean things in addition to ignorance, vandalism, and carelessness. Invitations to participate as a crowd member in digital projects are not all the same: a request to help proofread OCR'd materials, or to peer-review a scholarly essay, or to collaborate by adding new materials to complex editorial project sites are very different, requiring different skill sets from participants. One question to ask is: who constitutes an appropriate crowd? Beyond that I ask, perhaps rhetorically, is the concept of crowd sourced proofreading and peer reviewing based on the notion that we all have time to read and read again all of the essays and texts we examine in our studies--banking, as authors and as project conductors, on selfless community cooperation to make up for not doing the job right to begin with? Having read and collaborated once, must the crowd all come back to the finally-ready-for-prime-time version to read what might be called the real thing? Does crowd sourcing entail the notion that we are willing to leave to the crowd tasks we cannot be bothered to do ourselves--thereby trivializing accuracy, comprehensiveness, or even comprehensibility? Is it no longer true that what is everyone's job is no one's job? Although these questions worry me seriously, I raise them and lodge them on a back shelf, for now.¹

Crowd sourced intellectual endeavors have so much to recommend them that caveats raised about them are often dismissed as Luddite. In questioning the reliability of social scholarly editing I do not wish to suggest that we should not combine our efforts with others, including people we do not know; nor will anything I say be properly construed as an attempt to hold on to traditional methods--which, although holding high standards for accuracy, thoughtfulness, comprehensiveness, and production economies of compaction, were standards that allowed for notoriously complicated, slow, self-indulgent editions prone to eccentricity and error. I would not like that statement to apply to all print scholarly editions--just to those done by quacks--other quacks, I mean.

Scholarly Editing used to have a fairly precise definition. It applied to the investigation of the composition, revision, and publication of a single work, often in multiple variant forms, with the aim of producing an orderly history of the text and a reading text and apparatus that was reliable. Now it seems

¹ For an extended, related, discussion of error and efforts (both historical and current) to detect and remove them, see Paul Fyfe, "Electronic Errota: Digital Publishing, Open Review, and the Futures of Correction" *Debates in the Digital Humanities* <http://dhdebates.gc.cuny.edu/debates/text/4> 6 May 2013. Print: U Minnesota Press, 2012. See also chapter 10: "Ignorance in Literary Studies," in my *From Gutenberg to Google: Electronic Representations of Literary Texts* (Cambridge: Cambridge University Press, 2006), 189-99. (Misspelling in the title I take to be deliberate on Fyfe's part.)

to mean any kind of text production by anyone who is considered to be a scholar--hence, editing by scholars is scholarly editing. In particular the term has been used to apply to editing scholarly journals, to providing introductions and explanatory notes to classroom editions, and to digitizing collections of books. The last of these is in some ways the most interesting when one considers socializing the task of editing because its goal is to make searches possible for huge amounts of text--text collections so large that not only would it be impossible to vouch for textual accuracy, they are thought to be sufficiently redundant not to require accuracy. The simple act of producing with ocr a huge collection of texts, even though full of misreadings, is deemed useful because being able to find in a few seconds 1000 references to fish-mongering in 19th C periodicals is valuable even though 100 other references were missed due to OCR errors. Crowd sourcing the improvement of that kind of project is probably a good thing, but, having no experience in that field, I have nothing to say about it, except that if it is scholarly editing, it is a very different kind of scholarly editing from that which I do know about. My talk speaks only about digital scholarly editing aimed at creating the archive or edition of single works, where bibliography, textual criticism and scholarly editing are the methodologies in question--where the concept of "good enough" is generally abjured--where the object is to produce a body of virtual materials reliable enough for scholarly investigation.

So, then, what about reliability? I quote from a collaborative Internet project designed as a preservation repository for digital libraries:

"The partners ensure the reliability and efficiency of the digital library by relying on community standards and best practices, developing policies and procedures to manage content and services at scale, and maintaining a modular, open infrastructure."

That is to say, the partners have a disinterested (and probably also uninterested) crowd to back them up, but they have nothing to say about accuracy of texts and no up-front statement about the relation between virtual objects and the physical documentary objects they claim to represent. The statement of reliability claims basically: "this is as good as we can do". Reliable means, we've done what we can. Could one apply the quoted statement to LLBean or Land's End textile products--a field that relies on computers just as much if not more than we in the text business do. About their textiles could LLBean say:

The firm ensure the reliability and efficiency of our digitally woven fabrics by relying on community standards and best practices, developing policies and procedures to manage content and services at scale, and maintaining a modular, open infrastructure.

