Thresholds of Engagement: Integrating Image-Based Digital Resources into Textual Scholarship

by

Rebecca L. Niles

A thesis submitted in conformity with the requirements for the degree of Master of Information
Faculty of Information
University of Toronto

© Copyright Rebecca L. Niles 2012
Thresholds of Engagement: Integrating Image-Based Digital Resources into Textual Scholarship

Rebecca L. Niles

Master of Information
Faculty of Information
University of Toronto

2012

Abstract

In recent years, technological advances in creating, storing, and accessing digital facsimiles of print and manuscript documents has resulted in an explosion of digitization initiatives. While such initiatives commonly endorse the viewpoint that digital facsimiles either replace or successfully stand in for their physical originals, textual scholars, whose principle interest is in the text as material artifact, do not share this perspective. Thresholds of Engagement explores the ways textual scholars engage with textual artifacts, tests the limits of representation of digital facsimiles and of the interfaces that house them, and proposes a model for the relationship between physical texts and their digital counterparts that privileges the requirements of textual scholars. The digital-facsimile interface proposed in this study is designed to facilitate methods described by textual scholars in interview—methods of comparison, material analysis, pattern recognition, and modelling—using an open-source web-based approach that is accessible for individuals to innovate and build upon.
Acknowledgments

I thank the staff and faculty of the iSchool at the University of Toronto for the consistently prompt, thorough, and intelligent assistance they have provided me throughout the various stages of this thesis. In particular, I would like to recognize the contributions of Professor Sara Grimes, who provided me with both the foundational research methods used in this thesis and helpful feedback during the earliest stages, and of Professor Kelly Lyons, who provided me with training in methods of project management and provided feedback relating to prototype planning. I also thank Julia Flanders of Brown University, for performing as the external examiner for my thesis defense, and for her thoughtful—and thought-provoking—feedback.

Thanks are also due to Voytek Bialkowski, whose own thesis has operated for me as a model of a successful MI thesis, and whose insights into the thesis-writing process have offered both reassurance and commiseration.

I thank my friends and family, who offered love and support throughout the duration of this work, and patiently tolerated my absence from their lives when the work consumed me.

I thank Randall McLeod not simply for his guidance—and patience—throughout the process of writing this thesis, but also for years of instruction in bibliography and textual theory. It is to him I owe my introduction to the archeology of the book, and it is his unique perspective on textuality, his brilliant visual modelling techniques, and most of all his infectious enthusiasm for uncovering the hidden narratives of the book, that are the impetuses for this thesis.

I have left off thanking my thesis advisor, Alan Galey, until the end only because of the difficulty involved in finding words sufficient to express the extent of my debt to him. This thesis was born of conversations between us going back three years, conversations that also led to me applying for the Master of Information program at the iSchool. As my employer, Alan has made me feel like a valued member of his team and opened up opportunities that have introduced me to Digital Humanities and the minds that fuel the field. As my thesis advisor, Alan has offered his time and insights freely and without reservation. As my advocate, Alan has demonstrated faith in my abilities and has been instrumental in helping me launch my career. For all this, I think him most deeply and sincerely.
# Table of Contents

Acknowledgments ........................................................................................................ iii

Table of Contents ........................................................................................................ iii

List of Figures ............................................................................................................... vi

List of Appendices ....................................................................................................... vii

1 Introduction .............................................................................................................. 1

2 Literature Review .................................................................................................. 13

  2.1 Theoretical Foundations of Current Textual Scholarship .................................. 14
      2.1.1 The Advent of Modern Bibliography ......................................................... 14
      2.1.2 Movements Towards a Social/Materialist Approach to Textual Scholarship .... 19

  2.2 Theoretical Foundations of Image-Based Humanities Computing .................. 24

  2.3 Implementations of Image-Based Humanities Computing for Textual Scholarship ...... 28

      2.3.1 Projects ........................................................................................................ 28
      2.3.2 Tools for Managing Image-Based Data ....................................................... 33
      2.3.3 Best Practices and Techniques in Digital-Facsimile Creation ......................... 35

  2.4 Summary ............................................................................................................. 37

3 Methodology ........................................................................................................... 40

  3.1 Research Design ................................................................................................. 40
      3.1.1 What is a Textual Scholar? ........................................................................... 41

  3.2 Procedures and Data-Gathering ......................................................................... 42

      3.2.1 Systematic Journal Review ........................................................................ 42
      3.2.2 Target Informants ......................................................................................... 44
      3.2.3 Interviews ..................................................................................................... 45
      3.2.4 Data Recording, Reduction, and Analysis .................................................... 47

  3.3 Prototyping ......................................................................................................... 48
List of Figures

Figure 1: Beginnings of the physical archive of the Internet Archive, in Richmond, CA. ........... 4

Figure 2: Illustration of Proposed Digital-Facsimile Environment ................................................. 72

Figure 3: McLeod's diagram of remote offsetting in *Gulliver's Travels* (Huntington, 16126) .. 78

Figure 4: Greetham's diagram of the possible transmission of Trevisa........................................ 80

Figure 5: *Thresholds of Engagement*: Navigational Viewer Prototype page ................................ 83

Figure 6: Navigational Viewer—Default Page View ........................................................................ 84

Figure 7: Navigational Viewer—Page-Order View of *The Dunciad (1729)* .................................... 85

Figure 8: Navigational Viewer—Imposition View of *The Dunciad (1729)* ................................. 86

Figure 9: *Alice's Adventures in Wonderland* by Lewis Carroll, illustration by John Tenniel ...... 94

Figure 10: *Alice's Adventures in Wonderland* by Lewis Carroll, illustration by John Tenniel ... 96
List of Appendices

Appendix A—Ethics Approval........................................................................................................108
Appendix B—Call for Participation (Group 1) ....................................................................................109
Appendix C—Call for Participation (Group 2) ....................................................................................111
Appendix D—Letter of Informed Consent............................................................................................113
Appendix E—Interview Guide..............................................................................................................115
1 Introduction

In the title of this thesis, *Thresholds of Engagement*, the term *threshold* is used to describe the point of contact between spaces, as in the threshold of a building, which can either be barred to prevent entry, or be thrown open to admit free access. The spaces that I consider in this thesis are both the physical space of the rare book library, archive, or other homes to textual artifacts, and the virtual space in which digital facsimiles reside. The types of engagement I am interested in here are the ways in which textual scholars engage with physical artifacts on the one hand, and with digital facsimiles on the other. The phrase “thresholds of engagement” refers not only to the limits that must be surpassed when using digital images in textual research, but also to the digital interface as threshold, the design of which dictates the extent of a user’s engagement. In order to facilitate the use of digital facsimiles as tools for textual scholarship, the interfaces we use to view and manipulate them must be designed with the methods of the textual scholar in mind. In order to better understand these methods, this thesis includes extensive research into the investigative practices of textual scholars, discussion of design and technical requirements of digital resources for this community, and a prototyping phase in which one of the potential functions discussed is implemented in a web-based environment.

As the emergence of sophisticated, cost-effective methods of creating digital facsimiles of books and other documents causes librarians, archivists, scholars, and indeed the general population to reflect on the importance of preserving and making accessible the original physical texts, two models of the relationship between the original and the digital facsimile have emerged. On one hand is the view that with digital facsimiles available, no one should need the physical text. Librarians can now get busy clearing space in the stacks for more computer stations, and administrators can start shifting collections development money into the tech fund. On the other hand is the idea that digitization allows us to create stand-ins for print materials, thereby

---

1 Any similarities between this scenario and the disappearance of the A.S.P. Woodhouse Collection (ironically, containing many valuable texts on book history) from Robarts Library at the University of Toronto to make laptop-ready study space are purely coincidental.
allowing us to have a substitute that is *almost* like the real thing, while the real thing in question stays safely tucked away.

These models can be appealing to libraries. The benefit of added space—in the building and in the budget—that results from trading pixels for pages may be hard to resist, and, furthermore, limiting access to physical materials by redirecting readers to a digital doppelganger does indeed help preserve rare books. However, for one particular kind of reader—for whom the material text as artifact is of principle interest—neither scenario fits.

The primary focus of this thesis is to develop a third, alternative model for the relationship between physical texts and their digital copies, one that resists treating image-based digitizations of texts as either replacements of or surrogates for their physical originals. While existing approaches are suited to meet certain institutional interests, this third model is designed rather to privilege the interests of bibliographers, textual critics, and book historians (hereafter collectively referred to as *textual scholars*) through reimagining digital facsimiles as tools to aid direct examination of physical textual artifacts.

A recent development in the world of archiving serves to illustrate the negative consequences of treating digital facsimiles as replacements or stand-ins for the physical text, and may make it clear why I am seeking an alternative. Brewster Kahle, founder of the digital repository Internet Archive, has been receiving press recently for his latest archiving aspiration. His plan, ambitious nearly to the point of absurdity, may strike many as the library equivalent of billionaire Richard Branson’s Virgin Galactic luxury space-charter. Kahle is attempting to build a physical archive containing one copy of every book in existence. The massive undertaking was spurred by a very realistic concern about a growing trend that Kahle and his team witnessed, that when books were digitized, the libraries that previously held them saw less value in retaining the physical originals. Members of the Internet Archive noticed that upon digitization, some of the Archive’s partner libraries shifted the physical books returned to them into off-site repositories, no longer to be used. One library went so far as to discard books that had been digitized for them by
Google. The problem, according to Kahle, was that some libraries have come to treat the digital facsimile as a replacement for the book—an upgrade resulting in obsolete technology to be gotten rid of.

The disturbing trend of libraries doing away with books motivated Kahle to create a physical repository for all those books that were not being returned safely to the shelves of a library, and to take in donations of books from institutions and private citizens alike for safe-keeping. The archive uses a scalable system of 40-foot-long stackable shipping containers with climate-control features built in. Each container holds 40,000 books, packed tightly in cardboard boxes. No matter that this scheme makes retrieving any one book for use nearly impossible. In this design, the digital copies of the archive’s books would operate as the primary text for use, with the physical copy available as “the authentic and original version that can be used as a reference in the future,” such as when the representativeness of the digital copy is in question, when new technology necessitates the creation of a better surrogate, or to protect against the physical book’s extinction in the event that all other copies are destroyed in a catastrophe. The very design of the archive prescribes a particular relationship between the physical book and the digital facsimiles created from it. The extreme difficulty associated with accessing the physical book, compared with effortlessly accessing the digital copy over the internet, makes it clear that the facsimile is meant to behave like a surrogate. Ostensibly, the digital proxy increases access to texts, but its existence also provides a rationale for librarianship and archiving practices that impose limitations and barriers to readers interested in looking at a physical copy of a text.


3 Ibid.
Figure 1: Beginnings of the physical archive of the Internet Archive, in Richmond, CA. Photo retrieved from http://blog.archive.org/2011/06/06/why-preserve-books-the-new-physical-archive-of-the-internet-archive/.

The story of Kahle’s archive demonstrates the negative implications of those two prevalent schools of thought in the field of digitizing texts, referred to earlier. One the one hand is the perspective that digital facsimiles can replace the physical books from which they are derived. Such is the rationale of the library (which remained unnamed in Kahle’s blog post) that began to discard the books returning from Google’s digitization facility. It is this line of thinking that can make digitization unpopular among librarians, archivists, and others, who worry that digitization will inevitably result in the extinction of the books that have gone before. This concern over extinction also plays out in media coverage, such as that of the New York Times and the Daily Mail, which casts Kahle as a “latter-day Noah.”

On the other hand, where the second perspective on digitization of textual artifacts emerges in Kahle’s tale is where the biblical analogy begins to break down. Noah kept two of every animal, with the intention that once the water receded the creatures would be let back out into the wilderness. Kahle’s approach is closer to that of a museum of natural history, where a combination of careful preservation and protective glass may keep animal specimens in good shape, but precludes any opportunity to find out what they sound like, feel like, or how they might behave in the wild. While creating digital facsimiles as surrogates produces an alternative to direct access to physical texts, and thereby improves preservation, it can also foster the false impression that viewing an image-based surrogate is equivalent to experiencing the physical artifact itself. Furthermore, the relative ease of access of the digital surrogate can act as an institutional rationale for not enabling, and sometimes actively discouraging, access to the physical book itself. Finally, Kahle’s physical archive, along with the Internet Archive, and indeed most digitization initiatives, treats individual copies of books as representative of the work as a whole, rather than as potentially bibliographically distinct from one another. Without retaining multiple specimens, those scholars interested in the diversity of texts as manifested in physical artifacts would be as disadvantaged as Darwin if he had only studied a single finch.

Certainly from the perspective of the textual scholar, and to a great extent of a more general audience as well, digital facsimiles fall short of adequately replacing the original textual artifact. Furthermore, while treating digital facsimiles as surrogates for physical texts provides the dual benefits, usually found only in opposition, of increased preservation and access to content, the surrogate cannot embody the material qualities of the original. This problem is most acute in cases where the presence of a surrogate is used by special collections facilities as a rationale against ever accessing, or allowing access to, the physical artifact itself, since in this scenario all evidence not captured in the image becomes inaccessible to the researcher.

The purpose of this thesis is to propose a third scenario, one that rejects both the concept that digital facsimiles perform all the same functions of the original and therefore can replace them, and the idea that digital facsimiles capture enough of the essence of the originals that they can act as satisfactory representations. This may be true for some readers, but for textual scholars, who study the material features of textual artifacts, the barriers that these models erect are critical. For this reason, the present study envisions an alternative scenario, one in which physical texts and their digital facsimiles are treated as complementary, each fulfilling roles and
enabling behaviours that the other cannot. In this thesis I will profile the history and current state of textual scholarship methods, weigh the strengths and weaknesses of representation of textual artifacts through digital facsimiles, and demonstrate how digital environments for accessing, organizing and manipulating digital facsimiles can be designed with the needs of textual scholars in mind, to allow them to leverage the benefits of a virtual environment for handling texts, while simultaneously working directly with physical materials.

In order to get started, there are three questions that should be addressed at the outset. First, what is the difference between the text in the abstract, and the textual artifact, and why is the latter important? Second, how might we define the intellectual practice of textual scholarship in a way that is inclusive and yet specific enough to be useful? And third, what is meant by digital facsimile, and how is its role within rare book and archive settings evolving?

An important distinction must be made when discussing texts—manuscripts, books, and other written materials—as physical objects: the text-as-artifact, and the text-as-work. While the text-as-work maintains an existence outside of each individual copy, the existence of the text-as-artifact relies on the particular material instantiation of a document. The text-as-artifact, as Peter Shillingsburg notes, is a material object, which can occupy only one space at one time, while the text-as-work may be iterated an indefinite number of times through space and time, and still be considered the same text.\(^5\)

Another way to consider the distinction would be to apply Nelson Goodman’s concept of autographic and allographic arts.\(^6\) An allographic art is one in which, he writes, “a definite notation…provides the means for distinguishing the properties constitutive of the work from all contingent properties—that is, for fixing the required features and the limits of permissible

---


variation in each.”7 Applied to textuality, the text-as-work is allographic in so far as, so long as the “definite notation”—the sequence of characters forming the written language of the text—do not deviate appreciably from the ideal, the text-as-work is considered genuine.8 On the other hand, an autographic art is one wherein each property of the object is constitutive of its identity as such, and (again, in the words of Goodman) “no such feature can be dismissed as contingent, and no deviation as insignificant.”9 For the text-as-artifact, not only the words it contains, but also each and every property of its material existence is constitutive, and therefore its existence is necessarily unique. It is the text-as-artifact, the unique autographic text, with which this thesis is concerned.

