The Grub Street Project: Imagining Futures in Scholarly Editing*

Allison Muri

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

External Image
Please see:
http://rup.rice.edu/image/cnx_images/shapes-buybutton.jpg

1 The Grub Street Project: Aims and Objectives

Imagine an "edition" of eighteenth-century London, where a single page as zoomable map provides the interface for "reading" the city, its communications, its economies and texts, its literature, history, architecture, art, and its music. Imagine a new way of sharing a scholarly edition in the digital environment, not as a single annotated e-edition, but as an expanding library of selected and topographically encoded, searchable books, maps, and prints. Still very much in an early stage of development, the Grub Street Project aims to create a system to assemble and map the topography, publishing history, texts, and people of eighteenth-century London. The project is intended to create an open-access collaborative space where students, scholars, and members of the public (e.g., genealogical societies, high school classes, or gamers inventing a new space to imagine D&D style narratives) can add their own annotations, literary mappings, e-editions, or digitized documents to the infrastructure. The digital infrastructure will include a number of maps of London from 1720 to 1799, including John Strype's *Survey of London* (1720), widely understood to be the first authority on the history of London and its topography, and Richard Horwood's *Plan of the Cities of London and Westminster* (1799). Horwood's map, intended to present a complete view of the city with every individual property, street, and alley, will serve as the main interface, while the others will provide comparative views of the city over time. Horwood's thirty-two–page document is reconstructed as a single "zoomable" map that will be associated with data, including 11,000 place names and alternates, plus 5,300 place descriptions from the complete text of the public domain *Dictionary of London* (Harben 1918); addresses, trades, and tradespeople from Kent's London business directories published annually from 1732 to 1828; bookseller and printer locations derived from bibliographical data; and links to full-text e-editions of books, pamphlets, broadsheets, images, and maps printed and sold in the city.[1]

In terms of scholarly goals, by applying topographical markup to a set of digitized maps, texts, and images, the project aims to provide a means to search, navigate and visualize trends and relationships between material contexts (such as networks of print distribution, literary commerce and other trades, or particular historical events). This will allow us to read and visualize the history of print culture in this particular space.[2] But as a literary scholar I am also interested in the "immaterial" London: the London that is in the imagination of its citizens and visitors, the London that is both real and metaphorical topography

*Version 1.2: May 14, 2010 6:01 am -0500
†http://creativecommons.org/licenses/by/3.0/

depicted by contemporary writers such as Alexander Pope in *The Dunciad*. Maps can also help us to investigate how that space is represented by imaginary topos, for example a Dulness that is impersonated in localizable points such as Bedlam, Fleet Ditch, or St. Mary le Strand in Fleet Street and also lies like a cloud over the entire city. The utility of this concept for the study of literature and its spaces or topographies is that, as with Google Maps, annotated maps are not merely geography: they articulate the culture of places. By re-presenting the history of London as a network of literary communications, ideas, and physical-spatial relationships, by visualizing it as a heterotopia,[3] localized in maps of "the real" but simultaneously represented, contested, and inverted by literary metaphor and ambiguity, we can gain new understanding of the city and its literature.

In terms of the project's goals that are less "scholarly" but more crucial overall, this project aims to promote public access to scholarship and public domain documents. The site and the editions contained in it will be licensed under a Creative Commons Attribution-Noncommercial-Share Alike license that will enable the digitized documents and editions to be accessed and used freely. Making digital editions usable, accessible, readily annotated, and free of restrictions will help to ensure sustainability for non-commercialized projects. Perhaps the most important issue facing scholarly use of digital texts—possibly more important than TEI-compliant markup, machine-readable texts, digital tools for collocation and concordance, etc., as valuable as they are—is copyright, and the public access to high-quality, carefully edited, digital editions (whatever the future structure of digital "editions" might be).[4]

2 Copyright and Sustainability

Copyright issues are not new to digital projects, but the atmosphere of restrictions and controls over these materials is much more vigorous today. When scholarship is owned by commercial interests, we risk losing our ability to participate in our own discourses. Copyright has been, mostly, a fair means of protecting authors', publishers', and scholars' interests in the world of print. However, everything changes when digital technologies render all communications as copies. This situation has resulted in sometimes excessively restrictive user agreements written for companies that have built their business models on distributing books and journals in the age of print, when it was harder to copy and re-use words or images.