The textile business, of course, does no such thing. They hire inspectors to throw out any fabric with a flaw, just as we should throw out any verbal text with a flaw. In what sense can you as a scholar rely on a text with flaws? Can the terms "flawed" and "reliable" be applied to the same text at the same time? Of course if you knew where the flaws were or if you could recognized a flaw when you saw it, there would be little harm done. But which of you know without doubt whether Lily Briscoe, musing on her art in Virginia Woolf's *To the Lighthouse*, imagined a "transcendent beauty" or a "translucent beauty"? You could have recourse to the image of the manuscript, if you suspected a problem, but first you would have to suspect the problem. And then you too might misread the manuscript as did one transcriber. How often does that happen? We are talking here about what constitutes reliability, not how to achieve it. For that, let's turn to the crowd.

I do not need to add to or comment on discussions of the wiki model, the potential for vandalism, and ways to ring-fence or otherwise protect scholarly projects from malicious or inadvertent degradation. These are real issues, worthy of more attention than they have already received, but let us, for the nonce, consider that scholarly editing in the digital age is complex, has many separate parts, requires a multitude of skills, adopts a variety of useful, though sometimes incompatible forms, and serves a wide range of user needs and desires. If that is the case, and it is, it seems useless to talk generally about whether or not social scholarly editing can be reliable. Who amongst us has all the skills required? Who amongst us has hit upon the form for digital editions that is the true model form? Who has devised the best tool-set for the

construction of new editions? and, most important, who amongst us has devised the optimum fool-proof system for collaborative project content development and project enhancement that takes full advantage of crowd sourcing while also protecting projects from inadvertent damage? If there are specific and positive answers to those questions, I would want to declare the subject closed. But I don't think there are specific and positive answers, and that is what points up the (sometimes wonderful) eccentricities of the individual answers offered in print scholarly editions and raises the (potentially wonderful) eccentricities inherent in digital editions socially produced. In spite of the fact that scholarly editions are very complex, with many different working parts, and in spite of the fact that there are varieties of ways of conceiving what a work is that will be edited or what it is that is being edited--in spite of all that there are two certainties to keep in mind. First, it is not true that regardless of what you do, it is right. And second, it is not true that regardless of what you do, it is wrong. Some things are wrong, to be sure--such as saying one thing and doing another, or setting your crap detector in the off position. In addition, there are unresolved disputes about what is right. The Internet, like the ocean, accepts whatever is dumped into it. Oceans seem to have ways to purify or render harmless what is dumped into them; the Internet not so much.

And when it comes to digital editorial or textual projects, clarity about what is wrong and clarity about what makes what one does fall within the range of what is right cannot be harmful. I want to spend a bit of time on the range of modules that go together to make up a digital archive or digital edition and then make some suggestions about how and where crowd-sourced development and enhancement makes the most sense.

The Unbreakable Bits

In what follows I want to use certain words precisely and distinctly: the words "archive" and "edition" are sometimes confused with one another. The term "digital textual project" is also used for a range of things that are very different from one another. A "digital knowledge site" contains a surrogate archive, a critical edition, and a great many other things. And judging by what some folks say about scholarly editing, it has little to do with the history or accuracy of texts and is all about introductions, annotations, reception histories, and literary critical remarks or essays about works for which the text in hand (or on screen) is taken for granted as the text of the work. So, in dissecting the crowd-sourced digital editorial project, I want to do some naming of parts and assigning of roles.

A digital archive is a surrogate for something else. It stands in lieu of something. It gives digital access to materials that are related to one another but which, in the physical world, are often stored in geographically dispersed repositories, impossible to bring together for comparison. But digital access is not achieved simply; it is not a mechanical process. The digital archive aims to be a surrogate or stand-in; that is, as ARCHIVE it represents, not explains, materials. It provides surrogates for documents; it does not, as archive, analyze texts. An archive is not heuristic; it is a simulacrum whose first and most important function is to give access at any location to all the content of the physical archive or archives. It is a resource to be mined. Of course, the first persons (and alas sometime the only persons) who want to mine the resource are those who created the virtual archive. Given their position of power over the materials, it is often hard for them to keep their hands off. Paul Eggert says their modern hands are inevitably on, regardless of attempts to keep hands off, but there are degrees of involvement.² The deceptively simple remove, from the physical object source to the virtual object surrogate, is complex and fraught with potential pitfalls. The first of these is a popular notion that a digital transcription of a single physical item representing a work is itself the work. This is not a professional notion, for we know that any physical manuscript or printed text is one item, representing one manifestation of one expression of the work and that alternative expressions are manifested in other items all bearing the same title and looking, or sounding or reading so much alike that the popular notion asks, "What difference does it make?"

² In Eggert's contribution to this collection and in 'The Hand of the Present', *Textual Cultures*, 7.2 (2012; issued 2013), 3-19

rhetorically, expecting the answer to be "Nothing that matters."