Nearly every textual scholar is concerned with some aspect of the autographic instantiation of texts. This thesis follows David Greetham’s definition of textual scholarship. He groups a wide variety of disciplines—“enumerative bibliography, systematic bibliography, descriptive bibliography, analytical bibliography, historical bibliography and textual bibliography, textual analysis and textual criticism, textual editing, documentary editing, and social textual criticism”—under the banner of textual scholarship based on “the disciplinary relatedness of all aspects of the study of the text of a book.”10 To Greetham’s list I would add book historians, and in relation to Greetham’s definition I would reinforce that all of these disciplines relate to each other in their study of the text in its various material instantiations, or, the autographic text, rather than just the book. This is evident by removing Greetham’s reference to the artifact in

7 Ibid.

8 This concept may also be profitably applied to understand the objectives of certain schools of textual criticism, most notably the New Bibliography, to recover as much as possible the ideal sequence of the text when creating new material instantiations, i.e., new print editions. The potential for electronic editions to simultaneously represent multiple variants challenge the allomorphic nature of the text-as-work (see, for example, Alan Galey’s recent work representing multiple readings in his “Animated Variants” prototype, available at http://individual.utoronto.ca/alangaley/visualizingvariation/animated.html).


which the text is situated; “all aspects of the study of the text” broadens the definition to include all manner of literary scholars as well, while in its original form, “all aspects of the study of the text of a book,” the site of the text becomes part of the equation that delineates the field of textual scholarship.

G. Thomas Tanselle provides, in his 1989 article “Reproductions and Scholarship,” a definition of the term “reproduction” that will help to understand the nature of what is at issue when discussing facsimiles. He writes, “When I use the word ‘reproduction’ (and, more loosely, ‘photocopy,’ ‘photofacsimile,’ and so on), I mean the product of any chemical or electrostatic process that aims to represent with exactness … not only the text of a given document but also the details of its presentation, insofar as they can be duplicated on a different surface.”11 It is simple enough to extend this definition to include digital facsimiles, the product of a process of electronically sampling, processing, and displaying colour so as to produce an image of a document. Tanselle notes that the ability to reproduce parts of textual artifacts through photocopying has been transformative in the field of textual scholarship,12 and I would add that there is a modern-day correlation to this in the increasingly permissive rules in rare book libraries and artifacts allowing scholars to take their own digital images, on site. Creating digital facsimiles of large numbers of entire texts and making them available online is becoming a common practice among institutions, as Melissa Terras points out: “Digitization projects in libraries, archives, museums and other ‘memory institutions’ are merrily creating image representations of their holdings for display, sharing and distribution via the Internet.”13 When discussing “digital facsimiles” in this thesis, I will be referring to digital reproductions along the lines of Tanselle’s definition as adapted above, and when discussing digital-facsimile environments, I will generally be discussing the electronic, and usually online, interfaces used by providers and institutions to deliver the products of digitization programmes such as those described by Terras.


12 Tanselle, “Reproductions and Scholarship,” 25.

The research presented in this study has its precedents both in textual scholarship and in digital humanities. Amidst dire predictions of the death of the book, fields such as book history and modern textual scholarship have flourished, and the physical text as a primary subject of investigation and research has become a central element of this discipline. As the MLA’s “Statement on the Significance of Primary Records” demonstrates, access to textual artifacts continues to be indispensable to researchers, regardless of available alternatives such as microfilm and digital proxies.¹⁴

Image-based humanities computing, as Matthew Kirschenbaum notes, is an area of research and practice that descended from the social/materialist school of textual scholarship championed by Jerome McGann, D.F. McKenzie, and others, and takes advantage of the capacity of computing to act as a venue for representation to illustrate the physical properties of books.¹⁵ Currently, image-based humanities computing is in a unique position of opportunity, as digital technology increasingly enables high-quality, low-cost digitizations of textual artifacts as well as the ability to deliver these digitized artifacts to users.¹⁶

This thesis responds to the unique opportunities afforded by technological advances in image-based humanities computing for textual scholarship, and the need for advances in this area to be made with the practices of textual scholars in mind. While many studies of digital tools in the humanities have focused exclusively on their analytical and representational potential—that is, their ability to make research conducted within the rare book library accessible to those beyond—this thesis considers the use of digital tools within the rare book library. In other

---


¹⁶ Ibid.
words, this thesis proposes to think of digital facsimiles as tools to become integrated into the
textual scholar’s primary engagement with the textual artifact itself.

In her chapter for Blackwell’s *Companion to Digital Humanities*, Claire Warwick points out that
one of the major areas of difficulty for the relationship between humanities computing and
traditional humanities is a lack of understanding of how humanities scholars actually perform
their research, which can result in computing solutions that do not appropriately address the
problems they were conceived of to address. She concurs with D.A. Norman, in that “computers
are of most use when they complement what humans can do best,” and infers from this that “[a]
clearer understanding of what humanities scholars do in their scholarship is therefore
important.”  

In order to develop a model of the relationship between physical texts and digital
facsimiles that reflects the interests and needs of textual scholars, this thesis seeks to establish a
deep understanding of the practice of textual scholarship through direct engagement with textual
scholars and their work, and applying this understanding to the design of a digital environment
to facilitate textual investigation in a way that facilitates, rather than co-opts, the critical
activities of investigation, giving textual scholars, in the words of Warwick, “tools to think
with.”

In his 2000 lecture entitled “Scholarly Primitives: What Methods do Humanities Researchers
Have in Common, and How Might our Tools Reflect This?” John Unsworth proposed a shared
“list of functions (recursive functions) that could be the basis for a manageable but also useful
tool-building enterprise in humanities computing.”  

Unsworth describes activities such as

17 Claire Warwick, “Print Scholarship and Digital Resources in *A Companion to Digital Humanities*, Susan

18 Ibid.

discovering, comparing, and sampling, which he suspects are common activities of humanities scholars, and explains how a knowledge of these activities can contribute to a viable framework for designing widely applicable digital tools for humanities research. My vision for this thesis is to discover, through interviews with members of the textual scholarship community, what common methods or behaviours textual scholars engage in when investigating textual artifacts, and through small-scale web-based application prototyping, to theorize to what extent image-based digital resources can aid in these activities.

This study is presented in six parts. Following this introduction (chapter 1), chapter 2 contains a review of the relevant literature in the areas of both textual scholarship and image-based humanities computing. I will discuss the theoretical foundations of modern textual scholarship, beginning as early as 1870 and progressing through the establishment and eventual fracture of the New Bibliography and the introduction of the social/materialist school of textual scholarship, as well as the relatively recent theoretical foundations of image-based humanities computing. The review will then turn to a discussion of the implementation of image-based humanities computing techniques relating to textual scholarship, including some examples of past projects, as well as guidelines, best practices, techniques, and tools currently available to facilitate successful projects.

In chapter 3, I outline the methodology of this thesis. In one respect, this work fits the research model commonly encountered in the digital humanities. However, this thesis also borrows research techniques typically associated with the social sciences, in that data gathering through both systematic sampling of recent publications of the target subject group, namely, textual scholars, and in-depth interviews with select subjects, is based on techniques borrowed from the social sciences. Methods of professional project management also form the principal methodology of the prototyping exercise that is the subject of chapter 5.

\[\textit{Practice}, \text{King’s College London, May 13, 2000, http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html.}\]
In chapter 4, I discuss the findings of the interview sessions described in the previous chapter. Taking my cue from Warwick, I adopt the rationale that the basis for effective design of digital tools and interfaces is a thorough understanding of the target user-group. I plumb the interview participants’ responses for particular insights into their perspectives on the relationship between the physical artifact and its digital surrogate, common activities they perform that could be aided by digital tools designed just for them, and the technical considerations that should be accounted for. Following this discussion, I put my new understanding of the work of textual scholars into action, in the design of a digital-facsimile environment that complements the activities most important to textual scholarship. The ultimate goal of this experiment in design is to lay the foundations for a tool for the textual scholar to think with.

Of course, a theoretical model that purports to offer a useful way of integrating digital facsimiles into on-site textual investigation lacks validity without proving the model’s viability through practical application. In chapter 5, I report and discuss the outcome of a brief prototyping experiment, which serves three main functions. First, it acts as a partial test of the fitness of the digital-facsimile environment proposed in the previous chapter. Second, it reveals the potential to model various ways of thinking of the text as a material artifact through flexible display techniques. And third, it demonstrates the fitness of a lightweight, web-based approach for creating digital tools that are both effective and open to alteration and innovation by their users.

Finally, in chapter 6 I will return to the themes that give this thesis the title *Thresholds of Engagement*, by discussing the ways in which digital facsimile interfaces, operating as portals to the materials they house, can either limit or encourage engagement, and by looking forward to the ways in which doors can be opened to invite textual scholars to shape the future of how textual artifacts are represented and explored through digital environments.
2 Literature Review

During the introduction, I identified three simple questions that have a big role to play in this thesis: What is the text-as-artifact? What is textual scholarship? What are digital facsimiles? In this literature review, I return to these questions in greater detail. Here, I discuss the development of modern textual scholarship and the various stances on the text-as-artifact that have emerged, I provide an overview of the foundations of image-based humanities computing, and I review some of the current work being undertaken in the area of representing textual artifacts through digital facsimiles.

In the first part of this chapter, I take a look back at the theoretical and methodological foundations that shaped modern textual scholarship from shortly before the turn of the twentieth century to the current day. In order to understand more about the process of the material investigation of the text, I assess in particular how attitudes towards the physical document as a source of evidence evolved, and how methods of investigation took shape. By the end of this section I will have traced the way back to the current day, having shown how the shift in understanding texts as material objects created through a network of transmission placed a new emphasis on representing textual materiality and alterity in editions, such as through the use of digital facsimiles.

An examination of the advent and current state of image-based humanities computing follows the trajectory described in the previous section. Advances in computing resulting in the ability to produce and distribute high-quality digital facsimiles at lower cost have made the representation of texts in their various material forms through images far more viable. However, some concerns have been raised amidst the optimistic adoption of this approach, which serve as a reminder that digital representation is no replacement for the physical artifact it represents.

Next, I consider modern textual scholarship and image-based humanities computing together in practice. How has textual scholarship utilized the tools afforded it by image-based humanities computing? In what ways have textual scholars’ uses of the digital medium to represent textual artifacts been successful, and where is there still room for improvement? What useful tools have emerged at the intersection of textual scholarship and computing? And, if there are
improvements to be made, what are the standards and practices in the field of digital imaging and display upon which to build?

2.1 Theoretical Foundations of Current Textual Scholarship

As was discussed in the introduction to this thesis, the field of textual scholarship incorporates a wide variety of sub-disciplines, all of which are related through their central focus—the study of the text-as-artifact. In this section of the literature review, I discuss the history of modern textual scholarship, with an emphasis on two main areas: first, the methodological underpinnings of modern textual scholarship, and second, the shift in critical opinion as to how early texts should be represented. The era of modern bibliography is marked by the establishment of a methodology for systematically documenting and analyzing material evidence in physical texts, and the development of a system for analyzing such evidence. As this newly-developed methodology was established, its main objective became, perhaps ironically, to create editions of texts that are stripped of the features they acquired during the process of transmission, leaving behind an idealized conception of the author’s intended text. Only later, when social/materialist textual scholarship and unediting emerged, was there a resistance to this dismissal of the influences of material transmission as corruption, and with this resistance came a newfound enthusiasm for making early texts accessible to readers through facsimile editions. As I shall argue throughout this thesis, the next stage in this trajectory is to move from mere static representation of textual artifacts to the provision of digital environments that enable dynamic exploration of textual materials.

2.1.1 The Advent of Modern Bibliography

During the late nineteenth and early twentieth centuries, a shift occurred among those who concerned themselves with the study of the book, toward an emphasis on the importance of evidence derived from the physical text (as opposed to support derived from either historical assumptions or critical/aesthetic judgments), and the use of this physical evidence to support theories of production processes. The major change that marks the beginning of what is commonly thought of as modern bibliography is the notion that the physical features and phenomena in books, independent of the verbal content they contain, are of value as evidence for historical investigation, and provide information relevant to the reading of their literary
Those who adopted this new methodology were known as the New Bibliographers. G. Thomas Tanselle credits Henry Bradshaw as being one of the first textual scholars to stress the value of systematically collecting and analyzing physical evidence as a means of gaining insight into early printing practices, but it was W.W. Greg who was perhaps most influential in establishing the methods and techniques of the New Bibliography as standards in textual scholarship. Greg defined critical bibliography (as opposed to other forms such as enumerative and descriptive bibliography) as “the science of the material transmission of literary texts, the investigation of the textual tradition as it is called, in so far as that investigation is possible without extraneous aids. It aims at the construction of a calculus for the determinations of textual problems.” This sentiment captures nicely the emphasis on text creation as a process of transmission and the primacy of the physical text as source of bibliographic evidence. It also anticipates the relevance of using physical evidence to assess variant textual readings for critical editing.

The concept that the production history of a given text could hold evidentiary value for the editor of critical editions was relatively unconventional, until various methods of material investigation, such as headline analysis (tracking the reoccurrence of particular settings of type used to create the running titles at the tops of pages) introduced by A.W. Pollard in his 1909 *Shakespeare Folios and Quartos* demonstrated the ability to use evidence based on printing practices to come to conclusions about textual states. Emphasis on the importance of an understanding of the history of textual transmission for critical editing increased throughout the twentieth century. To this end, Phillip Gaskell’s 1972 reference work *A New Introduction to Bibliography* differs significantly in approach from McKerrow’s pioneering guide, *An


21 Ibid., 6.


23 Tanselle, *Bibliographical Analysis*, 16-17.
*Introduction to Bibliography for Literary Students* (1927), in that it focuses on the historical processes that were employed in producing books, devoting one of the three sections of the guide solely to describing how “an understanding of the history of book production may be used in approaching problems of identification, description, and the establishment of a text.”


Much of the drive towards establishing a systematic methodology for gaining knowledge about the processes of transmission was initiated by the desire to produce authoritative versions of texts, free from choices based on aesthetic judgment and therefore more resistant to dispute. Members of the New Bibliography emphasized the importance of using the new systematic methods they had developed to determine authoritative versions of texts, beginning with Shakespeare. The definition of an authoritative text, as will be discussed momentarily, was an edition that is thought to most closely resemble the author’s intended text, and that is cleared of variants, or “corruption,” which may have been introduced during the text’s transmission. In this way, the methodological approaches developed in the early days of the modern renaissance of textual scholarship effected the shift from examination of textual artifacts to shaping the representation of texts as idealized verbal content, independent of its material containers.

From the beginning of the revolution in textual scholarship at the turn of the century, champions of the new kind of bibliography made efforts to characterize it as belonging to the class of the sciences. In particular, early advocates such as Henry Bradshaw and W.A. Cosigner adopted the model of natural history, in which copies and editions have a shared genetic heritage, and in which the various branches of textual scholarship can be systematically classified. Indeed, the thematic similarities between disciplines that study natural history and disciplines that study


25 Greetham, *Textual Scholarship*, 2; emphasis in original.

textual history are striking. For instance, when Darwin wrote about determining geological history from available evidence, he compares the experience to reading a badly damaged text. He writes,

I look at the natural geological record, as a history of the world improperly kept, and written in a changing dialect; ... Each word of the slowly-changing language, in which the history is supposed to have been written, being more or less different in the interrupted succession of chapters, may represent the apparently changed forms of life, entombed in our consecutive, but widely separated formations.

The principal concern in this statement is with reading a history of variation out of the book of nature. Darwin used scientific methods to study difference, understanding such variation as the product of the transmission of genetic information over time. The ability to do the same with texts is promised by the methodology developed by early scholars in the modern era of textual scholarship, and is indeed realized later in the social/materialist approach to textual scholarship, discussed presently.

However, when the New Bibliography turned the systematic rigour of bibliography toward the task of isolating and removing textual “corruption,” and creating purified, authoritative editions, their campaign more closely resembled, not natural history the way Darwin saw it, but instead the literal—and at times brutal—application of Darwin’s concepts in the field of eugenics. As Joseph Grigely writes in his monograph *Textualterity*, “To the extent that eugenicists treat the body as a text (to be ‘edited’ eclectically), and editors treat the text as a body (describing it in pathological terms), there is in the two disciplines an interface of critical discourse that concerns

27 I am grateful to Professor Alan Galey for first alerting me to the similarities between concepts of genetics in natural history and variation in textual scholarship, during his Books 1000 course at the University of Toronto, Winter, 2010.

how we look at, categorize and describe human and textual differences.”