As an example, consider the legal notice for the *OED Online*, which reads as follows:

> *you may not ... systematically make printed or electronic copies of multiple extracts of OED Online for any purpose; ... display or distribute any part of OED Online on any electronic network, including without limitation the Internet and the World Wide Web (other than the institution's secure network, where the Subscriber is an institution); ... use all or any part of OED Online for any commercial use.*[5]

Merriam-Webster Online similarly limits the use of its text:

> *No part of the work embodied in Merriam-Webster's pages on the World Wide Web and covered by the copyrights hereon may be reproduced or copied in any form or by any means—graphic, electronic, or mechanical, including photocopying, taping, or information storage and retrieval systems—without the written permission of the publisher.*[6]

The commercial and legal motivations for such notices are understandable when the tools for copying and publishing are ubiquitous and easily employed; however, a dismaying trend here, by no means unique to these publications, is the explicit overriding of any principles of fair use or fair dealing. Aside from the obvious vagaries of terms such as "systematic" or "multiple," readers and writers are apparently forbidden to use *any* portion in an online work, in a commercial work, or in the case of Merriam-Webster, in any work whatsoever. An added irony is that a significant proportion of these dictionaries has been compiled from a long and profitable tradition of stealing, pilfering, and fair use. If one compares, for example, the definitions for *theft* provided by *OED Online* and *Merriam-Webster Online* to those from seventeenth- and eighteenth-century dictionaries, it becomes clear that each definition draws heavily upon and uses the same words as

its predecessors.[7] The *OED*'s "felonious taking away of the personal goods of another; larceny" comes almost unchanged from early printed dictionaries: Thomas Blount's *Nomo-lexikon* (1670) defines it as "Felonious taking away another mans moveable and personable Goods.... See Larceny and Felony," while Ephraim Chambers' *Cyclop* (1728) has "felonious taking away another man's moveable and personal Goods... See *Larceny*." Clearly, the copyrighted definitions that we may not freely use (if we were to obey the legal license in its strictest sense) have been appropriated from the public domain and potentially from fair use or fair dealing.

The issue here is not whether these publishers will allow or ignore quotations of excerpts (there is every reason to assume one could quote from the dictionaries in an online publication without fearing repercussions): rather, any author or editor of digital texts who needs to publish words from texts should be concerned that these licenses appear at all, claiming ownership of words that should be in the public domain, however insignificant a threat to scholarship they might seem in practice. As Howard Knopf, IP lawyer and Chairman of the Copyright Policy Committee of the Canadian Bar Association, has warned,

> *The better publishers will have to adapt [to the increasing use of open access texts in education]. The less adaptable will probably resort to litigation and, in Canada, greater reliance on collective licensing—even when there's little or no basis for it. Don't [expect] Access Copyright in Canada to go gently into the night as "open source" electronic resources in the educational and other sectors threaten their photocopy based foundation, which was never very solid anyway.*
>
> *One great worry, of course, is that the promise of better and...cheaper text books and resources could be turned into an Orwellian nightmare if DRM is deployed in ways that could allow for censorship, revisionism, "memory hole" deletion, and other means of control by state or private interests. Other means could include the prevention of fair use (fair dealing in Canada), the prevention of cutting and pasting, the prevention of "read aloud" features (as Amazon has also recently done), the prevention of access to the public domain, and other excessive exercises of copyright.*
>
> *The recent Kindle fiasco shows that all of this ... is not only possible but probable.*
>
> *This once again shows how we need protection from DRM much more than we need protection for it.*[8]

Knopf's argument emphasizes the imperative for those of us working in the digital humanities, in online research, and in education, to be aware of and participate in formulating the practices of online publishing, scholarship and education.

Also problematic are commercial databases such as Gale's *Eighteenth-Century Collections Online* (ECCO) and Chadwyck-Healey's *Early English Books Online*, commercial enterprises that have taken images of public domain texts and claimed copyright, thus limiting their use in digital editions and research. I will preface this commentary with the acknowledgment that my work has hugely benefitted from access to both EEBO and ECCO, and their involvement with the Text Creation Partnership (TCP) program will be of tremendous value to participating libraries. However, as an educator and a digital researcher, I have serious misgivings about Gale's Terms and Conditions:

> *The subscribing institutes ("Customer") and their authorized users, may make a single print, non-electronic copy of a permitted portion of the content for personal, non-commercial, educational purposes only. Except as expressly provided for in the foregoing sentence, you may not modify,*

[8] Howard P. Knopf, "DRM and the Demise of Textbooks?" *EXCESS COPYRIGHT*. http://excesscopyright.blogspot.com/2009/08/demise-of-textbooks.html. (<http://excesscopyright.blogspot.com/2009/08/demise-of-textbooks.html>)

> *publish, transmit (including, but not limited to, by way of e-mail, facsimile or other electronic means), display, participate in the transfer or sale of, create derivative works based on, or in any other way exploit any of the Content, in whole or in part without the prior written consent of Gale and (if applicable) its licensor.... In the event of any permitted copying, redistribution or publication of the Content, such use shall be for personal, non-commercial, educational use only and no changes in or deletion of author attribution, trademark legend or copyright notice shall be made.*[9]

Normally, creating a derivative work based upon a facsimile of an eighteenth-century text without needing to ask for special permission would be a right in any country that honours public domain, but the preceding agreement would seem to preclude such use.[10] At the least we might be concerned about workflow, having to make a special request for each derivative work. And while it is highly unlikely that Gale would deny any scholar the right to create a derivative work based on these resources, clearly the ease of copying and publishing online has resulted in more restrictive user agreements than books ever inspired. The agreements for accessing commercial databases engender a certain caution about actually using the large volumes of material that are available by subscription. Databases represent important and powerful sources of volumes of information, and in terms of literary studies, enable for the first time the possibility for many researchers of conducting what Franco Moretti has termed "distant reading" of literary texts.[11] But this analysis is impossible if we cannot access, download, and manipulate large bodies of data.