The professional view is that a digital archive represents each of the items which research identifies as a significantly different form of a work. (Don't ask me to define "significantly different" here; that is to be determined by each project based on its examination of the materials and its goal in constructing the digital surrogate.) The digital archive's goal is to give access to all the significantly different forms so that users can investigate the archive in order to write histories, edit versions, and write critical assessments that are informed by the relevant historical facts. Professionally, our answer to the question, "What difference does it make?" is that variant texts represent a work variously, and it is, therefore, a good idea to know with which text of a work one is working and how it differs from variant forms of the work.

The digital archive is a repository of materials serving as surrogates for--please forgive me for this expression--the real things. It consists of images, which give the look of the original; and transcriptions, which enable analytical and manipulative functions. Together, images and transcriptions provide as good a surrogate for the real things as we currently can have.

That is to say, they provide a good surrogate to the extent that the images are of high quality and the transcriptions are accurate. If the images are not high quality and the transcriptions are inaccurate, then, I would say your surrogate is no surrogate; your digital project is worse than worthless because it will be trusted in matters you do not vouch for; by mounting it you pollute the Internet. These are the bits that cannot be broken. This is what cannot be left to chance. This is what cannot be launched before it is done: ready, signed, sealed, and worthy of reliance. My question is, who constitutes the appropriate crowd that will produce the reliable virtual archive? And if your project has texts, which you cannot be bothered to make accurate before launching them onto the mercies of a long-suffering crowd, your project may be social, but it will not be reliable, and it is doubtful whether it represents scholarly editing. Furthermore, accuracy is the sine qua non of textual virtual archives because the texts (plural) of a work form the basis of any additional investigation, insight, analysis, and commentary that might be conducted or added to the Internet site that hosts the archive. Inaccurate archives are sand upon which to build critical and explanatory sand castles.

What is the difference between an Archive and an Edition?

The scholarly representation and analysis of literary texts, whether in analog or digital media, has always had, three distinct categories of content: a) the material documents, both historical and new; b) the texts, both accurate and inaccurate or revised, that found expression and endurance in the material documents; and c) the scholarly analyses and critical interpretations representing engagements with the first two. Physically, the first two categories (document and text) seem inseparably intertwined, though texts can be reiterated and yet seem to be the same text but documents, when duplicated, seem obviously new and different. In virtual archives, the document is reduced to its visual aspects; and, in order to be truly useful digitally, texts must be transformed by transcription--which, incidentally renders them new editions of the work. In the material world, interpretive, explanatory engagement with literary works is, perforce, separated from text in the form of footnotes, endnotes, or apparatus or separate essays; but in the virtual world, critical engagements and enhancements are frequently embedded in the text in the form of encoding, to appear as pop-ups, drop-ins to text or as links which, famously and proudly, interrupt the linear with the radial, if not exactly radiant. It is important to note, however, that in the construction of scholarly textual projects, it is a greater sin to mix text and annotation in our storage systems than in our display interfaces. But that is an important question of design for which there are various solutions, not all as

elegant as others--and is not the primary subject here.³

The differences between an archive and an edition can be stated as propositions about these three categories: document, text, and analysis.

Archives purport to collect and give access to primary materials (documents with texts) inevitably mediated by interpretation at the transcription level, but minimally so--restricted to questions like, is this a comma or a period or a fly speck? is this word underlined or crossed out? or is this word actually illegible?

Editions are scholarly/critical arguments about what the archive means or about how the archive should be read. Do not confuse this argument with arguments about what the text or texts mean--that is literary criticism, not scholarly editing. But scholarly editing does involve the creation of critical arguments about what the archive means or how to read it. A narrative history of composition and revision justifies whether or not the texts are emended, and explanations are given for editorial interventions.

Analytical and explanatory commentary are often added to editions in ways that make it difficult to maintain the separation of the text and the analysis of text. It is worth keeping the distinction between text and analysis alive in the construction of virtual archives and editions, as I will endeavor to show. One quick way to do this is to note that historical authoritative text is finite and new critical analysis is infinite.

The first proposition--that the archive is unmediated--is untrue in some ways. Decisions about what does or does not belong in the archive reflect interpretive, critical judgments. Critical judgment as well as finances affect decisions about how many copies are needed of ostensibly identical copies of a work--editions, printings, or states of printed objects. Whatever texts they contain, an archive, as opposed to an edition, offers its ware in "as is" condition--to the extent that well-informed transcription skills allow.