Indeed, both initiatives espoused an ideology equating difference with corruption, and, perhaps even more disturbingly, both were fueled by an underlying sense of moral justification, and even obligation, to purify their subject of such corruption.

Fredson Bowers, a major figure in the New Bibliography, was perhaps one of the most vocal, and most aggressive, supporters of this viewpoint. For Bowers, adherence to proper bibliographic method was the only way for the editor to avoid perpetuating textual corruption, a failing that Bowers frequently equates with moral degradation: “Every practicing critic, for the humility of his soul, ought to study the transmission of some appropriate text,” Bowers preaches in the Sandars lectures of 1957–58, and, further, “One can no more admit ‘just a little corruption’ to pass unheeded in the transmission of our literary heritage than ‘just a little sin’ was possible in Eden.”

For Bowers, bibliographic methodology is the fiery sword in the hand of the redeemer, New Bibliography, and what is at stake is not simply the soul of the critic, but rather the soul of humanity in the form of its literary heritage. This editorial regime aggressively eschewed textual variation and sought to quarantine the text from the mediating influences of its material conditions, believing this to be possible through rigorous application of bibliographic method. As we will see presently, the remaining history of textual scholarship leading up to the present day is characterized foremost by a rejection of the notion of the purified, idealized text, and a reevaluation of the study and representation of the material instantiations of texts.

Ultimately, it was the New Bibliography’s confidence in the infallibility of conclusions derived from a systematic analysis of physical evidence that became the focus of criticism during its decline. Growing skepticism regarding the limits of tenuous conclusions purportedly based on solid analysis of textual evidence, but in reality also relying heavily on assumption and conjecture, could be heard in publications like Studies in Bibliography, and culminated in D.F.


McKenzie’s 1969 “Printers of the Mind: Some Notes on Bibliographical Theories and Print-House Practices.” An approach to bibliography was on the rise that placed greater emphasis on investigating the historical contexts of text-creation, and, with it, an eventual critique and reimagining of the methods and limits of remediation and representation of textual artifacts.

2.1.2 Movements Towards a Social/Materialist Approach to Textual Scholarship

In the 1980s the social/materialist approach to textual scholarship appeared. This approach, characterized in the writings of D.F. McKenzie and Jerome McGann, privileged the physical text as its primary site of investigation, and prioritized the understanding of processes, rather than just products, of textual transmission. However, the social/materialist approach diverged from New Bibliography significantly in two fundamental and related ways. While the New Bibliography considered the pursuit of a text that represents the final authorial intention for a work to be the primary purpose for bibliographic investigation (i.e., recovery of the ideal text), the social/materialist approach questioned authorial intention as an organizing principle, and instead advocated for a focus on the text as an entity created as much through the processes of transmission as by its author or authors. As a result of this shift in focus, social/materialist bibliographers placed a greater and more inclusive emphasis on the processes of textual transmission.

In the era of New Bibliography, the concept of authorial intent acted as a guiding rationale for textual editing. What made a text corrupt, in the eyes of New Bibliography, was the infiltration of non-authorial influences. Scholars such as Bowers and Greg devoted much of their critical work to establishing methods for distinguishing between authorial inscriptions and the influences of other agents in the transmission process. According to Bowers, “if we respect our authors we should have a passionate concern to see that their words are recovered and


currently transmitted in as close a form to their intentions as we can contrive.”

When concerns regarding the fallibility of New Bibliography’s methods emerged, so too did skepticism regarding the true authority of the texts they created using these methods, as well as the concept of authorial intent as a governing principle for textual editing.

Rejection of authorial intention began to take root, which formed the basis of the social/materialist approach to bibliography. One of its primary proponents has been Jerome McGann, who in 1983 asserted that editorial principles grounded in authorial intent “so emphasize the autonomy of the isolated author as to distort our theoretical grasp of the ‘mode of existence of a literary work of art’ (a mode of existence which is fundamentally social rather than personal).” McGann argues for a socialized concept of authorship and textual authority, claiming that since texts are not constructed in a vacuum, “the location of authority necessarily becomes dispersed beyond the author.” This is especially true for McGann from a bibliographic sense, in that he recognizes that the material instantiation of the text, which is influenced heavily by the various agents of the transmission process, contributes to the literary effect of a work.

McKenzie similarly stressed that the physical forms of texts affect meaning, and advocated that bibliography be a study of not only the technical, but also the social processes of transmission. Like McGann, McKenzie situated textual creation as an ongoing polyvocal process, rather than an ideal, univocal product of a monolithic author-figure. This emergent insistence on the


35 Ibid., 8.

36 Ibid., 84.


constitutive role of material and social factors in the creation of the text raised new considerations for the representation of these texts, as we shall see unfold particularly in relation to the representation of textual alterity and material situatedness, discussed below.

Along with the shift in focus from the authorial to the social as a locus of textual creation is a widening of the scope of the process of transmission. In the social/materialist approach to textual scholarship, the scope that the word “transmission” referred to, particularly in McKenzie’s writing, was extended to include a much wider range of activity, including not only production, but also reception. In McKenzie’s view, bibliography should be thought of as “the study of the sociology of texts,” and in so doing, he extends the already inherently diachronic study of the text as a record of its own transmission into a nearly limitless engagement. While Greg could dismiss the study of bookplates in 1912, claiming that “bibliography only concerns itself with processes that leave their mark on the character of the finished book,” McKenzie sought for a scope that recognized transmission, distribution, reception, and the resulting creative regeneration by audiences. This change in scope, of course, introduced the issue of how to represent a larger range of states of a text, as well as the importance of including paratextual materials in the representation. The use of reproduction technologies such as photographic facsimile to represent texts as artifacts is a direct response to this issue, and the application of digital-facsimile technology to this problem is a continuation of that agenda.

There is a strong correlation between the shift in emphasis from New Bibliography to methods and theories characterized as social/materialist, on the one hand, and the rising popularity of book history on the other. Approaches adopted by the successors to New Bibliography mirrored those of book history in its wider interest not only in textual artifacts, but also in the social conditions within which these artifacts emerged, as well as the social impact of textual events. In a way similar to how bibliographers of the early twentieth century recognized the need to establish a base upon which to build a systematic bibliography, works such as Lucien Febvre and Henri Martin’s *The Coming of the Book: the Impact of Printing 1450-1800* and Elizabeth


Eisenstein’s *The Printing Press as an Agent of Change* laid much of the groundwork for considering the roles of texts and text-technology in human history. However, the correlation between the social/materialist school of bibliography and book history is most keenly felt in the model of textual transmission first outlined by Robert Darnton and later adapted by Thomas Adams and Nicolas Barker. In these models, each of the various stages of transmission of a text (characterized as “actors” in Darnton’s model and as “events” in Adams and Barker’s) is influenced by social, political, and intellectual factors. These models serve as illustrations of very similar views of textual transmission to those of McGann and McKenzie, namely that texts are not created solely by their authors, but are shaped also by a network of both physical and social influences. From the relatively new but increasingly influential field of book history came an interest in the material text as situated within a network of transmission, and therefore also an increased pressure to represent these texts in dynamic, inclusive ways that the classically-edited texts of the New Bibliography were simply ill-equipped to support.

The mounting pressures coming both from social/materialist bibliography and book history to recognize the text as an product of a systematic process of transmission with its own bibliographic as well as linguistic codes comes together in the movement described by Gabriel Egan as the New Textualism. This movement rejected the hierarchical model of authoritative vs. corrupt texts espoused by the New Bibliography. Choosing instead to see the variant forms of texts as, in the words of Leah Marcus, “different instead of debased,” they embraced this


This position led to a rejection of eclectic editing, and a desire to fight back against editorial whitewashing through a self-conscious return to the material context of the early editions. Textual scholars began to turn their attention to “depict[ing] a highly unstable culture of texts, whether that instability is discerned in original renaissance objects and in their modes of composition and production, or in their editorial representation and transmission up to the present.” In this mission, photographic facsimiles (both analogue and digital) began to play an important role.

While facsimile technology was itself not new to the world of textual scholarship, it gained a new significance in the era of social/materialist bibliography and New Textualism for its ability to reproduce not only the words, but also the material features of early texts. Diplomatic reprints, such as Holderness and Loughrey’s *Shakespearean Originals: First Editions* series, were completely overshadowed by photographic facsimile editions, such as the third series of the *Arden Shakespeare Editions* in 1995 and the launch of Early English Books Online in 1998, both of which took advantage of image-based reproduction to enable users to access surrogates of textual artifacts. Other facsimile editions were created that questioned the myth of the authoritative edition and laid textual alterity bare, in an act of “unediting” wherein idealist interpretations of texts are, in Marcus’s words, resituated as “one interpretive agenda among others,” and in which the goal is “to place that current of opinion within a wider network of varying opinion.” Michael Warren’s *Complete King Lear* is one example of such an edition,

---

44 Quoted in Egan, “Materialism,” 190.


46 Ibid., 191-92.

and Cyndia Susan Clegg’s *The peaceable and prosperous regiment of blessed Queene Elisabeth*, with textual commentary by Randall McLeod, is another.\(^{48}\)

Facsimile editions played, and continue to play, an enormous role in achieving the goal of enabling virtual access to original artifacts with little remediation (by contrast with the critical editions of the New Bibliographers), while allowing social/materialist textual editors to represent texts both as artifacts and as products of a complex system of transmission. As we shall see, efforts to represent textual materiality persist, in the field of image-based humanities computing, and digital facsimiles continue to play a powerful role.

### 2.2 Theoretical Foundations of Image-Based Humanities Computing

As the creation and storage of digital facsimiles became less expensive through innovations in imaging and storage, and the means to deliver them to a public became more convenient through the emergence of the web, the struggle for the text increasingly shifted into the digital medium. At the same time, the dynamic nature of digital representation seemed poised to fulfill social/materialist textual scholars’ objective of representing the plural nature of textual witnesses in a non-hierarchical way.\(^{49}\) In this section, I discuss the use of the computer as a venue for representation of textual artifacts, and touch on some caveats and objections to this use that have been raised.


At the intersection of Jerome McGann’s work in social/materialist theories of textuality and digital textuality is the representation of material texts through digital images. In “The Rationale of Hypertext,” McGann’s principal argument is that using codex forms as critical tools for analyzing other codex forms presents an inherent structural limitation, from which hypertext, a “machine of knowledge” that exists on a higher scale of complexity, promises to free the critical edition. Similarly, although codex facsimile editions allow for investigation of some of the material elements of the original, McGann laments that “Because the facsimile edition stands in a one-to-one relation to the original, it has minimal analytic power.”

The obvious implication of McGann’s position in respect to hypermedia extends in part to representing material textuality through image-based digital editions, or archives, as he prefers to refer to them. However, throughout much of his work, as in The Blake Archive or The Rossetti Archive, McGann repeatedly points out the critical distinction between character-based and image-based digital texts: “The practical objection to image-based editing centers in the electronic structure of the digital image. The information reconstituted in electronic images cannot be searched and analyzed—at least not in the way that electronic texts can be searched and analyzed.” It is this problem that has spurred attempts to create methods of digital annotation of images such as McGann’s Inote, and other image markup tools (discussed in section 2.3.3). For McGann, the seeming inability to submit images to computer analysis is a critical obstacle.

The 2002 special issue of Computers and the Humanities on image-based humanities computing lays out a brief history of the field, with particular attention paid to how it relates to textual scholarship. Matthew Kirschenbaum, in his introduction, characterizes the origin of image-based approaches for representing texts as a product of the intersection of social/materialist


51 Ibid., 62.


theories of textuality (such as those discussed in section 2.1) and the advancement of imaging and computing technology that has allowed for the creation of high-quality, low-cost images and the means to disseminate them electronically.\textsuperscript{54} Kirschenbaum also identifies a benefit of image-based humanities computing that circumvents the stumbling block that McGann confronts, that images cannot be searched and analyzed electronically. Kirschenbaum points out that while computers are generally treated, in the field of digital humanities, as a tool for analysis, their use as a “venue for representation” should not be overlooked.\textsuperscript{55} This assertion speaks profitably to McGann’s observation in “The Rationale of Hypertext” that the value of facsimiles lies in their ability to make available the material properties of physical texts, and increases access to rare works.\textsuperscript{56}

There is, it should be noted, a strong theme of skepticism regarding the visual representation of textual artifacts through digital means in the literature on image-based humanities computing. Siân Echard’s chapter on the British Library’s Turning the Pages initiative in \textit{Printing the Middle Ages}, for instance, raises several concerns about the allure of the digital facsimile as a proxy for the physical text.\textsuperscript{57} One of her objections is that high-quality digital facsimiles are often uncritically substituted, in the rhetoric surrounding such projects as well as in the minds of many viewers, for the artifact itself, without considering what information is discarded in privileging of visual aspects of the artifact over all other considerations. This condition leads to an illusion of direct access to the material artifact. Brian Winston voices a similar caution in “‘The Camera Never Lies’: The Partiality of Photographic Evidence,” that too often viewers conceive of the image-based digital text as an infallible representation of a book, neglecting to recognize not only that transmission of any sort, including digital facsimile, involves some form of transformation, but also that the book the images represent is but one unique copy of the text.

\textsuperscript{54} Kirschenbaum, “Editor’s Introduction,” 3.

\textsuperscript{55} Ibid., 4; emphasis in original.

\textsuperscript{56} McGann, “Rationale of Hypertext,” 56.

Finally, Echard notes with apprehension the paradox of the digital facsimile, that in the attempt to simulate the original and thereby bring the experience of it closer to the viewer, not only is the text alienated from its physical context, but the digital proxy provides a rationale for the holders of the original to further restrict access to it. Echard’s perspective echoes that of Jay David Bolter and Richard Grusin, whose 1999 book titled *Remediation: Understanding New Media* outlines the danger inherent to the act of translating information from one medium to another, that the desire for authenticity in the proxy leads to a denial of the original. In the double-logic of remediation, Bolter and Grusin assert, “our culture wants to both multiply its media and to erase all traces of mediation: ideally, it wants to erase its media in the very act of multiplying them.”

The optimism and objections that have been voiced regarding the representation of textual artifacts through digital means suggest that a balanced approach is needed in managing the relationship between the represented object and its representation. McGann and Kirschenbaum both note that the computer’s power to represent rich, high-quality images of textual artifacts, outside the confines of the codex medium, offers distinct advantages to enabling access and exposure to many forms of the material text. On the other hand, Echard, Winston, and Bolter and Grusin are right to warn of the slippage that can occur when we begin to equate representations, image-based or otherwise, with the originals that they represent. It is important to acknowledge that no digital representation can or should take the place of the material artifact as the primary site for investigation, but rather should be used in conjunction with it, while still providing the benefits of its own medium as a virtual object.

---


2.3 Implementations of Image-Based Humanities Computing for Textual Scholarship

Given the benefits, as well as potential pit-falls, of using digital means of representing textual artifacts described in the previous section, how have researchers and institutions fared so far? In this section, I assess several major projects that use digital facsimiles to represent textual artifacts, and in particular identify the ways in which these projects succeed, and the ways in which they fail. I also examine some tools that enable users to manage or enrich their image-based data. Finally, I review the techniques and best practices associated with such initiatives.

2.3.1 Projects

One the most influential projects in digital image-based textual artifact studies is Early English Books Online (EEBO). EEBO is a web-based repository of over 125,000 texts, digitized from an original microfilm collection of titles from the Pollard and Redgrave Short Title Catalogue (STC I) and Wing STC II, along with the Thomason Tracts and the Early English Books Tract Supplement. EEBO is also working with the Text Creation Partnership (TCP) to create character-based versions of the image-based digitized books currently in the EEBO repository. As of 2011, 14,823 texts have been keyed in, and EEBO and TCP intend to apply the text-creation process to the entire EEBO collection, making the text of each book computer-searchable.

As of 2009, nearly a thousand institutions have access to the repository, and during the 2008-2009 academic year, it was accessed over 1.33 million times.