For example, I have been told by a librarian at my institution that printing or downloading a non-substantial portion of a database "(an article, a chapter, a results list)" is permissible, but "if repeated over time with the purpose of duplicating and archiving a substantial part you are probably morally infringing... The vendor assumes that researchers acting in a normal scholarly way in relation to e-resources will not infringe. For example, harvesting or using robots or spiders to download/copy would definitely not be normal research," and that in any case "As a researcher you are not inconvenienced since the database is available online." On the one hand, these instructions are by no means punitive or extraordinary, but what if a scholar's (arguably quite "normal") digital research concerns reading, conducting OCR, and running analysis tools on a large set of documents offline for a systematic study of a large body of works, or is dependent upon gathering facts (which are not copyrightable) from a database, which may or may not be protected by copyright?[12] These restrictions create an imperative for digital scholars to build new databases of early modern materials with an open access license. The Text Creation Partnership (TCP) with EEBO, ECCO, and Evans offers one exciting possibility, as the license eventually "allows scholars to use texts in their entirety to reproduce or create new editions."[13] However, it is unclear whether this agreement would allow for truly releasing all the texts into the public domain. Does public access to the texts actually mean that users could download the set of 25,000 texts in its entirety? Public access is one thing, but full access might well be limited by page views, expiring sessions, and export limitations to a certain number of records, pages, or documents at a time, as various online resources have shown[14]. My institution's library, a partner in the Early English Books Online—Text Creation Partnership, specifies on every page linking to a database such as EEBO, ECCO, or ESTC that "Systematic copying or downloading of electronic resource content, including the downloading of a full issue, is not permitted by Canadian and International Copyright law."[15] However, the Canadian Copyright Act [C-42] does not refer to systematic downloading or to database rights, and though the library cautions that every user is responsible to be aware of copyright, there is no information available about how much downloading is considered systematic, and which International Copyright law forbids it.

The above is not meant to be a criticism of our librarians: they are protecting the interests of their patrons, and rightly so (though perhaps, due to pressure from vendor agreements, leaning more to limiting access than to freedom of information in this regard). Nor is this a criticism of ESTC, which has generously made the catalogue available online, or the TCP, which was initiated with the laudable goal of eventually allowing libraries to release the resulting textfile to the public domain. These are commendable initiatives.

[11] Franco Moretti, "Conjectures on World Literature," *New Left Review* 1 (January–February 2000). http://newleftreview.org/A2094.

As Mark Sandler, Collection Development Officer at the University of Michigan University, explained in 2003:

> *as co-owners of the text-file, a partner library has the right to open its servers to nonauthenticated users to permit access to the text files. By contract, five years after the completion of production, partners can freely distribute some or all of the texts to neighboring high school students, state residents, or world-wide. Since ProQuest owns the EEBO page images, those will not be freely available, but the texts are useful in themselves and public domain access has been protected through the TCP licensing agreements.*[16]

Neither this explanation nor the current TCP website is entirely clear, however, about how this will work, particularly, how many text files can be accessed, downloaded, and published elsewhere by individual users. According to Aaron McCollough, TCP Project Outreach Librarian, "The first 25,000 encoded texts are set to become free in January of 2015. If we are able to raise enough revenue to encode the rest of the collection (roughly 44,000 unique items), then the entire EEBO archive (as TCP text) would become free in 2020."[17] Work such as this has tremendous potential for new modes of digital scholarship and public access to that scholarship, and if it is the case that the entire dataset is "free" in five years' time, there are many reasons to be excited, and many more reasons to start planning how they might best be integrated into digital projects so that the many projects currently online, and those created in the future, could attract readers, interact with one another, and forge new communities of scholarship and general interest.

It is also important to recognize a potential impediment to conducting digital research and publication on individual projects, specifically, that user licenses have shifted expectations and rights that we have come to expect as readers and writers. As Howard P. Knopf has noted, regarding so-called "use" rights,

> *If I buy a book, I need neither further permission nor further payment to read it, or to quote or copy short or insubstantial excerpts from it, for any reason, or even longer portions, if such copying amounts to "fair dealing." I can resell my copy of the book, lend it, rent it, put it on reserve in the library, copy chunks of it for interlibrary loan, and, generally get my money's worth from it and share it with others without further worry. If my law firm buys a copy, everyone can share it. Since the time of Gutenberg, this is how the world has gotten smarter.*
>
> *Now, if someone puts that book into electronic format, all of a sudden everyone gets excited and worried and wants to create new rights and limit old uses.... it seems that publishers are trying to move users into electronic media with high initial acquisition and/or high per-use costs, with little or no cost savings over traditional paper versions and, in some instances, an overall cost premium. Make no mistake. A use right is the dream of any intellectual property monopolist, and it is a quantum change from intellectual property law as we have known it until now. It is not merely a preservation of previously held rights.*[18]

Published a decade ago, these comments are still relevant to some of the most crucial challenges facing digital scholarship: untangling the rights to freely use and publish public domain texts (and images of them) in digital databases.