The proposition about editions goes much farther. Mounting an argument about the meaning of the archive involves choosing some text of the work to foreground; and, choosing to emend or not to emend that text of the work is a decision that goes beyond collecting and preserving--it invokes a scale of values and a sense of the purpose of texts. Some editors claim they give only a guided tour of the archive, but guides always interpret the data for the tourist.⁴ Choosing not to emend is just as critical an act as choosing to emend; both require explanation.

For me, the greatest and most appropriate scope for crowd sourcing scholarship is in the third area: the analytical and explanatory commentary and critical engagement with works. Textual awareness is provided by the virtual archive's documentary and textual range and accuracy, but without many people engaging with the work, its nuances and anomalies and secret pleasures will remain hidden. What the

³ A more detailed analysis of these three components of a modular "knowledge site" is forthcoming in my "Development Principles for Virtual Archives and Editions" in *Variants*.

⁴ See Dirk Van Hulle and Peter Shillingsburg, "Orientations to Text, Revisited" *Studies in Bibliography* (forthcoming).

digital world does for literary scholarship is to give access at any computer or digital access point to the conversation, the well-informed conversation of crowds of persons seriously engaged with RELIABLE TEXTUAL EDITIONS and ARCHIVES which they will enhance socially and endlessly to the benefit of us all. They might, by the way, also find errors or gaps in the basic bibliographical and textual scholarship that created the archival foundation for everything else. And if they do, collaboration might lead to corrections.

But, the fact that text is finite and must be accurate, while analysis is infinite and cannot be definitive, means that if analysis is encoded and embedded in the text files, the text is vulnerable to inadvertent change every time analysis is added. Most of you have heard me tout stand-off markup as one solution to this (and other problems with embedded encoding) but that's not my subject now. I simply repeat what I've said before, embedded encoding is the enemy of collaboration.

Scholars and students, to the extent that they rely uncritically on scholarly editions, are really just tourists visiting the edition or archive site casually or as scavengers picking up this and that to carry off. It is, therefore, one function of digital textual archives to provide tourists with a rich experience. I suppose one could say that a tourist visiting a well-organized theme park in which all the rides work all the time have a rich experience. And tourists sometimes invoke such experiences in order to judge the entertainment value of a museum or, in our case, a textual knowledge site. It should not be difficult for them to navigate the site. The site should not assume that documents and texts explain themselves. But if we are building theme parks with high entertainment value, we should remember that when visitors go to a petting zoo or to a tourist gold mine where they can pan for gold in salted sands, they know in their hearts that it is just entertainment. Biology departments do not take field trips to petting zoos; and the school of minerals does not send students to tourist gold mines. Although it is natural for most scholars and students in much of their engagement with literature to be tourists, reading for pleasure or background information, from time to time they need to be focused in a serious way and will want answers to questions. They require a belief that what they are relying on is actually reliable, not just pretty. In order to serve as the surrogate basis for such investigations, the virtual archive or scholarly edition must provide accurate materials with forthright explanations of the principles for document collection and representation employed.

Two more thoughts: If we replace the word Reliable in my title: Reliable-Social-Scholarly-Editing with the word Useful, and combine it with the warning: "to be used with caution because digital textual projects are never more than a convenience, allowing one to do vast amounts of preliminary study before visiting the archives of the real thing to verify one's findings" then I might relax a bit and get off my idealistic high horse. But either way, as scholars, we must not replace hard-nosed admissions about the limits of our accomplishments with hopeful wishes.

I end by recalling the old saying about the desirables in almost any project of importance: We want it quickly, cheaply, and of high quality. In scholarly editing, whether digital or not, whether done by an individual or a crowd, one can still have only two of those at a time: if it is cheap and fast, it will be of low quality; if it is of high quality, it was either slow and cheap or fast and expensive. Look at the Bichitra digital archive of the 80 to 100 works in multiple forms by Rabindranath Tagore completed at Jadavpur

University in the space of about three years, with a team of thirty or so dedicated transcribers, programmers, and editors, supported by a huge grant from the Indian government.⁵ Compare that to the already six years we have spent on the WoolfOnline knowledge site for Virginia Woolf's *To the Lighthouse*, working with three part-time editors, one part-time programmer, and a series of part-time research assistants.⁶ It is comparing apples and oranges to compare the digital archive of all the works of one writer with the full-scale scholarly digital knowledge site for one work by another author, but the principle demonstrated is that, if you want reliability, the process is either slow or expensive. The cheap quick fix is for tourist traps.

⁵ Watch an introduction at <http://www.youtube.com/watch?v=GRNT9pf-sWA&feature=youtu.be> or visit the archive at bichitra.jdvu.ac.in (accessed June 16, 2013).

⁶ www.Woolfonline.com (launched 1 Oct. 2013), using the same URL previously used for the pilot project.