61 Magedanz, “Early English Books Online,” 2084; “EEBO-TCP: Early English Books Online,” Text Creation Partnership (http://www.textcreationpartnership.org/tcp-eebo/); one feature that makes this project unique is that texts are transcribed by humans, rather than generated via an optical character recognition program.

As a tool for textual scholars, EEBO has both strengths and weaknesses. EEBO provides access to an enormous selection of print artifacts in facsimile, providing scholars in institutions without large rare-book facilities materials to work with. EEBO also makes for an effective teaching tool, such as to illustrate lessons and provide material to students.\textsuperscript{63} Furthermore, the digitized material has significant advantages over its microfilm predecessor, including ease of use, clearer images, and lower costs for institutions.\textsuperscript{64} On the other hand, critics have pointed out that a slippage occurs both in the marketing of EEBO and in scholars’ (especially students’) understanding of the resource, causing the following three misleading impressions:

1. \textit{Examining a text in EEBO is equivalent to examining the physical text.} In fact it is a heavily remediated reproduction, several steps in the reproduction chain removed from the physical text.\textsuperscript{65} Furthermore, it does not contain pertinent physical information such as size, texture, or—critically—colour.\textsuperscript{66}

2. \textit{EEBO can be used as a bibliography, equivalent to the English Short Title Catalogue (ESTC).} In fact it is nowhere near as inclusive,\textsuperscript{67} and using it as a bibliography can easily confound any effort at comprehensive research.

3. \textit{The copy reproduced in the facsimile is the same as all other copies of the edition.} In fact, on top of the fact that no two copies of an edition are likely to be identical

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{63} Ibid., 680, 688.
\item \textsuperscript{64} Ibid., 685.
\item \textsuperscript{66} Gadd, “Use and Misuse,” 682.
\item \textsuperscript{67} Ibid., 686.
\end{itemize}
\end{footnotesize}
in every respect, no attempt was made in the initial microfilming process to select a copy that resembled a common state of the edition. 68

Diana Kichuk also points out a variety of critical failures in the digital-image creation itself, including content amputation, page distortion, deterioration of microfilm, substandard resolution, and additional loss of detail caused by scanning the already extremely high-contrast microfilm only as bitonal digital images. 69

*Electronic Beowulf* (now in its third edition) and *Electronic Boethius* are prime examples of the use of a digital-image-based approach to facilitating access to extremely rare, delicate, and/or valuable material, by creating a surrogate. Both of the manuscripts that these projects digitize are the sole surviving witnesses of the texts they contain. The primary source in the *Electronic Boethius* is British Library MS, Cotton Otho A. vi, the only surviving copy of the prose and verse translation, and *Electronic Beowulf* is based on British Library MS Cotton A. xv., *Beowulf*’s sole surviving medieval witness. Both manuscripts are also extremely fragile, having been badly damaged in the Cotton Library fire of 1731. 70 These projects are closely related to Kiernan’s initiative with James Griffioen and Brent Seales, the *Digital Atheneum Project*, which seeks to develop restored surrogates of the Cottonian Collection using specialty lighting techniques and two- and three-dimensional digital image capture and processing. 71

68 Ibid., 686-87.

69 Kichuk, “Metamorphosis,” 298-99; the word “bitonal” refers to a digital image with only black and white pixel information.


These digital editions are valuable resources for Beowulf and Boethius scholars, in that they include full translations, extensive XML markup (using the University of Kentucky’s Edition Production and Presentation Technology, described in section 2.4), and images created using sophisticated techniques, such as ultraviolet and fiber-optic backlit imaging. While this kind of project represents an enormous asset, particularly to the study of witnesses that are extremely rare, fragile, or inaccessible, this approach is particularly expensive, requiring specialized imaging technology, and researchers/staff knowledgeable in using the EPPT software and working with XML.

Armadillo Systems’ Turning the Pages software (TTP; now at version 2.0), as used by the British Library for their Treasures collection, is of interest as a digitization project relevant to textual scholarship insofar as the company produces digital surrogates that are primarily designed to supply the experience of encountering the artifact in situ. According to their company descriptions, Armadillo Systems “have developed a methodology of approaching bibliographic projects that encompasses thinking about … the book as object.” TTP is a platform for digital books that strives for realistic three-dimensional manipulation, particularly page-turning, and enables magnification, windowed commentary, audio accompaniment, and searching, among other features. The original TTP, as implemented by institutions like the British Library, is accessible online using a computer running Windows Vista or assisted by a plug-in.

While TTP demonstrates the aesthetic appeal and multimedia potentials of creating digital exhibitions of rare books, there are significant problems with using this resource for textual scholarship. The most fundamental issue relates to the distinction between realistic and real. While the apparent goal of a TTP digitization project is usually, as in the case of the Wellcome Library’s implementation of the software, “to make as realistic a digital version of a precious


book as possible,“ the TTP digitization process and software achieves this, ironically, through heavy modelling and remediation of the subject images. The images themselves are cropped, colour-corrected, and downsampled, and special areas such as gilding are isolated and overlaid with a specularity map. The pages are filmed being flipped, and this information is used to create a three-dimensional CAD model for the animation effect that simulates page-turning for the user. The animated page-images are darkened and lightened in relation to simulated light conditions, and areas treated with a separate specularity map behave differently than the rest of the page in relation to simulated light sources, such as in the instance of gilding, where that part of a page is replaced with an imitation “gold” effect. While this approach may be crowd-pleasing, it renders the digital facsimile practically useless to the textual scholar wishing to examine the artifact, as any evidence that could be gleaned from the images themselves is thrown into doubt by the high degree of simulation. In this sense, Echard is astute in her observation that TTP distances audiences from the medieval texts they have digitized, through their very attempt to make the artifacts present.

A promising project in the use of a digital-facsimile approach for the examination of structural relationships within a textual artifact is Neal Audenaert and Richard Furuta’s Annotated Facsimile Editions (AFED). This project’s primary goals are similar to those of the present project: openness to new data and research paradigms supplied by users; nonhierarchically structure to accommodate the inherent nonhierarchical nature of some features of texts; the ability to structure the representation in ways other than those permitted by the physical artifact; and the ability to align a given facsimile with other representations of the same work to enable comparison. AFED enables both sub-image-level annotation and annotation across multiple

---


75 Ibid., 102-103.

76 Echard, “Coda.”

image files. AFED’s primary concern is for enabling navigation rationales other than those imposed by physical artifact structure, such as shifting directly from chapter head to chapter head, or from a position in one witness to the parallel position in another witness. It is less concerned with visually modelling the nature of the relationships among navigational points, and leaves aside other possible applications of digital-facsimile markup, such as the ability for the user to sample and/or manipulate portions of digital facsimiles.

From examining several projects in digital-facsimile representation of textual artifacts, several conclusions emerge. The example of EEBO shows that accessibility of large amounts of digitized texts is often accompanied by poor image-quality. On the other end of the spectrum, Kiernan’s electronic editions demonstrate that extremely complex, value-added image-viewing and handling environments can be useful for particularly rare materials, but are too costly and labour-intensive for the average individual scholar to generate without a team and considerable funding. The implication here is that there is a balance to be struck between labour-intensiveness and quality, resulting in a resource that offers high-quality images but has a simple design that makes development and use feasible on a small scale. Turning the Pages demonstrates that manipulation of images destroys their value as evidence; while TTP may please viewers looking to be wowed by stunning visual effects, for textual scholars looking for evidence this collection of digital facsimiles is useless. AEFD represents a promising development in the area of creating environments for active engagement with the materials, such as reorganization and user-defined navigational rationales, but more work can be done in facilitating modelling activities, as well as sampling and manipulation of images.

2.3.2 Tools for Managing Image-Based Data

A popular tool for creating semi-structured data for images is the UVic Image Markup Tool, a free open-source tool that allows users to create XML (eXtensible Markup Language) markup for images, and either generate XHTML output (a combination of XML and HyperText Markup

78 Ibid., 145.

79 Ibid.
Language, or HTML) using the program itself for immediate use, or incorporate the markup in other projects. The design of the tool was prompted by the observation that existing tools either require expert knowledge to use, such as Kiernan et al’s EPPT (discussed below), or use a proprietary input file format, such as the University of Virginia’s Inote. Therefore, the team at the University of Victoria built a tool that was simple to use, accepts a range of file formats, and creates XML that is TEI P5-compliant. Unfortunately, thus far the UVic Image Markup Tool is available for the Windows operating system only.

A more sophisticated toolkit, gradually developed by Kevin Kiernan and others over nearly ten years, is the Electronic Production and Presentation Technology software (EPPT). Consisting of a group of plug-in applications running on the open-source Eclipse platform (originally created by IBM), the EPPT consolidates many stand-alone tools Kiernan and his various teams have developed over the years for creating Electronic Beowulf, Electronic Boethius, and other digital editions. The tools incorporated into this kit enable image-tagging, which is the basis for their model of side-by-side image-and-text transcriptions, as well as many other functions, such as the ability to overlay images of the same artifact created using different photography techniques, glossary generation, and even hypothetical reconstruction modelling. Despite EPPT’s claim that “humanities scholars and their students, who typically have little or no prior knowledge of XML/TEI markup or encoding, can set to work with EPPT with very little training,” the toolkit is daunting even for someone with experience in both bibliographic description/analysis and XML, and usability issues are exacerbated by the toolkit’s fairly inflexible input demands. Furthermore, EPPT is, like the UVic Image Markup Tool, currently available only for Windows.


82 “Edition Production & Presentation Technology (EPPT).”
While EPPT is an impressive achievement as an image-based approach to representing textual artifacts, it is not simple, flexible, or accessible enough to accommodate the casual user.

2.3.3 Best Practices and Techniques in Digital-Facsimile Creation

This section of the literature review collects several professional perspectives on standards and practices in digitization. This information is relevant to the current study insofar as a digital environment for accessing, organizing, and manipulating digital facsimiles should ideally incorporate the best-possible calibre of input. The considerations laid out by these authors offer a framework for evaluating this input.

Melissa Terras’s *Digital Images for the Information Professional* provides an overview of not only the technical and operational considerations in creating digital image-based resources and repositories, but also the theoretical, historical, cultural, and ethical issues of digital images. Terras identifies two particularly pressing issues in the production of digital facsimiles of physical documents: colour, and image quality. Terras writes, “One of the trickiest issues surrounding the production of digital image surrogates of analogue documents, objects and artefacts is the issue of colour accuracy in reproductions.”

Terras advocates that digitization programs adopt a commonly-held colour profile, such as Adobe 1998 or sRGB, ensure proper colour calibration of all devices, and include a standard reference at some point in the digitization, such as the ISO’s IT8 target, a


84 Ibid., 185.

85 Ibid., 186-87.
colour chart to be included in the set of digitized images and photographed under the same conditions.  

Regarding image quality, Terras adopts the same general standard as Czeslaw Jan Grycz (discussed below), that “quality … can be judged by how accurately the digital image reflects the original artefact it represents, and the relationship of the surrogate to the original.” Because the accuracy of a digital facsimile cannot be judged in any other way (a limitation that will be discussed in chapter 4), Terras advocates that the digitization should always be verified against the original artifact.

Czeslaw Jan Grycz offers a particularly detailed discussion of the digitization of rare books and manuscripts within a rare-book-library context, in his chapter for the reference book *Digital Heritage*. Grycz’s discussion centres on the digitization of a rare text as the creation of a surrogate, to be used in various contexts where access to the physical artifact is either impossible or unadvisable. Grycz lays out three qualities of digital images of rare materials that he offers as guidelines for creating digital facsimiles, and which can similarly be used for evaluating such materials: 1) faithfulness, 2) completeness, and 3) authenticity.

Faithfulness entails representing the book’s native characteristics as closely as possible in its digital surrogate. An example of a guideline for maintaining faithfulness is to retain the opening, or double-page spread, as the default view of the text, as it is with the original artifact. Creating faithful representations also requires filming in light that represents the artifact’s colours and textures most accurately, which means natural full-spectrum lighting set off to an

86 Ibid., 189.
87 Ibid., 192.
89 Ibid., 41.
angle to create a “raking” effect.\(^{90}\) Completeness describes including all phenomena present in the original. Here Grycz reveals considerations that are also at the heart of the social/materialist bibliographer, in that he includes in this list post-print paratextual materials, such as binding, fly-leaves, and “any other artefact or ephemeral item that may be closely associated with the book.”\(^{91}\) Authenticity requires that the digitizer capture and provide the information necessary for the user to infer details of the original artifact. This requirement includes metadata (descriptive textual data associated with a file, similar to catalogue information associated with a book) on dimensions, colour values, and details on the imaging technology and process used.

Grycz’s criteria for creating quality digital facsimiles are helpful guidelines not only for digitization, but also for evaluating the quality of digitization initiatives such as EEBO and Internet Archive. However, it must also be borne in mind that these guidelines not only describe conditions in a best-case scenario, but also describe requirements for a digital facsimile to be used as a surrogate, in the absence of the original. Some requirements he lists, such as the need to shoot the opening as a default unit of viewing, may be unnecessary, and even a hindrance, in the context of the current study where the goal is to imagine the use of digital facsimiles in versatile tools to be used alongside, and thereby verified against, the physical artifact.

### 2.4 Summary

This literature review was designed to provide a background for the present study, which seeks to tackle the problem of how to design digital environments that deliver facsimiles of rare books in a manner that supports the needs and interests of modern textual scholars. In order to do so, I have examined the history and current state of modern textual scholarship, and in particular the evolution of its methodologies, and the ways in which the representation of text has evolved over time. What my review has revealed regarding methodology is that the systematic and detailed examination and analysis of the textual artifact as a physical object created through a complex process of transmission has been central to the methods of textual scholarship since the

\(^{90}\) Ibid., 44-5.

\(^{91}\) Ibid., 46.
turn of the twentieth century. Over time the scope of what should rightly be included in this investigation has increased as the scope of what is considered part of the network of transmission has expanded. What my review has revealed regarding representation of the text describes a significant shift in opinion about what constitutes the text. Initially, bibliographic methods were deployed in order to determine, and ultimately deliver to readers, a single, authoritative, idealized text, isolated from the mediating factors of non-authorial agents in the transmission process, and from the various material instantiations of the text. However, this position was challenged as the rejection of both the spectre of the authoritative text and the traditional editorial agenda took hold, coupled with a new appreciation of the role and scope of non-authorial agents in the transmission process as well as the significance of the material context (the “bibliographic codes”) of a text. What resulted was a mandate to give credence to the multitude of material forms that texts have taken over time, a mandate that is reflected in the use of facsimile technology to deliver representations of the text that express textual alterity and material situatedness.

The review of the foundations of image-based humanities computing, particularly in relation to the representation of textual artifacts, reveals that the agenda just described was one of the driving factors in the development of image-based humanities computing. As McGann and Kirschenbaum both note, computers offer a powerful tool for the representation of the text as a material object. However, this review also revealed that there are limitations to representation of texts through images, so much so that it is misguided to consider digital images, rather than the objects they represent, as an acceptable primary object of investigation in textual scholarship. The limitations suggest that this thesis’s proposal, that digital facsimiles should be used as tools in the process of on-site examination of textual artifacts, may be the only way to properly verify the evidence delivered by the digitizations.

Finally, this review examined the current state of digital-facsimile initiatives in textual scholarship, including current digital-facsimile environments, tools for managing image-based data, and standards and considerations for evaluating the quality of digital facsimiles. The review of other digital-facsimile environments revealed that accessibility of large amounts of digitized texts is often accompanied by poor image-quality, and also that the creation of extremely complex, value-added image-viewing and handling environments can be useful for particularly rare materials, but are too costly and labour-intensive for the average individual.
scholar to take on. Turning the Pages acts as a cautionary tale that optimizing a product for one audience may ruin it for another, and demonstrates how the relatively common activity in digital imaging, of applying effects, ruins the value of the facsimile as evidence. Promising developments have been made in the area of creating environments for active engagement with the materials, but more work can be done in facilitating modelling activities, as well as sampling and manipulation of images. The review of tools for managing image-based data revealed that while simplicity and support for open-source file formats are preferable for the general user (as a comparison between UVic’s Image Mark-Up Tool and the University of Kentucky’s EPPT software showed), limitations relating to operating system compatibility still persist. Review of professional perspectives regarding digital facsimile creation established that in order for digital images to be effective representations, they must be faithful, complete, and authentic representations of artifacts, with particular attention paid to the issue of accurate reproduction of colour.