While many important projects already mentioned have begun this process, there is still much for literary scholars to do, especially in terms of re-imagining what an "edition" is or ought to do in a digital environment. A scholarly edition need no longer be imagined as a single text; it might be something closer to an archive or a system of archives. Assuming that the texts truly are released into the public domain, a potentially useful source for page images is Google Books; however, while out-of-copyright books are available to freely search, link to, or download, the image quality is low and editorial/curatorial selection and bibliographical data have not been priorities: there are no filters, for example, that limit search results to items catalogued in the ESTC. Furthering the general uglification of e-books, subsequent use or distribution requires that the Google watermark remain on every page. The tremendous good that is the result of Google Books for early modern

works is that it has released these images of public domain texts for the public to freely use. The Google Books project, like the work of Wikimedia in asserting public domain rights, will benefit all scholars and readers, insofar as their unrestricted availability on the Internet establishes a precedent for public access.[19] However, as Robert Darnton has argued, we are facing with Google Books "a fundamental change in the digital world by consolidating power in the hands of one company."[20]

One key to sustainability for non-commercialized digital projects might have to be a willingness by textual scholars (including those who do not consider themselves "digital humanists") to negotiate the rights to study and analyze, to quote from, and to re-distribute digital documents—whether as faithful replications of works now out of copyright or as quotations from recent digital texts, recordings and movies. Projects such as the Internet Archive could to a great extent alleviate some of the problems with access to images of historical documents, though opaque organization of early modern materials, limited finding aids, dirty OCR, and lack of editorial apparatus or commentary presently make it difficult to find, read, and use the archived works.[21] In any case, partnerships with or links to open access projects such as the Internet Archive or Wikipedia in addition to ECCO, EEBO, and other proprietary databases seem to me crucial for establishing the longevity and accessibility of large digital projects. For these reasons, the Grub Street Project aims to explore some of the new possibilities inherent in creating and publishing a scholarly digital edition/collection that is, in short, as *usable* as possible for as many readers as possible.

3 Usability, Technology, and Community

At the outset, determining the potential for integrating an individual project with other collections and projects at the global level will ensure a project's longevity, but my first concern before developing this strategy has been attending to decisions about software, programming, and markup. TEI is the gold standard for scholarly digital texts, but these encoded documents are mysteries to most of their users, and still end up being transformed into HTML and CSS for display in a web browser. Perhaps for this reason, scholarly projects for the most part can only dream of the tremendous volunteer participation by the free culture community all over the world in projects such as Project Gutenberg, Wikipedia, or Creative Commons. Here again, I turn to accessibility and usability: the first collection of free electronic books, Project Gutenberg, founded by Michael Hart in 1971, are distributed in "Plain Vanilla ASCII," HTML, and a few other experimental open frameworks. They are still both accessible and readable on any operating system today, and augmented by other popular and usable formats such as HTML. Most important, especially in terms of ASCII and HTML, they are easy to make. The first web page, written and published by Tim Berners-Lee in 1990, is still both accessible and readable on any browser on any of the major operating systems today,[22] and could be easily written by anyone today in a text editor or a more advanced WYSIWYG HTML editor, readily available in a multitude of free and commercial versions. Scholarly works focusing on the more rigorous process of markup in SGML have occasionally fared less well. For example, the proprietary e-book software DynaText is now out of business, and books in this format cannot be read on recent Macs unless they happen to be running Windows. I am not saying anything new or remarkable here, but digital editions and projects published in any format that is not open source (modifiable by current and subsequent developers of the editions, and unencumbered by DRM locks), or at least readily accessible and compatible with the majority of browsers or e-book readers online today, will not be sustainable unless the developers have the means to continuously update and rewrite the edition. According to Google, "PDF formatted files are the most popular after HTML files,"[23] clearly suggesting that usability and availability of both production and reading tools will encourage the creation and use of digital documents. Publication of digital editions and text that are not open access will also limit their use and their contributions to public knowledge.

Another major difficulty, then, for digital projects is establishing long-term access, ease of use, ease of contributions, and long-term compatibility of the interface with web browsers. Should I gamble on any scripting language or display that is not the most ubiquitous of current web technologies? Will I need to reconsider how my maps, jpg tiles displayed via Adobe's proprietary Flash technology, are accessed ten years from now, should Flash be superseded by more elegant software that will work on all e-book readers? (Imagine a scholar in the past having to decide, in addition to producing an intellectually challenging and

original work, what is most likely to be the most sustainable medium for publication: who ever had to wonder, Will my book open in an old house? Will it open in ten years if its owner moves to a new house? Will it open faster if I change the book's binding in five years?) While my interface, in HTML and delivered from a MYSQL server, and my data, output from the server as XML files, are assured to be readable for many years, the "Zoomified"[24] maps that I have created as one access point to the digital edition of London will need to be constantly monitored and tested as Flash develops over time.

The next important issue concerning sustainability, and one, I think, that is less attended to, is determining how readers use and interact with the digital texts we create. The success of both Project Gutenberg and Wikipedia in continuing to attract so many volunteer contributors is not only because they contribute so much to public knowledge, but also because they are built on unremarkable technologies that are easy to use. I can teach my undergraduate students how to edit a Wiki in just a few minutes. Creative Commons licenses, both human-readable and readily applied to works easily published on the Internet Archive,[25] have been applied to more than 130 million works since CC was founded in 2001.[26] I will need to know how many users of my interface are frustrated and give up because they cannot find what they are looking for / do not have the appropriate version of a plugin installed or enabled / cannot navigate a complex interface, etc. At the minimum, readers of online scholarly works should be able to:

- access texts online from any computer;
- copy, paste, and reformat text freely;
- download / print documents in their entirety (not only page by page or limited number of pages);
- move the text easily and freely from one reading device to another;
- experience pleasure in encountering the design, typography, and layout of the text;
- conduct both simple text searches and Boolean searches across multiple texts;
- make their own links between and within texts;
- annotate texts, and share their textual annotations and related images or movies easily (as, for example, Flickr, Diigo, YouTube, Twitter, etc.);
- share their adaptations of texts.

More sophisticated users will want to apply text analysis tools to digital texts, but these texts should also appeal to those who, apart from keyword searching, for the most part read them in a linear fashion or annotate them with few other demands than typical word processing. Ideally, these texts should also appeal to, and be adaptable to, Web 2.0 editing procedures—though I am certainly conservative enough to argue for means of filtering these annotations by peer review or by group (e.g., a particular genealogical society might add highly useful documentation to an edition of London, but neither require or want scholarly peer review; a scholar seeking tenure or promotion might submit a set of annotations and materials to a peer review process. These groups or review tags are easily added, and could be readily filtered from a set of results to exclude, say, "My favorite early modern zombie mashups"). Arguably, any digital text should aim to be as usable as, and more useful than, a book: to be as usable, it should not need instructions for opening, paging through, searching, or bookmarking; to be more useful, I think (to be terribly opinionated here), means scholarly texts should be available for categorization and annotation in the wild. These, at least, are my own aims, and certainly the most successful online digital projects share many of these characteristics, other than the relatively new prospects for Web 2.0 (e.g., the William Blake Archive, the Rossetti Archive, the Proceedings of the Old Bailey, the Perseus Project).[27] Ideally, usability testing (especially focusing on prospective readers who are not digital scholars) would be ongoing throughout and beyond the development of any digital project, though like every other aspect of digital publishing there is little or no institutional infrastructure to support and reward such activities. Questions we need to be attending to are: how vital is highly structured TEI in these early stages of preservation and sustainability? Do we know how many readers/users make advantageous use of Boolean search techniques, full-text proximity searches, or wildcards? How many searches benefit from painstaking markup (now and years from now when eventually intelligent programs will surely recognize contextual indicators of "Paris" as place versus "Paris" of Greek mythology)? How would we make these searches more effective? (Drop-down menus are one method that has been tried, but they can unnecessarily limit the search options.) Elaborate markup and data mining offer tremendous

advantages, but is this the optimal way to encourage *all* readers to use and participate in digital projects? Thus, three significant issues that a project such as my own should address, where the intent is to build and test a system that will engage and involve students, scholars, and the public in online research, publication, and discourse: first is scholarly credibility; second is usability; third is community.

As Jerome McGann reminded us in his recent article, "Our Textual History," there are "two very broad obligations that scholars have by virtue of their vocation as educators. We are called to surveille and monitor this process of digitization. Much of it is now being carried out by agents who act, by will or by mistake, quite against the interests of scholars and educators—and in that respect, against the general good of society. So we must insist on participating."[28] However, the challenges are many, ranging from the intellectual complexities and presumptuousness of any small individual project by a new scholar imagining new ways to edit, read, publish, and share scholarly editions, to getting funding at the national and international level, to the particular administrative assumptions and infrastructures at the departmental and university level. What is involved in making this happen? The following pages will assess some of these challenges aside from the issues of copyright I have been discussing, particularly in the context of a new scholar working in a mid-sized university in Canada.

4 Preserving and Sustaining the Life of Digital Projects: Institutional Obstacles

The University of Saskatchewan has been active in digital publishing and research initiatives since the 1990s. However, the economies of doing this kind of work have meant that, until recently, we have remained a number of independent scholars building standalone projects or editions that have only occasionally been peer reviewed, and are housed on the English Department webserver space. These early projects were done on PCs in offices with the minimum of necessary hardware—student-owned scanners did at least some of the graphics work—and they were completed on shoestring budgets. As the digital humanities have become more complex in orientation, the need has grown for material infrastructure, programming assistance, and continued revisions and updates to the resources as web standards change. Ongoing efforts over the years have resulted in a number of important initiatives, notably the creation of the Digital Research Centre (DRC), a space nominally reserved for research, with a projects room, a performance room, and three A/V booths.[29] Canada Foundation for Innovation (CFI) infrastructure funding has helped to ensure that the space, in much demand, is truly reserved for digital research and training. The DRC was first imagined and promoted by Peter Stoicheff, now the Vice-Dean of Humanities and Fine Arts, in the 1990s. Without his ongoing commitment and efforts to educate university administrators and promote the potential for such a space, it would not have been achieved. Others, of course, were invaluable to the creation of the DRC, now operating at capacity. The financial support of the university and the College of Arts and Sciences in dedicating the space to digital computing has meant that researchers now have access to good equipment, some limited programming assistance if they have funding, server space, and a community space. However, the centre and its staff are likely to continue to be dependent on external funding and will need committed endorsement and financial support from the university.