Throughout the remainder of this thesis, I will return to the issues discussed in this literature review. In particular, the themes of the central importance of the textual artifact as a primary site of evidence of a text’s transmission history, the social/materialist goal to represent textual alterity and material situatedness of the text, particularly through facsimile, and the inherent limitations of facsimiles to represent the textual artifact, will be focuses in the discussion chapter of this thesis (chapter 4). Further, the critique of concurrent digital-facsimile environments and image-markup tools will inform my own design of a digital-facsimile environment to facilitate textual investigation (also chapter 4), and the guidelines set out regarding qualities of responsible digitization practices will be taken into consideration when selecting input for the prototyping experiment (described in chapter 5).
3 Methodology

The overarching goal of the present study is to develop a model for the relationship between textual artifacts and their image-based digital counterparts that privileges the perspective of the textual scholar by establishing the digital facsimile as a tool for exploring the physical text. As Warwick tells us, too many projects in humanities computing fail to take into account how their target users actually perform their researches, resulting in ineffective digital tools. Therefore, developing a clear and nuanced understanding of the perspectives of textual scholars themselves is essential. Is direct access to physical texts really of paramount importance, and if so, why? How do digital facsimiles help, and how do they hinder the examination of textual artifacts? What activities are most important in textual research, and how can a digital infrastructure support these activities? These are the questions that the present research seeks to answer, using a combination of systematic journal-article analysis, interviews, and digital prototyping.

3.1 Research Design

As I discussed in the introduction, the research design for this thesis is largely informed by John Unsworth’s concept of scholarly primitives, “a list of functions (recursive functions) that could be the basis for a manageable but also useful tool-building enterprise in humanities computing.”92 The present study uses the concept of scholarly primitives as a guiding principle, and seeks to test the hypothesis that there are scholarly primitives that are common, not to all humanities researchers generally, but, rather, particularly to textual scholars. The research component of this thesis is designed to uncover these scholarly primitives, through reviewing the techniques described in a selection of recent articles published on topics of textual scholarship, and through conducting a series of interviews with both established and aspiring textual scholars.

The research module of this thesis consisted of three major components. In the first component, I aimed to establish preliminary hypotheses about what types of research activities, or scholarly

92 John Unsworth, “Scholarly Primitives.”
primitives, are shared by textual scholars. In order to achieve this, I reviewed a sample set of recent publications in the fields of book history, bibliography, and textual criticism. In the second component, I sought to verify and/or challenge the hypotheses I had developed, and cultivate a more sophisticated understanding of the research habits of my target group. To achieve this, I conducted a series of interviews with both established and aspiring textual scholars. In the third component of the research process, I synthesized the data and knowledge gathered from the literature review, recent publications review, and interview process, and used it to inform the design of a digital-interface prototype that allows users to access, organize, and manipulate digital facsimiles in ways that embody the prevalent methodologies observed. As part of this latter component, I implemented a portion of the digital model in the form of a web-based prototype, which serves as a means to test the efficacy of the proposed model.

3.1.1 What is a Textual Scholar?

This study aims to identify the research behaviours associated with textual scholarship, and to imagine ways in which digital platforms can be structured so as to aid these behaviours. Therefore, it is important to have a sense of what interests and activities are part of being a textual scholar, at least in the context of the present research.

Textual scholarship, in this study, refers to the research carried out by bibliographers, textual critics, book historians, and editors of scholarly editions. This list is informed by the content of the literature review (chapter 2), but does not claim to be exhaustive. Rather, since the focus of this study is on how the relationship between the physical textual artifact and its digital counterpart can be modeled to privilege textual scholars as a user-group, this study focuses particularly on textual scholars for whom the physical artifact is a primary site of investigation. This definition can include forms of textual scholarship that deal with born-digital texts; however, since these forms do not typically include digital facsimiles of their subject matter, they will not be dealt with directly in this thesis.  

93 Notable works in these areas are Stephen E. Jones’s *The Meaning of Video Games: Gaming and Textual Strategies* (New York: Routledge, 2008), and Matthew Kirschenbaum’s *Mechanisms* (Cambridge, MA: MIT Press, 2008).
3.2 Procedures and Data-Gathering

3.2.1 Systematic Journal Review

While the literature review for this study allowed me to observe methodologies practiced throughout the history of modern textual scholarship, such a review lacks observation of how textual scholars practice their research in the current day. Accordingly, as a precursor to the interview process, I undertook a review of a systematic sampling of recent publications in six leading journals of textual scholarship: *The Library, Textual Cultures, Studies in Bibliography, Papers of the Bibliographic Society of Canada, Papers of the Bibliographic Society of America*, and *Book History*. In order to strike a balance between comprehensiveness and manageability, I selected one article from each year’s worth of published articles for each title, from 2006 up until each journal’s last full-year’s worth of publication.\(^\text{94}\) For most journals at the time of the selection process, that year was 2010.\(^\text{95}\)

The articles were analyzed for whether common activities exist among textual scholars that should be facilitated in a digital toolkit for performing textual research. Analysis revealed several activities that were common to many of the articles reviewed. The most common activity was that of comparison. Scholars engaged in comparisons both synchronic (such as comparison of the text of multiple copies of a single edition) or diachronic (for instance, comparison across multiple editions of a text). Second-most common was the employment of modes of graphic or visual representation of findings, including tables, charts, graphs, and models of hypothetical material relationships within individual texts and among multiple texts. Related to this common activity were two other activities, which rarely appeared in isolation but

\(^{94}\) For each year’s worth of articles for each journal, the articles were numbered, and the article to be reviewed was selected using a random number generator (www.random.org). This ensured that the article chosen as the sample for each year from each journal was selected impartially.

\(^{95}\) *Studies in Bibliography* has been publishing sporadically of late, and yielded only two articles for the time period, while *The Library* had published their full 2011 offerings at the time sampling occurred, and so yielded six samples, for a total of 28 articles.
commonly accompanied activities of graphic/visual representation. One was the quantification of observed data, predominantly counting and calculating. This quantification frequently took the form of a chart or graph. The other activity was that of pattern-seeking. References to correlations among observable data were present in many of the articles reviewed, but nowhere was this behavior more prevalent than in conjunction with acts of graphic or visual representation.

The other major activity that I observed was the analysis of the physical structure of the artifact, which appeared as the third-most common activity, after comparison and visual/graphic representation. In this activity, different discrete parts of the text as a physical object are related to one another in order to glean insight into the material structure of the textual artifact, and sometimes to reveal the design rationales that went into the book’s structure. Throughout this thesis, I have termed such activity *material analysis*, based on an analogy between this activity and material analysis in archaeology: “Material analysis,” writes R. Gebhard, “has two main parts: the characterization of the material of which objects consist and the characterization of the technical treatment or the manufacturing of the objects. The first can, for instance, give information about the provenance of the material, while the latter helps to reconstruct ancient techniques.”⁹⁶ This definition is applicable to the activity of textual scholarship under discussion particularly in respect to the latter part of this quote, in that material analysis in textual scholarship tends to be concerned with revealing elements of the production of the artifact.

Findings of this stage in the research process, combined with conclusions drawn during the literature review, form the basis for the lines of inquiry during the interview sessions.

---

3.2.2 Target Informants

Participants in the interview portion of this study fall into two groups. The first is comprised of textual scholars who are established in the field, those who are no longer students and who either hold a post teaching some form of textual scholarship or who have published on topics in textual scholarship. The second group is made up of aspiring textual scholars, who are attending courses in textual scholarship or pursuing a topic in textual scholarship in their studies. Since this research treats of an emergent factor in textual studies, namely the widespread availability of digital facsimiles, I was interested in gaining a perspective on the issue both from scholars who have been practicing since before high-quality digital facsimiles were commonly available, and from students who have always had such resources at their disposal.

Participants were gathered using three primary methods. Initially, some authors of the sampled journal articles, whose work exemplified one or more methodological trends (described above), were approached to participate. The second method was to put out a general call for participation on the University of Toronto Book History and Print Culture (BHPC) listserves. These listserves deliver mail to both students and faculty of the BHPC program, which is a collaborative program that offers courses in book history, bibliography, and other forms of textual scholarship. The third participant-gathering method was a “snowballing” technique, in which participants or others peripheral to the study approached colleagues and suggested participating in the study.

Group 1 is comprised of a total of seven participants (listed alphabetically): David Greetham, Distinguished Professor at the Graduate School of the City University of New York (English); Andrew Murphy, Head of the University of St. Andrews School of English; William Sherman, Professor of Renaissance and Early-Modern Studies at The University of York; Markus Stock,

---

97 Because this portion of the research program involves human subjects, approval of the research plan was obtained from the University of Toronto’s Office of Research Ethics. All measures were taken during the course of this study to ensure the security and privacy of the participants, including proper secure data-storage practices and a clearly-worded informed-consent process that included the ability to indicate either confidential or credited participation in the study. A copy of the Office of Research Ethics approval is provided in the Appendix.
Associate Professor of German and Medieval Studies at the University of Toronto; David Vander Meulen, Professor of Bibliography and Textual Criticism and Eighteenth Century Literature at the University of Virginia’s Department of English; and two confidential participants. Of these two participants, one holds a post at a Canadian university in the subjects of both bibliography and digital services, and the other is Professor Emeritus of English at a Canadian university.

Group 2 is comprised of five students: Elisa Tersigni, PhD student in the English Department at the University of Toronto; Matthew Schneider, also PhD student in the English Department at the University of Toronto; Kirstyn Leuner, PhD student at the University of Colorado; and two confidential participants. Of these participants, one is a PhD student in an English department, and one is a Master’s student at a school for Library and Information Studies.

3.2.3 Interviews

Generally speaking, the interview process performed the following three functions: first, it enabled me to test the hypotheses established through the literature-review and systematic journal-review stages of this study. Second, it allowed for a more in-depth look at the processes, rather than simply the products, of textual scholarship, including the social processes not visible in the systematic journal review. Third, the interviews presented an opportunity to perform a type of user-needs assessment, wherein participants were able not only to discuss the benefits and failings of current digital tools, but also to describe the sorts of functions they require in such a resource.

Interviews lasted approximately one hour each, and, while they were in person whenever convenient, since the majority of participants lived outside of Toronto, many interviews were conducted via Skype-to-Skype or Skype-to-telephone connection.

Since preliminary hypotheses were established prior to the interviews, during the journal-article analysis, a rigidly structured set of interview questions based on these hypotheses would have carried the risk of leading the interview subject to particular responses, and, ultimately, would have produced biased, predetermined results. Therefore, a merely semi-structured approach was used, in which participants were asked to comment on themes and reflect on their own perspectives.
The interviews were guided by a set of 14 questions or topics of discussion (the full interview guide is included as an appendix). The first three questions were very general, aimed at determining the nature and scope of the participant’s involvement in textual scholarship, her or his history within the field, and the sorts of methods, tools, and techniques he or she has used.

The next two questions sought to determine the researchers’ perspectives on the importance of access to textual artifacts, and their opinions on the utility of digital facsimiles. These questions were designed with the two prevalent models of the relationship between the physical artifact and the digital facsimile in mind (discussed extensively in chapter 1). In other words, the question regarding the importance of direct access to the textual artifact was designed to determine the extent to which the current trends of using digital facsimiles as replacements for physical artifacts, on the one hand, or as surrogates for the physical artifacts, on the other hand, satisfy the needs of textual scholars. Furthermore, the question pertaining to the utility of current digital facsimiles was designed to determine not only what current benefits researchers identify, but also what is left to be desired. The ultimate goal of such a line of questioning is, of course, to assess what qualities would be needed in an image-based digital resource to make it useful to researchers.

The third line of questioning in each interview pertained more specifically to the observations made during the systematic journal-review, namely, the hypothetical scholarly primitives of comparison, material analysis, use of quantitative data, and modelling/visualization. This line of questioning was designed to verify the findings of the systematic journal review, but it was also important to determine whether these behaviours were truly part of the investigative process, or whether their prevalence in the journal articles was influenced by conventions of scholarly publication. For instance, participants were asked to reflect to what extent they employ graphic or visual representational techniques, and, if so, whether they employ them only during the presentation of their research or whether they play a role in the analytical stage as well.

A final, fourth line of questioning regarded the participants’ current use of technology in their research. Participants were asked not only to describe the extent to which they use computing devices in the various stages of their work (discovery of materials, investigation, analysis, and presentation), but also to list the devices they either own or have access to, and whether they employ any of these devices in the context of the rare-book room or archive. The rationale
behind this line of inquiry is to identify what device and/or platform would be the most appropriate choice to design digital tools for, based on researcher preference and access.

Several other questions that do not fall into these main groups were also put to researchers, including one about the extent to which secondary resources played a role in the investigatory stage of their studies. This question was designed to provide insight into how important it is to researchers that digital resources should be designed to integrate other resources, online or otherwise.

3.2.4 Data Recording, Reduction, and Analysis

Interviews were audio-recorded, and the recordings were then subsequently reviewed and summarized in point form. Interview data was analyzed using a thematic-coding-and-analysis method adapted from the *Sage Encyclopedia of Qualitative Research Methods*. This approach is most appropriate for this study for several reasons. First, thematic coding begins with a range of themes known or anticipated to be in the data. Since interview participants were questioned on a variety of topics that were thematic in the systematic journal review, thematic grouping of the interview material facilitated the verification of those preliminary findings. Second, thematic coding and analysis emphasizes not only the continuities among participants, but also the correlations of themes within individual participant responses. In relation to the search for scholarly primitives, thematic encoding provided a useful tool, since simple aggregation of the data relating to particular behaviours is not enough. For instance, a majority of participants may endorse one technique, and a majority endorses another technique, but if there is little correlation for individual participants between these two techniques, the case for considering them both to be essential practices of textual scholarship is weak. The results of this analysis forms part of the discussion chapter of this thesis, chapter 4.


99 Ibid.
3.3 Prototyping

In order to intelligently design tools for textual scholars, the research component of this thesis asks the question, *What is the foundational methodology of textual scholarship?* The answer to this question is found in reviewing productions in the field and conducting a series of interviews. The other major question addressed in the research portion of this thesis, *What model of accessing, organizing and manipulating digital facsimiles supports the methodological foundations of textual scholarship?* is pursued through experimental prototyping.

Later in this thesis I present a hypothetical digital interface that enables the digital facsimile to be used as a model for expressing and exploring issues in textual scholarship. Both the overarching rationale, and the various individual features of this environment are predicated on the information about methodologies either inherent or prevalent in textual scholarship. As a way of both testing and demonstrating the fitness of the hypothetical digital environment, this thesis includes a constrained prototyping exercise, in which I select a portion of the proposed environment and carry it out.

The rationale for treating prototyping as a scholarly endeavor is laid out extensively in Alan Galey and Stan Ruecker’s “How a Prototype Argues.” They recognize that “thinking through making” is a key activity within the digital humanities, and that the design of prototypes such as digital interfaces constitutes an act of interpretation, just as does the creation of a scholarly edition of a work. Consequently, Galey and Ruecker argue, “digital artifacts themselves—not just their surrogate project reports—should stand as peer-reviewable forms of research, worthy of professional credit and contestable as forms of argument.”

It was important not simply to create a mock-up or illustration of the prototype, but actually to create a prototype that performs some functions of the larger proposed toolkit, because it allowed me to also test the technical feasibility of such a project. As will be discussed in


101 Ibid.
extensive detail in chapter 5, part of the design rationale for the toolkit is that it is made in such a way that other textual scholars, with limited technical expertise and time, may be able to implement their own additions and alterations to the source code. Therefore, in order to evaluate whether it is realistic to expect that such active involvement on the part of the tool’s users is feasible, I undertook the creation of the prototype, under conditions of constrained time, expertise, and resources.

The methodology of the prototyping exercise, as an activity, draws primarily from the realm of project management. The typical steps included in responsible project management are surprisingly complementary to the act of prototyping as a scholarly endeavor. Just as prototyping can be thought of as a critical interpretation, and in this case is an interpretation of both background research and data acquired through investigation of the target group, responsible project management begins with developing an understanding of the context of the project area, gathering data on the particular problem to be solved or opportunity to be exploited, and projecting and refining a response in the form of a project proposal based on an interpretation of this information. A project-management-based approach also includes a stage wherein a proposed solution (in this case embodied in the prototype) is evaluated against an established set of criteria, which is analogous to the discussion of the proposed digital resource in chapter 4. In a field such as the digital humanities, where so much of its academic output is in the form of projects, the use of methods of professional project management should be standard practice.

The prototype is web-based, uses freely downloadable page-images acquired from the Internet Archive, and is built using a combination of XHTML, JavaScript, and CSS (Cascading Style Sheets). As will be discussed in greater detail in the discussion following this section, the choice of a web-based approach is predicated on several factors. First, when participants were asked about what types of computing devices they used, the responses varied widely, from desktop


103 Ibid., 32.
computer to smartphone, and from Mac to Windows to Linux operating systems. This diversity necessitated a platform that would operate on as many of these devices and operating systems as possible. A web-based application fulfills this requirement. Second, as referred to above and discussed in greater detail in chapter 5, the design of the prototype is influenced by the idea that other textual scholars should be able to make their own improvements to it. Web languages such as those used in this prototype are relatively open to alteration and addition, can be improved upon incrementally, and can be learned relatively easily using online teaching tools such as those offered at www.w3schools.com. The time allotted for the prototyping experiment was constrained (for reasons referred to above and discussed in more detail in chapter 5) to six weeks, which included three weeks of design, planning, and acquiring necessary technical skills, and three weeks of implementation.

3.4 Conclusion

The research design for this thesis combined methods and techniques from several disciplines. The literature review that preceded this chapter on methodology was carried out in a manner common to research in the humanities, while the second stage of the research, consisting of data gathering through both systematic sampling of recent publications of the target subject group, and in-depth interviews with select subjects, is based primarily on a social-sciences model. Interviewing members of a target community and synthesizing their responses into a theoretical framework that will inform a theory about that community is an increasingly common approach within fields that use qualitative research methods, such as anthropology and sociology. Techniques from professional project management also form the principal methodology of the prototyping exercise. Because this thesis is steeped in the humanities background of textual scholarship, requires the investigation of the current, real-life functioning of a particular professional community, and is intended to achieve its primary results in part through project-driven academic scholarship in the form of prototyping, a synthesis of techniques from the humanities, social sciences, and project management was essential.

In the following chapters, I will discuss the findings of this research, and put these findings to work in a design for a digital-facsimile environment that responds to the particular needs of textual scholars and takes full advantage of the digital medium for representing textual artifacts through images.
4 Discussion

Digitization projects have revolutionized our access to resources such as images or manuscripts. … However, the projects that are most innovative are those that use digital resources not for reconstruction or improved access, though these are of course enormously valuable, but as tools to think with.

- Claire Warwick¹⁰⁴

So far during this thesis, I have been concerned with the types of engagement that textual scholars undergo with their materials, and the sorts of thresholds for representation presented by digital facsimiles and the environments created to provide access to them. In this section, I undergo a more detailed analysis of these issues, with reference to the lived experience of both seasoned and beginner textual scholars, and respond critically to these findings through the design of a digital-facsimile environment aimed at facilitating exploration of textual artifacts.

As I described in the introduction to this thesis, Claire Warwick notes that one of the dangerous shortcomings of many projects in the digital humanities is that they fail to fully take into account how members of a target user-group, usually humanities scholars, actually perform their research, and, as a result, creators of new tools expend considerable amounts of time and resources into development only to receive a lukewarm reception from their intended audience. Only by carefully considering how textual scholars investigate their objects of study, and actually applying this knowledge in the design of a new resource, can that resource be expected to make a meaningful contribution. This is the rationale of the present chapter. In order to develop a clearer understanding of what textual scholars do, I have conducted interviews with both aspiring and established textual scholars, and in this chapter I discuss my findings. Following this discussion, I put my findings into action, in the design of a digital-facsimile environment that complements the activities most important to textual scholarship. The ultimate

¹⁰⁴ Warwick, “Print Scholarship and Digital Resources.”
goal of this experiment in design is to lay the foundations for a tool for the textual scholar to think with.

The following discussion is comprised of two major parts. First, I will analyze the responses of the twelve interview participants, and consider the implications for the relationship between physical textual artifacts and the digital facsimiles designed to represent them. Of course, with such a small number of participants, and with the lack of structure of the interviews, no one would venture a claim to the generalizability of these findings. Rather, the interviews are treated here as case studies, in which, as Joachim puts it, “a few instances of a phenomenon are studied in depth.”[^105] The “instances” in this study are textual scholars, and the “phenomenon” under investigation is how these scholars use particular tools and techniques to interrogate the objects of their research. The interview responses allow for greater insight into themes of material investigation, methodology, tool use, and the limits of representation. In the second section, I propose a new environment for accessing, organizing, and manipulating digital facsimiles to facilitate the investigation of textual artifacts, which responds critically to the findings of the interview analysis and to the successes and failures of existing digital resources for textual scholars (discussed at length in section 2.3.1).

### 4.1 Analysis and Discussion of Interview Results

In the following paragraphs, I identify and discuss themes that emerged in the twelve hour-long interviews with both practicing textual scholars, and textual scholars in training. Main topics of discussion in these interviews included the importance of direct access to physical artifacts, the benefits and drawbacks of digital facsimiles, methods of textual investigation, and technical considerations such as computing platforms and imaging technologies.

4.1.1 Physical Artifacts, Digital Proxies

A theme emerged when participants were asked about the evidentiary value of physical artifacts, on the one hand, and digital facsimiles of these artifacts, on the other. Although the individual responses were nuanced, the overall message was clear: while digital facsimiles held promising opportunities for investigation and analysis, particularly through the ability to access content remotely and take advantage of functionality such as extreme magnification and text searching, respondents were skeptical about the reliability of digital images as sources of evidence in general, and as substitutes for access to the physical artifact in particular. Participants overwhelmingly insisted that access to the physical artifact was essential, for reasons ranging from the assurance that comes from being an eye-witness to textual evidence to the benefits of being in the physical space of the library in which textual artifacts are housed. Ultimately, while scholars see potential for using digital facsimiles in their research, there appears to be little possibility that facsimiles can or should replace on-site research. The question for the design of digital-facsimile environments should therefore not be how quality and access can be improved to the point of making the originals obsolete (although quality and access is itself very important), but rather how can interfaces be designed that can enhance the investigatory work that textual scholars perform with the physical artifact.

4.1.1.1 The Role of Physical Artifacts in Participants’ Research

Since individuals were requested to participate because of their use of books and other documentary material as physical evidence, it is not surprising that nearly all participants reported that direct access to physical textual materials was important to their studies. Even in the case of Matthew Schneider, a student of textual scholarship who focuses almost exclusively on born-digital material, he reported that access to the physical containers of data, such as discs and cartridges, was extremely important to his analysis.

This nearly unanimous response concurs with Phyllis Franklin’s findings when she polled a group of members of the Modern Language Association as to whether they would prefer to consult virtual or print materials. Ninety-five percent of her respondents responded to the effect
that while digital holdings are useful, they cannot satisfy all of their scholarly needs. In other words, for these scholars it is not a matter of preferring one medium to the other, but of valuing both. Franklin discovered, after soliciting responses from a much larger group of MLA members, that 57% of the 169 respondents felt that physical primary materials are important to the work of bibliographers, textual editors, and literary critics, while 39% indicated their importance for audience/readership studies, and 28% spoke up for their value to book historians.

Participants often associated the value of direct access to material artifacts with the physical context in which they are housed. Four participants indicated that working in a setting that brings them into contact with physical artifacts allows a productive form of unstructured investigation, which included physically browsing special collections and/or simply absorbing the details of an artifact without a particular agenda. In this case, exposure to the artifact was necessary to develop the familiarity needed to begin noticing patterns of evidence. Two respondents referred to benefitting from the expertise of library staff, while one indicated the benefits of stack access given to her/him by librarians. These observations resonate with Unsworth’s observation that in performing the scholarly primitive of “discovery,” lack of structured investigation can increase the potential for unpredicted, novel, and interesting discoveries. On the other hand, electronic access to information (including digital facsimiles) is usually highly structured, and is designed to be navigated by searching for particular data rather than browsing. Like many who advocate the concept that enabling researchers’ access to physical holdings will set the scene for serendipitous browsing (whereby unexpected connections are made among materials that are physically nearby each other), Unsworth also points out that unstructured browsing that leads to unexpected discovery often is directly related


107 Ibid., 115-116.

108 Unsworth, “Scholarly Primitives.”

109 Ibid.
to the physical space of the archive or the library stacks. Furthermore, like those interview participants who cite the staff as one of the most valuable features of going on-site to study an artifact, Unsworth notes that “one of the most effective methods of discovery is still, and has always been, conversation with others who share our interests or who are simply interested in sharing.” The importance of these sorts of social processes was not revealed by the literature reviewed for this study, but was a clear trend in these interviews.

4.1.1.2 Digital Facsimiles: Limitations of the Medium

Many participants framed their rationales for desiring direct access to textual materials in the context of what they found unsatisfactory about replacing the experience of the physical artifact with a facsimile. Many participants expressed deep mistrust in the facsimile’s ability to faithfully represent the artifact, along with skepticism that any form of reassurance was possible to dispel this mistrust—short of a comparison of the facsimile to the original. As David Vander Meulen pointed out in interview, “One of the basic points of facsimiles, you can use them provisionally, but one never knows what one has.” In other words, one cannot work from facsimiles alone, because one must verify the fidelity of the facsimile to the original.

Another common concern surrounding facsimiles is in their inability to deliver necessary medi-richness to perform responsible textual analysis. Textual information such as watermarks, chain lines, paper thickness, and binding were cited as crucial textual evidence that facsimiles are at best unlikely, and at worst incapable, of representing. Participants also noted difficulties in determining scale, and found that even when a measurement device was included in the facsimile it was not a useful tool for envisioning scale. Determining format, particularly in the absence of visible chain lines, and measuring characters for the identification of type, are activities that require knowledge of the artifact’s size, and therefore are confounded without

110 Ibid.; for a recent discussion of serendipitous browsing, see Bruce E. Massis, “‘Serendipitous’ Browsing Versus Library Space.”

111 Unsworth, “Scholarly Primitives.”
Even if a ruler is included in an image, the act of representing a three-dimensional object in a two-dimensional format causes distortions that make the sort of accurate and consistent measurement required of most bibliographic exercises impossible.

One participant, David Greetham, made the point that direct consultation of textual artifacts is a prerequisite in creating a reputable scholarly edition. He noted the criticism Hans Walter Gabler experienced due to his reliance on facsimiles for parts of his copy-text for *Ulysses*. In 1988, John Kidd called Gabler to task for consulting facsimiles instead of primary materials in his critical editing of Joyce’s work. In this critic’s view, Gabler could not have distinguished between the markings of Joyce and those of his helpers or printers through reference to facsimile alone. Such allegations demonstrate the low opinion within the textual scholarship community of any critical editing that relies solely on facsimiles.

For instance, earlier this year I was invited to help Daniel White at the University of Toronto’s Book History Print Culture program determine the format of a particular book by looking at a digital facsimile version of it on Google Books. Without any data on size or indicator of scale, the format appeared to be a simple quarto, since its signatures indicated gatherings of four leaves. Only with catalogue data on the book’s dimensions could it be determined that the book was very likely actually an octavo-in-fours, created using half-sheet imposition. I thank Daniel for sharing his interesting research with me.


Hans Walter Gabler, “The Scandal of ‘Ulysses’: An Exchange,” *The New York Review of Books*, August 18, 1988, np. A response from John Kidd caps off this exchange, in which he employs a humourous conceit as a means of both commenting on Gabler’s and others’ responses to his original argument, and stating his bibliographic case in a way that is simultaneously more thorough, and more explicit regarding the importance of material investigation. The opening volley of Kidd’s response:

In the year 2088, the general editor of *Selected Papers of the Joyce Wars* has her hands full. [S]he finds the documents (paper, electronic, and plasmic) surviving from 1988 incomplete, contradictory, error-prone, stylistically archaic and a touch comic. Letters in one issue of *The New York Review* keep Professor Futura and her textological lexica busy for weeks.
These major areas of concern—doubt in the reliability of the digital representation, dissatisfaction with the level of detail captured in a representation, inability to determine aspects of the relationship between original and facsimile such as scale, and a general discontent with the sole use of facsimiles for performing critical editing—all have their root in the limited representational capacity and reliability of the digital image as medium.

The trepidation caused by the various ways in which the digital facsimile cannot provide the type and level of detail required to truly “know what one has,” as Vander Meulen put it, is symptomatic of the fact that as consumers of a steady stream of images, we have long grown out of the salad days of photography, in which Edgar Allan Poe could observe the details in a daguerreotype in 1840 and declare that they “are those of truth itself in the supremeness of its perfection.”\(^{115}\)

Digital photography today is treated with a great deal more skepticism. As Anne-Marie Willis, W.J.T. Mitchell, and Melissa Terras each point out, the ease with which adjustments and alterations are made to digital images contributes to an overall shift in the cultural acceptance of the trustworthiness of digital images. As Willis writes, “The ease with which … transformations can be made—using menus and a mouse, as on any personal computer—mean that digitized imagery allows a much greater potential for manipulation than traditional photography.”\(^{116}\) Mitchell points out, “the essential characteristic of digital information is that it can be manipulated easily and very rapidly by computer … Digital images are, in fact, much more susceptible to alteration than photographs, drawings, paintings, or any other kinds of images.”\(^{117}\)

Digital images are also altered unintentionally, through the process of storing, accessing, and displaying them: the algorithms built into each individual image-viewing program, which may

\(^{115}\) Quoted in Mitchell, *The Reconfigured Eye*, 5.


\(^{117}\) Mitchell, *The Reconfigured Eye*, 7; emphasis is Mitchell’s.
variously interpret the instructions that make up the image, or, as Lev Manovich points out, the ways in which compression leads to a newly-interpreted image each time it is copied, or even the variability of individual monitors, all affect the manifestation of the image in ways that make it less than completely reliable as evidence.\textsuperscript{118} Terras worries that there is a cumulative effect of manipulating digital images: “We have already seen that images are powerful, and that maintaining image-accuracy is crucial: manipulated images, whether intentional or accidental, change our understanding of reality and influence our cultural perception.”\textsuperscript{119}

There are also practical limitations to the amount of information that even the most unremediated digital image can provide. The level of detail in a digital facsimile is limited to its resolution, and therefore without recourse to the original, there is a limit to investigation. So far from being characteristic of Poe’s hyperbolic statement, that “the closest scrutiny of the photogenic drawing discloses only a more absolute truth,” the digital image, too carefully scrutinized, pixelates, and discloses only the truth that the image is a composite and an illusion.\textsuperscript{120} As Gombrich points out, “The standard of truth [is] … related to the medium. The image cannot give us more information than the medium can carry.”\textsuperscript{121} Furthermore, as the medium in question is, effectively, two-dimensional, qualities that can be determined through three-dimensional observation, such as the impression caused by chain lines, bearing type, hammering, and dog-earing, are difficult if not impossible to interpret in a two-dimensional image.\textsuperscript{122}


\textsuperscript{120} Quoted in Mitchell, \textit{The Reconfigured Eye}, 5.

\textsuperscript{121} Gombrich, “The Standard of Truth,” 192.