Turning back to the sustainability of the individual project, I will describe my own experiences to date, which indicate that institutional support, both financial and administrative, as well as federal support (in terms of funding) are still very much at a new stage and will continue to benefit from further education, planning, and development over the next number of years.

Jerome McGann begins "Our Textual History" with a question: "Why does textual scholarship matter?" The question of how (or whether) textual scholarship matters is playing itself out in my own university, where its importance is apparent but not well understood by administrators who can clearly see the value of medical science, ecology, public policy, and so on. We are currently advertising for a prestigious Canada Council Tier I Research Chair for outstanding researchers acknowledged by their peers as world leaders in their fields. This year our Digital Textuality proposal was well received by the university's oversight committee, but the committee decided to make the competition ultimately between applicants for the position in three areas rather than advertising in a single area: thus the university is oddly interviewing for a single position that will be in Digital Textuality, Water Policy, or Innovation Policy. In some way or other, a committee

of members from a variety of disciplines will decide on the relative merits of one over the others. So, a very interesting question will be answered here, I suspect: how "valuable" is text compared to water or to business? An unfair question, perhaps, but it points to some of the challenges when it comes to educating our colleagues and others about why textual scholarship matters.

This is not at all to say that the university here is not supportive of digital humanities scholarship, but we have many challenges as we build not just our discrete research projects, but also a systematic framework for support. Identifying and communicating needs is the first. There are three key groups at the university level that would ideally develop a system to foster successful digital scholarship and long-lived projects in the digital humanities: (1) for the intellectual content and long-term oversight, the scholars intent on conducting and disseminating their research in a digital environment; (2) for the long-term storage and serving of data, Information and Technology Services (ITS), the unit that supports research and educational IT infrastructure on campus; and (3) for the long-term cataloguing (and potentially for archiving digital works), the library. The case here at my institution is not a lack of support or enthusiasm for digital humanities projects at the top levels of administration: indeed, it is through the efforts of support in the Dean's office in my college that I have been fortunate to not only have access to a dedicated space, resources, and assistance in the Digital Research Centre but also to startup funding (normally considered less essential to the humanities) as in-kind contributions toward external funding. And while the funds available are significantly lower than generally available in the sciences,[30] it is a good start.[31] However, what we face is a matter of developing nonexistent policies and systems for continued financial, material and human resources. Every aspect of digital research seems to come down to a matter of education and building, long before one gets to actual research. The startup grant provided by the university, for example, required a reversal of decision by a top-level administrator, since my application had been turned down in the apparent assumption that a startup grant of $12,000 was unnecessary to literary research. It will take some time to establish a milieu where it is well understood that literary scholars will need more equipment than a word processor, a few books, and some chalk!

Outside of the process of grantsmanship, determining the foundational level of support that can be accessed and built upon by individual projects has proven to be a challenge. College IT, which has a share in the DRC, is a separate entity from university ITS, each carries out different roles and has different regulations concerning their activities and ours. These are not always clearly defined, and sometimes that means that the initial response to a request—"No, you can't do that"—is not the only response available to a researcher. Part of the process of establishing this work at an institutional level is, first, communicating with university ITS about specific needs that go beyond the somewhat limited services currently available and second, establishing clearly evident roles and responsibilities of the various units and staff at all levels of the university.

According to Dr. Rick Bunt, Chief Information Officer and Associate Vice-President, Information and Communications Technology, research support for digital projects in the humanities is an area where ITS needs to develop a better understanding of how to provide support. Support for research is a high priority for the next two years and, he explains, it is not merely "providing more cycles and getting supercomputer access." He explains that ITS needs researchers to articulate their needs and contribute to the development of a foundational level of support infrastructure committed to research in the various areas of information technologies. Data storage, he says, can be accommodated easily and cheaply, but how long we keep the data and how we organize it still needs to be determined. When I suggested I want it to be accessible and running forever, we both laughed. But longevity is what I want, and one aspect of the process seems to involve repeatedly explaining to those outside of the humanities what "data" is to us. Unlike other disciplines, the texts produced in the humanities are not immediately superseded by new discoveries. We expect any text or project we produce to have a long life, and this means that, *in addition to* support and infrastructure for building and storing these new forms of editions, we will need to have access to means for upgrading and maintaining them long after any initial funding has been depleted. "Curation," Bunt suggests, will be the responsibility of the scholar; I suspect it should be a joint responsibility of the library and the scholar, but the library too will need a significant investment of resources and infrastructure if it is to ultimately host digital projects over the long term.