\textsuperscript{122} For demonstrations of the bibliographical significance of the indentations left by bearing type (uninked type used for platen stability in a forme), fold-marks left by systematic dog-earing performed during binding, and the marks in the paper left by the binder’s tools, see random cloud, “Where Angels Fear to read,” in \textit{Ma(r)king the...}
While the various representational limitations, as well as the inherent unreliability, of digital images is well-documented by theorists such as those referred to here, the responses as to how these shortcomings directly affect scholars’ ability to use them in their research puts these failings into a practical context, allowing me to query individuals regarding the depth of investigation at which point consultation of facsimiles becomes inadequate.

4.1.1.3 Digital Facsimiles: Responsibilities of the Medium

Participants raised additional objections to facsimiles that reinforce the need for access to the textual artifacts themselves, and they pertain to the choices that the creators and providers of digital facsimiles themselves make. These objections demonstrate the difficulties and dissatisfaction caused when the objectives of the creators and producers of a service are fundamentally different from those of its users.

Participants commonly complained about the lack of comprehensive digitization of the artifacts they were interested in. Participants cited the omission of blank pages (pages that did not include printed material but nonetheless are relevant to bibliographical analysis), book jackets, ephemeral material associated with the text, and pages that held no printed text, but contained manuscript notes. This problem is well documented in articles such as Kichuk’s “Metamorphosis: Remediation in Early English Books Online,” in which she discusses the practice of “content amputation” on the part of the creators of EEBO (discussed in chapter 2). As she points out, the original Early English Books microfilming project’s practice of omitting bindings, endpapers, special front matter, consecutive blank pages, and cropping margins, “results in the loss of valuable details about book production, illustration practices, and reading practices gleaned from page notes and inscriptions.”123 Unlike Greg, who in the early days of modern bibliography set aside the study of ephemeral materials such as bookplates as falling

---

123 Diana Kichuk, “Metamorphosis,” 298.
outside the concern of bibliography, textual scholars today, as revealed in the interviews, place great importance on paratextual materials.

Several participants expressed disappointment in the overall quality of some digital facsimile resources, such as EEBO and Eighteenth Century Books Online (ECCO). Five participants out of the total twelve expressed strong displeasure with quality of the images themselves, of the associated searchable text, and of the metadata, expressing reactions that ran the scale from mistrust to open hostility. While I had previously recognized that services such as EEBO fail to satisfy the standards for quality described by Grycz and Terras (discussed in section 2.3.3.), responses from interview participants revealed that the issue of quality has a much deeper impact on textual scholars than I had predicted.

These two areas of concern, lack of coverage and lack of quality, relate directly to the standards of the institutions creating the facsimiles. What the producers of the facsimiles consider to be the essential aspects of the texts to be captured appear to rarely match those of the textual scholar. The reasons for this frequently reside in the fact that the producers and providers of digital facsimiles tend to develop standards and practices that privilege the exigencies of efficient and cost-effective production over the needs of investigators of material texts. The closely cropped margins of some texts in the pre-digital Early English Books microfilm collection, as Kichuk observes, were often the result of attempting to fit the image to the 35 mm constraints of the microfilm, or to correct the skew of a hastily performed image capture. These problems were inherited by EEBO when the microfilm was digitized. While Internet Archive produces far superior digital facsimiles to those found in EEBO, the semi-automated image-editing software, combined with quotas for workers that encourage speedy editing, nonetheless similarly promotes the cropping out of margins from the images, thereby discarding material of interest to many textual scholars. The worker, operating in concert with the software, is able to edit images at the highest rate of speed if she or he sets the program to crop images mid-way between the edges of the text block and the fore-edges. Therefore, the tendency to willfully

124 Ibid.
discard valuable portions of the image, in this case half of the margin on each side of the text block on each page, is endemic to the Internet Archive’s workflow.

These two examples epitomize the influence of commercial/production models taking precedence over the needs of particular groups of users. The prevailing rationale for creating digital facsimiles is that image-based digitization is a cost-effective way to deliver verbal content, and what value the medium has for the scholar interested in the materiality of texts is largely coincidental. Susan Leigh Star’s well-known axiom is applicable here, that infrastructure tends to remain invisible until it breaks down, or in other words, that a system’s failings are typically apparent only to those whose perspective it does not privilege.\textsuperscript{125} While to textual scholars clipped margins, absent binding images, and missing pages may be obvious violations of the notion of the facsimile as a representation of the artifact, these violations are unlikely to concern the content-minded producers and providers of most digital facsimiles.\textsuperscript{126}

### 4.1.1.4 Digital Facsimiles: Contributions to Textual Exploration

Among interview subjects, the most popular positive observation about digital facsimiles was that having them available provided greater and more convenient access to verbal (“written”) content. Access to texts otherwise unavailable due to the scholar’s geographical location, or the limits of the scholar’s home institution’s collections, has been a major selling point of services such as EEBO, and constitutes one of the major rationales for creating digital facsimile repositories of rare texts.\textsuperscript{127}

Indeed, working with a digital facsimile was sometimes preferred, due to the added functionality of the electronic proxies. Participants specifically mentioned that they enjoyed working with

\textsuperscript{125} Susan Leigh Star, “The Ethnography of Infrastructure,” \textit{American Behavioural Psychologist} 43.3 (1999), 380.

\textsuperscript{126} A case in point is from my time working at Internet Archive Canada as a digitization technician. My fellow Internet Archive co-workers did not understand my motivation for retaining as much of the margin as possible during editing, and considered it an irrational waste of time.

\textsuperscript{127} Magedanz, “Early English Books Online (EEBO) [Review],” 2084; Grycz, “Digitizing Rare Books and Manuscripts,” 34.
digital facsimiles because of the ability to magnify closely, and to search full text when the facsimile is accompanied by a transcription. Some participants also noted the beneficial qualities of specialty imaging, such as beta-ray photography. Digital facsimile producers, such as Internet Archive, Google Books, and EEBO through its relationship with the Text Creation Partnership (TCP), have recognized and responded to the interest in having fully searchable text, and each has invested, to varying degrees, in creating some form of character-based (and therefore computer-readable) transcription of the digital facsimiles that they provide. Each of these producers also provides the ability to magnify their images for closer scrutiny. Just as the accuracy of the transcribed text determines how reliable the search results will be, resolution of images determines how much it can be magnified before it pixelates.128

Slightly under half of the participants volunteered that access to digital facsimiles of artifacts allow them to perform collation of multiple copies. Andrew Murphy pointed out the ability to use digital tools to make comparisons of textual variants, David Vander Meulen indicated the value of digital facsimiles as potential witnesses for collation, and Markus Stock described how digital facsimiles, which he both keeps on his hard drive and accesses online, allow him to perform collations on site in rare book rooms and archives. In this respect, Stock noted in particular that digital facsimiles allow him to bring together, albeit virtually, multiple witnesses to a text that would never normally be collocated, through arranging these facsimiles in multiple windows on his computer. Elisa Tersigni similarly noted the benefit of being able to collate digital facsimiles on screen against printed texts on site, and Matt Schneider specifically referenced his extensive use of imaging software for mechanically collating digital textual artifacts.

Finally, several participants reported the positive impact of using digital facsimiles in the classroom. Andrew Murphy, David Greetham, and William Sherman all reported that they use digital facsimiles in the classroom, in particular for the ability to examine details closely using the zoom function (Greetham) and because of the interactivity of many digital facsimile edition projects (Murphy). Indeed, many institutions are taking advantage of the use of high-quality

128 Interestingly, while Internet Archive has the lowest quality of character-based transcriptions, it provides the highest-resolution and best quality images.
digital facsimiles in online teaching platforms, such as the National Center for the Humanities’ “toolbox library,” entitled America in Class, which combines both text transcriptions and digital images of primary documents with structured lesson plans to teach subjects in American history and the arts.\(^\text{129}\)

Responses such as these demonstrate that, for all the dissatisfaction that participants expressed regarding digital facsimiles, there are particular activities in which digital facsimiles are valued for what they offer that the physical text itself cannot. The ability to scrutinize the artifact more closely than is possible with the naked eye, search entire texts for particular words or phrases instantly, the power to compare texts that are separated by great distances, and the chance to bring artifacts into the classroom virtually are all benefits made possible only through digital facsimiles. The implication of this finding is that a digital-facsimile environment designed to cater to textual scholars should be designed to take advantage of these strengths, rather than try to replicate the experience of direct examination of the textual artifact.

### 4.1.2 Methods of Textual Investigation

As is discussed in chapter 1 and previously in this chapter, one of the hypotheses of this study is that textual scholars have a shared set of characteristic activities, or scholarly primitives, to use Unsworth’s term, that they perform in their investigations.\(^\text{130}\) A review of recent article publications (described in section 3.2.1) revealed activities of *comparison, material analysis, pattern recognition, and the creative representation of observed patterns*, often through visualization and modelling. Accordingly, questions pertaining to these four activities were put to the participants in order to test these hypotheses.

#### 4.1.2.1 Collation and Material Analysis

The majority of participants performed at least one of these two activities as part of their investigation process. Of the participants that confirmed they did perform either comparison or

\(^{129}\) http://americainclass.org/sources/

\(^{130}\) Unsworth, “Scholarly Primitives.”
material analysis, or both, as part of their studies, there was a surprisingly even split between those who perform one or the other. Three participants reported engaging either mostly, or exclusively, in comparison as opposed to material analysis, and three participants claimed the opposite, while two reported engaging in both practices. Of those in the comparison group, two were more senior scholars, and one was a student, while in the material analysis group, two were students and one was a more senior scholar. This distribution may be related to the fact that the established textual scholars from the comparison group had performed a significant amount of critical editing throughout their careers, while no participant in the material analysis group had yet engaged in that form of textual scholarship. Participants in the material analysis group tended to be more interested in analyzing parts of texts not explicitly verbal: illustration, typographical braces, and moveable book parts such as volvelles.

Two participants observed that considering comparison and material analysis to be discrete activities might not be a helpful distinction. David Vander Meulen, for instance, pointed out that observations relating to a text’s structure need to be verified by comparing multiple copies, while differences observed when collating multiple witnesses are often profitably explained by investigating the way the texts are constructed.

The implication of these findings is that while both comparison and material analysis may not be considered scholarly primitives in the sense that they are activities inherent to the methodology of all textual scholars, the fact that nearly every participant engaged in one or the other activity indicates that both are important behaviours in performing certain kinds of textual investigation. A digital environment suited to all forms of textual scholarship should therefore facilitate both of these activities, and provide the scholar with the ability to privilege one or the other. Vander Meulen’s point that these activities are often closely related, and even interdependent, is also well taken, and therefore the ability to perform both activities within a single interface should be a feature of the digital-facsimile environment.

4.1.2.2 Pattern Recognition and Representation

In its broadest sense, pattern recognition is an activity that is inherent to all forms of analysis, not just textual scholarship. However, pattern recognition is an important activity to focus on in the present study, insofar as it is also a distinguishing characteristic of modern textual
scholarship within the history of the discipline. As was revealed in the literature review for this study (chapter 2), modern textual scholarship is defined in part by the shift to analyzing patterns present in the physical artifact in order to understand the process of transmission and establish the origins of textual variation, rather than relying on assumptions about the history of printing and using aesthetic judgment to select among textual readings. Within the sample articles from the review of recent publications in textual scholarship (described in section 3.2.1), textual scholars demonstrate a variety of tactics for representing these patterns, which tend to be expressed quantitatively. These tactics range in complexity from simple ratios to full-scale interactive modelling. To obtain more in-depth information about the importance of pattern recognition and representation in textual scholarship, participants were asked to reflect on the role of pattern recognition in their work, and the ways in which they represent their findings.

The majority of participants claimed the identification and analysis of patterns to be part of their work as textual scholars. Most described their investigation as an attempt to establish patterns and base their arguments on these patterns. Two participants in this group also described the importance of grouping evidence and of establishing correlations and trends as a means of understanding patterns. Others were more interested in identifying points at which patterns are disrupted.

Clifford Siskin, in his article for *Textual Cultures* entitled “Textual Culture in the History of the Real,” claims that in order to distinguish textual criticism from other forms of textual analysis such as literary criticism, textual critics must adopt alternatives to older tools of textual analysis, such close reading, by which “we have moved in closer to the text in order to make more of it.”

As an alternative, Siskin advocates pattern recognition, and cites Marshall McLuhan, who pointed out in 1994 that in conditions of information richness, the potential for pattern recognition is greatest. In the context of most forms of textual investigation, when a scholar is dealing not only with the literary content of a text, but also with the bibliographic information


132 Ibid., 128.
embedded in an artifact (not to mention the historical context of its transmission), the opportunity to profit from analysis based on pattern recognition is great indeed.

When asked to speak to the ways in which the participants represented their findings, and whether quantitative representation or modelling is relevant to their work, again the majority of participants agreed. The participants’ definitions of modelling varied, and they described activities ranging from the more or less straightforward graphical representation of quantifiable data, such as graphs and charts, to predictive engines containing large data sets, such as the application of corpus linguistics to reveal the semantic structure of an undeciphered ancient document. While most of the approaches described are in essence forms of visualization, and while the more complex modelling appears to perform a more explicitly predictive function, what is interesting is the extent to which both approaches were used as part of the analytical process, not simply to represent findings to others. Of the eight participants who use some form of visualization/modelling, six described this activity as having a role in the analytical phase of their work. For these participants, visualization/modelling is an ongoing activity that is used to work through ideas, test hypotheses, and reveal relationships in the data.

4.1.3 Technological Considerations

4.1.3.1 Platform

Participants were asked about the types of computing devices they currently use or would be willing to use in a rare book room or archive. The majority responded that they either currently use or would be willing to use laptops, in particular, in these settings. No one particular operating system was favoured among the participants. The most-commonly used operating system was Macintosh OS, however Windows operating systems were also used, and one scholar reported using the Ubuntu OS. Therefore, there appears to be no one favourable platform for which to design digital resources, and consequently a non-platform-specific approach is preferable.

4.1.3.2 Digital Imaging

Throughout the interviews, respondents repeatedly made reference to their use of both personal and institutional imaging devices. Fully half of the more senior scholars listed digital cameras as
part of their technological equipment. Within the student group, three participants owned
digital cameras, and three regularly employed a digital scanner owned by the institution at
which they performed their investigations.

Of note here is the oft-repeated observation that changes in institutional policy regarding the use
of digital cameras in the rare-book room and archive has changed the respondent’s approach to
the research that he/she does. Half of the participants that used imaging devices in the rare-book
library or archive noted that the institutions they visit have recently begun allowing photography
or other imaging to take place, which has significantly changed their research methods. Three
conclusions should be drawn from this finding. First, there is indication of a widespread shift in
policy at rare book and archive facilities relaxing restrictions on digital imaging, and even
facilitating digitization activities. Second, incorporating the products of this imaging activity
into the scholar’s workspace will be an important feature of a digital toolkit designed for these
users. The third and final conclusion I want to highlight is that in designing digital tools for any
user group it is important not only to take into account technological considerations, but also to
investigate the social/cultural conditions in which the digital tool is to be used, such as, in this
case, rare-book room etiquette. One of the benefits of interviewing textual scholars for this study
is that it provided me with multiple perspectives on these conditions, and provided me with an
understanding of conditions as textual scholars currently find them.

133 This proportion may in fact be higher. I was unable to ask two of the members of this group for their full
 technological equipment list due to time constraints.
4.2 Adjusting the Thresholds of Engagement: A Proposed Digital-Facsimile Environment for Facilitating Textual Scholarship

[I]n designing tools we are designing ways of being.

- Winograd and Flores, 1987.134

Willard McCarty notes in his 2008 article “What’s Going On?” that there has been a growing emphasis, since the creation of the Web, on mass digitization, as the millions of titles boasted by entities such as Google Books and Internet Archive attest.135 However, as McCarty points out, along with these digitized materials also comes a range of methods to work with them that are inherited from the past, and which need to be rethought. Interfaces such as those of EEBO, the Internet Archive, or the British Library’s Turning the Pages software reproduce the two-page spread, or opening—life-like in the sense that it mimics the experience of looking at a codex, but practically devoid of any attempt to benefit from the flexibility afforded by the digital medium. Conceived of exclusively as facsimiles, these materials perhaps come closer to their goal by recreating the represented material’s organizational structure as closely as possible. However, the digital medium allows for new forms that go far beyond those inherited from inscription technologies of the past. As McCarty points out, “at this historical moment the inventing function is or should be prominent, since it is quite clear that the imitation of the codex is for us a mug’s game.”136 In respect of the cache of digitized materials themselves,


inventing new ways by which they are accessed, organized, and manipulated, is precisely the work of this section.