The university has made definite improvements over the last ten years: at one time an individual scholar was not allotted enough server space for a single edition of scanned pages; currently, for a class or small project where the demands are not too high, the DRC server can meet basic requirements (PHP is installed, and 5–10 GB of server space is readily available). However, the ordinary needs that are normally quite inexpensive and readily available through a commercial provider are prohibitively expensive, cannot be managed by the researcher due to security constraints, and require wait time for assistance in a university setting (i.e., larger amounts of free server space are not available: due to the requirements imposed by external funding, the DRC cannot provide for free what other projects pay for through particular grants; hosting a domain is unusual, creating a database and adding users, creating ftp accounts or new directories, providing levels of access for different groups, a programming environment that as a matter of course would allow researchers to develop with Perl, ASP, .Net, JSP and Ruby on Rails, etc.). I pay about $2 per month for hosting my site on iWeb, a commercial service in Quebec, with these capabilities built in. To get access to a virtual server in the DRC, with some of these capabilities, would cost $500 per year. I am fortunate to have some of this support from ITS in the form of server space, a domain, and MYSQL databases which I have access to login and manipulate myself, in part because I am known and trusted by the webmaster, but inquiries about hiring programmers from ITS have taken months for a response. This is understandable, albeit frustrating. They have many priorities and are in much demand: one small project with limited funds is unlikely to be at the top of the list.

Ideally, we need a separate server for experimental work so that there is no potential for a security hole on the larger system. We need an institutional repository that is not based merely on archiving electronic versions of printed articles and books (as important as that is). Even at the most basic levels, we need a reliable capital funds allocation to buy and upgrade servers and desktop computers, and (however unlikely, given the economic constraints all universities are facing right now) hire a dedicated programmer / sys admin whose job description is to support faculty research. Presently, each digital project is required to provide its own funding for markup and programming. This has often amounted to hiring students in the past, but without a dedicated research support programmer / sys admin, smaller projects without access to external funding must be planned to require little by way of updates to code; they are also jeopardized by the fact that programmers and student assistants hired for the short term will move on, leaving a researcher (sometimes with little expertise in the necessary techniques) with the problem of not being able to easily upgrade or modify the existing project.

I have been fortunate to receive startup infrastructure funding through a Canada Foundation for Innovation (CFI) grant to create a Computing and Media Studies Research Lab within the DRC, which has been both enormously beneficial and enormously time-consuming. Aside from the very complex and demanding application process that has generated roughly eight hundred email messages between me and various constituents at the university, I find—even with assistance from research services—that research time is eaten up with budgeting; determining, or negotiating, with various levels of administration concerning who does which job (one cannot, I find, purchase a camera without the involvement of several separate units); finding programmers who can actually do the work and will answer my emails; filling out and photocopying forms; justifying my expenses and sending identical documentation to both Research Services and Financial Services Division, and so on. Because the CFI traditionally funds innovative research in the sciences and social sciences, applying for funding in the humanities and fine arts has entailed a huge learning curve both for applicants and research services. One most memorable exchange arose from the stipulation that books are not an eligible expense for infrastructure. Because I wanted to scan original eighteenth-century editions and maps, I found myself arguing that, for the purposes of my studies, these books could be considered "core data," which is an allowable expense. The example I gave was Strype's 1720 edition of *The Survey of London*. Given their formal structure, with "tabular data" (consistently organized short entries categorized under headings and subheadings) the textual entries in this text, I argued, are in fact an early form of a database: for example, each subheading corresponds to what I have designated a "placename" field, and each description corresponds to a "notes" field. I am very grateful that Research Services and CFI conferred on this to ultimately conclude that books in this case could indeed be considered data. These issues, humorous in retrospect, show how much negotiation and education must go into the inserting of literary scholarship—

traditionally requiring minimal financial support—into funding and support structures built on models of the sciences and social sciences.

The CFI grant covers the costs of initial programming and database development, but not for the research itself. This funding, in turn, is very difficult to come by. The main source of funding for humanities research in Canada is the Social Sciences and Humanities Research Council (SSHRC), and up until relatively recently, it has not specifically allocated a grant to digital humanities projects. Image, Text, Sound and Technology grants were initiated in 2000 but offer significantly lower amounts for a shorter duration than the traditional Standard Research Grants. Moreover, the program does not support digitization of collections, routine computer applications, creation of stand-alone major research tools, or development of technological infrastructure[32]—all necessary to the sustainability of most projects in digital humanities.[33]

Mostly, small research projects in this area have been reliant on Standard Research Grants, relatively few of which are awarded to textual or literary scholarship (see Appendix). These have tended to be projects traditional in scope: indices, concordances, and editions. Without knowing the relative numbers of applicants and their research histories, the reasons for this are hard to gauge. My own experience of applying each year for the past five years has been discouraging and, though the reasons for this range from my own newness and inexperience as a scholar to the way I have described the project, certainly some bias in the traditional humanities against digitization has resulted in very odd criticisms of my proposals. For example, one assessor last year asked, quite as if Google Maps, other digital mapping technologies listed in my bibliography, and basic keyword searching had never existed, "How can verbal comments be mapped on a map of London?" and commented, "It would be helpful to explain how the user would find information. For example, if one wanted to find Ivy Lane on the map? If one wanted to find the Gun in Ivy Lane? If one had simply a reference to the Gun, without mention of Ivy Lane?" The simplest explanation is that in some ways the interface will resemble Google Maps, where short annotations can be overlaid on the map, and can link to much longer texts in new windows. Much as in Google Maps, searching for "Gun" results in several hits; from these the viewer will select the most likely option, and click on provided links to more textual information about the place, or to the precise location or street on any given map. It is hard to know, indeed, how much basic education about computer interfaces must displace rationale and methodology in any grant application that is not dedicated to digital projects. We cannot rely solely on the tenuous possibilities of grants to build sustainability for either a group of researchers pooling resources or for any individual project, especially those that challenge traditional notions of "doneness" when a monograph or edition is published.

The final challenge I will address with regard to sustainability for digital scholarship at the institutional level is addressing our own structures for conferring tenure and promotion. The demands of this work require expertise in both traditional humanities research and digital methods. The members of my department mostly recognize this, but even though the first digital editions were published online in our department in the 1990s, it is only this year that the department is officially incorporating digital scholarship into its Standards for Promotion and Tenure document. In my experience of working in the digital humanities in the Canadian context it is still difficult to explain what it is that I do, why it is important, and why it is worth as much as the work that other researchers do. Recognition for this work at the departmental and college level is key: until digital scholarship has established itself as legitimate, and as equal to more traditional forms of book production, sustainability is uncertain, as digital researchers will have to turn their attention consistently to fighting local battles about worth and merit rather than focusing on research and development.

The Modern Language Association's 2006 report on tenure and promotion noted that new modes of scholarship include digital archives and humanities databases, though a major focus is on monographs and journal articles in electronic format (rather than new forms of publishing entirely in a variety of inventive digital project forms). Even these, relatively comparable to refereed publications in print, were not seen by a majority of departments as "important" for earning promotion and tenure. Furthermore, as the report concludes, "It is of course convenient when electronic scholarly editing and writing are clearly analogous to their print counterparts."[34] In our department we have started the process of defining what digital humanities projects entail, but we have not yet determined how "valuable" or how meritorious scholarly work is when it is ongoing, experimental, and without a formal peer review system. Given that many reviewers do not have

experience in the practices of digital projects and research, how will they tend to treat the significant but to date largely unacknowledged differences between creative and original digital project design, conception, and implementation, and the industriousness or "sweat-of-the-brow" projects that are more concerned with gathering and marking existing data (and all the various levels in between)? Should peer-reviewed external funding "count" as a meritorious indication of a successful research program, or is it merely an essential component of our practice? Some of my colleagues have invested many hours of investigation and experimentation into tools that are not yet publicly released. Others have incorporated student editions (teaching the editorial process) into their research. In terms of my own situation, I worry that in an online environment there is no need for an edition to represent a single work or stand-alone collection by a single editor or small group of editors, there is no need for an edition to be "finished": indeed, an edition need not be book-like at all. I imagine a multimedia work that encompasses a database of factual information never before integrated within a single system, an expanding corpora of networked texts, images, and unmoderated Web 2.0-style editorial commentary, a rigorous application of documented usability guidelines and testing, and high-quality graphic and interface design that is regularly updated; the resulting publication is an experiment, a teaching tool and a research tool, and quite possibly a decades-long commitment for its principal creator(s). All of these components might well make an innovative, important, sustained and usable contribution to scholarship, research, and teaching, but how will I, as principle investigator, present this work to the College Review Committee, which oversees tenure, promotion, and merit increases, in the first year? in the third year? in the tenth year? These issues are starting to be addressed at my institution, and they will certainly evolve as other universities work through the same processes of examining assumptions about peer review, the relative merits of digital versus print publications, and work that does not fall neatly into traditional categories of research and pedagogy.[35]

To conclude, every instance of setting up a research program entails a process of education both for me and for the people who ultimately help to support my work. Moving humanities scholarship, historically and typically a relatively inexpensive practice, into digital projects, which often require expensive equipment, space, and personnel, means one must educate one's colleagues, one's department or college, one's internal sources of funding, one's research office, and one's potential grant reviewers about the need for resources in an area that typically requires books, word processing, and thought, what those resources legitimately ought to be, and whether digital scholarship itself has any "value." I have discussed some of the obstacles that digital projects encounter, but I will end with what I think is likely to happen as institutions catch up to individual scholars working in this area. My goals, and many others, are accomplishable, and digital projects are likely to go forward in productive ways as long as the necessary collaborations between librarians, archivists, scholars, and commercial publishers are strategically pursued and achieved. While failed or stalled projects and collaborations, and probably a few lawsuits lie ahead, there is every indication that many other digital projects and collaborations will go forward, that new business models and new models of scholarship are emerging, and that they will benefit all stakeholders.