In this section, I outline a proposal for a digital environment that is predicated on facilitating textual scholarship, and, in particular, the investigation of textual artifacts. This digital environment puts the lessons learned from the interview process, as well as the analysis in section 2.3.1 of other digital facsimile interfaces, into practice. The primary goal here is to demonstrate how innovative interface design and added functionality can recast digital facsimiles as tools, both for performing transformations not possible with the physical artifacts themselves, and also for modelling hypotheses about the materials under investigation. Such an environment challenges prior conceptions of digital facsimiles as discrete from the physical artifacts they represent, either as replacements or surrogates for the originals, and alters the thresholds of engagement between the original and the representation, between the physical space of the library and the virtual space of the digital environment, and between the textual artifact as physically fixed in its material form, and the digital proxy as mutable within its virtual environment.

In his chapter on modelling in *Humanities Computing*, McCarty builds on Clifford Geertz’s distinction between *models of* and *models for*. A *model of* is a representation of a pre-existing set of relationships that facilitates examination and experimentation, while a *model for* is a design for bringing something new into existence.\(^{137}\) The digital-facsimile environment that I propose in the present section is primarily a *model for* organizing and interacting with digital facsimiles in such a way as to privilege in particular the activities and needs of textual scholars, and enable new methods of investigation. However, this environment will itself contain several *models of* various relationships of the text, both to other texts and to itself.

One conclusion resulting from the analysis of the interview sessions that should be emphasized at the outset is the indispensability of direct physical access to textual artifacts. As responses have demonstrated, not only is direct access to the physical artifact crucial, but digital facsimiles are also felt to be inherently incapable, by virtue of their medium, to replace this access.

Furthermore, both the space and the resources of the facility itself that houses the artifact were noted as important, if not indispensable, for research into the material nature of texts. Therefore, since direct, on-site access to textual artifacts is indispensible, one measure of this resource’s suitability as an aid to textual scholarship will be whether and to what extent it supports investigation of artifacts on site.

Building on the strengths of resources such as the Internet Archive, which provides high-resolution full-colour facsimiles under a Creative Commons license, the digital tool I propose in this study operates by incorporating images from open repositories such as that of the Internet Archive, or images from users themselves. This approach takes into account the increasing interest expressed by interviewees in taking their own digital images of textual material and using them in their studies.

In order to create a digital-facsimile environment that presents significant advantages over other available environments, the principle of this proposed model is that it includes the ability to organize and manipulate the materials in such a way as to privilege the relationships of interest to the textual scholar. This environment would come equipped with several different templates for modelling the materials at hand, which facilitate the two major activities identified in the data-gathering portion of this study, namely, comparison and material analysis. As an aid to investigating artifacts on site, digital page-images that can be abstracted from typical codex page-order, combined in different sequences, cut up, annotated, and overlaid provide the opportunity to do virtually what cannot be done with the physical text itself. In this sense, rather than attempting to do the impossible—replicating the three-dimensional codex as a two-dimensional image—this model takes advantage of the mutability of virtual space by allowing transformations not possible to perform with the physical object itself. Thus, this environment establishes a complementary relationship between original and facsimile.

Edward Tufte, in Visual Explanations, provides a superb primer for organizing visual information in meaningful ways. “By extending the visual capacities of paper, video, and computer screen,” Tufte writes, “we are able to extend the depth of our knowledge and
experience.” If McCarty provides the rationale for why modelling is a productive scholarly activity, Tufte provides a guide for how particular relations in a model might be carried out visually.

One of the principal techniques that Tufte describes is that of parallelism. “Spatial parallelism,” he writes, “takes advantage of our notable capacity to compare and reason about multiple images that appear simultaneously within our eyespan. We are able to canvass, sort, identify, reconnoiter, select, contrast, review—ways of seeing all quickened and sharpened by the direct spatial adjacency of parallel elements.” Tufte’s analysis of visual rhetorical devices as essential acts of visual meaning-making is reminiscent of Unsworth’s concept of scholarly primitives, espoused throughout this thesis as a guiding rationale. Since comparing (one of Unsworth’s primitives) is one of the two primary activities this model proposes to facilitate, and material analysis, which is in essence reasoning about material relationships and phenomena within a text, is the other, spatial parallelism as Tufte describes it presents an ideal basis for the visual organization of information in this model. As the following illustration of the proposed digital-facsimile environment shows, spatial parallelism is deployed in this environment in various ways, both in the far left and far right frames. The following paragraphs will discuss how these areas are designed to operate.


139 Ibid., 80; my emphasis.
4.2.1 Navigational Viewer

The left-hand frame represents the navigational view of an entire text. This area contains thumbnail-versions of each digital-facsimile image and is used as a navigational aid for selecting particular pages to scrutinize in the central frame. There are three layouts for this navigational viewer, which represent three distinct models of the text: page-order view, which privileges the book as a container for textual content, and thereby presents the page-images in a conventional linear order (this organization is pictured to the extreme left of the illustration); imposition view, which highlights the structure of the text as a physical artifact by recasting the individual page-images in the orientation they were originally printed in (pictured in the left frame of the environment mock-up); and the witness view (not pictured), which casts multiple copies, or multiple editions, of the same text concurrently, to show the individual witnesses as steps in an ongoing process of textual transmission. The navigational viewer is the particular
focus of the prototyping experiment associated with this study, and will be discussed in much greater detail in chapter 5.

4.2.2 Workspace View

From the navigational view, users select a particular page for closer scrutiny, which appears in the workspace view in the central panel. This area is designed to provide users with the benefits of working with facsimiles that were noted by the interview participants. This view provides options for magnification, and offers tools for manipulation, including annotation (both textual and graphic), image-sampling, horizontal and vertical flipping, and rotation. The workspace view enables users to label their interventions, which are collected and organized in the collection view, described below.

4.2.3 Collection View

The collection view, depicted in the far right frame in the mock-up (figure 2), is designed to facilitate the behaviours of pattern recognition and visual/graphic representation that were identified in the data-gathering portion of this study as principal activities in the work of many textual scholars. Collection views are essentially folders containing the results of the interventions that scholars perform and label in the workspace view. This view enables users to isolate instances of a particular feature in a text or collection of texts, and to compare and contrast these instances. Based on Tufte’s assertion that colour is an effective technique for encoding information, one particular method of comparing and contrasting, and therefore describing patterns of phenomena in a collection, is to encode similarities and differences through the use of colour.140 In the hypothetical scenario being played out in the mock-up illustration, the researcher is tracking the repeated use of single type-settings in the headlines of the Dunciad (a technique often referred to as headline analysis), and has chosen to mark up one set of headlines that appear similar in red, and another in blue.

4.2.4 Platform

In the interview stage of this thesis, participants were asked to describe the computational equipment at their disposal, and it was revealed that participants used not only a wide range of devices, but also a variety of operating systems. This diversity presents a problem for the application designer, as most software runs on a specific operating system, and creating packages for multiple operating systems is both costly and time-consuming. Furthermore, since software applications are stored on a particular device’s hard-drive, a version of the application would need to be downloaded on each of a user’s devices if she or he wanted to use the digital-facsimile environment on multiple devices. Finally, the information architecture of mobile computing devices is significantly different from that of a laptop or desktop computer, and would require the design of an additional environment specifically for mobile use.

A viable alternative to the software application is to create an application using web languages, such as XHTML, CSS, and JavaScript, which can then be run using a web browser. A web-based application (or web app) differs from other documents and entities found online in that it is a small program or utility that runs in a browser and is designed to perform a finite range of tasks with a particular goal or purpose in mind.141 Web apps are not operating-system specific, and web apps accessed through browsers on mobile devices operate similarly, and in most cases identically, to those on larger computers. Furthermore, since the application is stored on a server and accessed through the web, a user may access it on any device that has an internet connection. Since the ability to access the digital-facsimile environment in the rare-book room or archive is an important requirement, a design approach should be selected that enables

141 This definition is adapted from John Barber, “Overview of Course, Tools and Methods,” a slide-set presented in the Digital Humanities Summer Institute 2012 course entitled “Creating Mobile Apps” (available online, http://dtc-wsuv.org/dgrigar/dhsi/day1-barber-overview.key); though Barber uses this definition specifically to describe mobile apps (which can be either native to the device or accessed through the web), it also applies to web apps generally. See also the ongoing discussion on the distinction between web-apps and websites on the blog for AppsLab, a think tank at Oracle Applications Strategy (http://theappslab.com/2010/11/12/website-vs-web-app/); here, Rich Manalang points out that websites are a collection of documents that may contain multimedia, whereas a web-app is an application (i.e., performs some type of function for the user).
researchers to use the tool on any portable device, and even on the facility in question’s own computers. Therefore, for the proposed digital-facsimile environment, a web-app design is highly preferable to a software-based approach.

4.3 Conclusion: Testing the Thresholds of Engagement

This chapter has discussed the results of this study’s interview process, including textual scholars’ positions on the importance of textual artifacts, the drawbacks and potentials of digital facsimiles, and tracks methodological themes for the sake of developing a profile of essential functions in a digital-facsimile environment to be used particularly by textual scholars in their on-site investigations. I have also proposed a model for accessing, organizing, and manipulating digital facsimiles in a way that is designed particularly to respond to the feedback acquired in the interview process, and builds on the strengths, while avoiding the pitfalls, of digital-facsimile environments.

However, a theoretical model that purports to offer a useful way of integrating digital facsimiles into on-site textual investigation lacks validity without proving the model’s viability through practical application. In the following chapter, I will report and discuss the outcome of the prototyping experiment described in section 3.3, which will serve as a partial test of the fitness of the digital-facsimile environment proposed in this chapter, reveal the ability to utilize digital facsimiles to model various perspectives on the nature of the text as a material artifact, and demonstrate the fitness of a light-weight, web-based approach for creating digital tools that are both effective and open to alteration and innovation by their users.
5 Testing the Thresholds of Engagement: Prototyping the Navigational Viewer

“So Angels read with ease. But the book remembers its other narratives.”

- random cloud

In “Where Angels Fear to read,” random cloud (a.k.a. Randall McLeod) contrasts the conventional approach to reading a text, that is, as prescribed by the narrative order of its verbal content, with an approach that examines the substance of the text as a physical object with the goal of revealing the narrative of its coming into being. The metaphor of the angelic and the earthly, or even demonic (or at least impish), readers works as a parodic comparison between traditional bibliography’s pursuit of the ideal text, and the reconsideration of the substance of textual transmission, the physical phenomena of textual artifacts, that scholars like McLeod propose as an alternative. Just as McLeod and others of his tribe used, and continue to use, many traditional methods of textual analysis to reveal and relate alternative narratives of the book, so too does this prototyping experiment attempt to show how the building blocks of more traditional forms of digital-facsimile display may be used to build new ways of expressing the alternative narratives of the book.

This prototyping exercise implements only a small part of the digital-facsimile prototype described previously. The goals of this prototyping experiment are threefold. The first goal is to demonstrate one of the central principles of the proposed digital-facsimile environment, namely, that although digital facsimiles cannot be used for deep material investigation of textual artifacts, they do have qualities, such as the ability to be freely organized and manipulated, that physical artifacts do not. Because one may transform a proxy within a digital environment in ways that are impossible with the physical artifact (or at least, not possible without destroying

digital facsimiles can be used to investigate textual artifacts and represent both synchronic and diachronic relationships within single texts and across multiple copies or even editions. The second goal is to demonstrate how a digital-facsimile interface can fulfill criteria for *both* of McCarty’s descriptions of modelling, discussed in chapter 4. The navigational views prototyped here act as both *models of* distinct perspectives on the nature of the material text, and *models for* performing different activities in the investigation of textual artifacts. Finally, this prototyping experiment seeks to show that the design and/or innovation of digital tools for textual scholarship can be undertaken by motivated textual scholars with limited expertise, resources, and time. This last feature is particularly important since, as will be discussed in further detail, there is a strong tradition within textual scholarship of researchers building and innovating the tools and techniques they use in their investigations, and one of the goals of this study is to put forth an approach for the design of digital tools for textual scholarship that are open and conducive to this sort of innovation.

Before we begin, I should like to reinforce that this prototype represents only one small part of what the digital-facsimile environment proposed above could be, and only in a very rudimentary state of development. The OED defines the term “prototype” as “a preliminary version made in small numbers for evaluation, or from which improved or modified versions may be developed.”¹⁴³ This is an apt description of the navigational-viewer prototype: the small number of functioning test cases included in it are intended to be evaluated by myself and the readers of this thesis, with the understanding that the ultimate goal of a prototype is to act as a preliminary testing ground, to be learned from and improved upon in the future.

¹⁴³ OED, “prototype,” Def. 3.
5.1 Leveraging the Potentials of Digital Facsimiles as a Dynamic Medium

Figure 3: McLeod's diagram of remote offsetting in *Gulliver's Travels* (Huntington, 16126)

The effectiveness of the navigational viewer prototype as an aid in investigating the physical artifact hinges on the fact that it takes advantage of the ability to manipulate the proxy in ways that one cannot with a physical artifact. Indeed, much of the drive towards modelling and visual representation within textual scholarship can be viewed as the attempt to represent relationships within the fixed medium of the textual artifact with models that use distinctive visual rhetoric to bridge the gaps of space and time. Two examples may clarify this point.

Looking at the various diagrams and models in the work of Randall McLeod (an example of which is shown to the left), one may easily recognize visual rhetoric designed to transcend the form of the text that is fixed in place by both the physical structure of the object itself and by pervasive custom of privileging the verbal content of the text and approaching it in a prescribed linear fashion. In the diagram shown here, McLeod visually
dismantles Jonathon Swift’s *Gulliver’s Travels* (housed at the Huntington Library, call #16126), in order to show the patterns of remote offsetting that occur in this copy.\(^{144}\) The one- or two-headed arrows that crisscross the diagram trace out the physical disposition of the sheets that make up the book before binding, in an intermediate state—or perhaps multiple intermediate states—of collation. In this way, diagrams such as McLeod’s demonstrate how modelling in textual scholarship frequently acts to expose relationships that are occluded by the physical structure of the textual artifact.

The second example demonstrates how modelling is used in textual scholarship to represent relationships in time. This diagram (shown below) is derived from David Greetham’s 1984 essay for *Studies in Bibliography*, entitled “Models for Textual Transmission of Translation: The Case of John Trevisa.”\(^{145}\) In this diagram, Greetham arranges textual variants found in different editions of the work of Trevisa, and demonstrates potential paths, through transmission and translation, that resulted in variant readings over time. Greetham implies a loose Cartesian organization, with time as a linear quality that runs down the page from earliest to latest.

\(^{144}\) The term *offsetting* refers to instances when a page bearing undried ink is pressed against another page. The result is that traces of the text of one page appear in reverse on the other. *Remote offsetting* describes when this transference occurs between two pages that do not face each other in a bound text. This type of evidence can reveal much about the organization of freshly printed pages in the printing workshop, including details of print order, and even whether the text as a whole was printed in conjunction with other, separate texts.

Since the diagram is concerned principally with textual priority, which is a relational rather than an absolute quality, neither a quantification of time passed (such as in the form of a legend), nor a distribution of variants to scale with the amount of time passed between them, is necessary. Variants, standing in on this diagram for the textual witnesses in which they appear, are arranged horizontally and are also relational rather than absolute. A vertical distance between them represents difference, but does not depict the extent to which they differ. The third factor in Greetham’s model is the connecting lines, which denote relationships among the various witnesses. Greetham’s model is very much in keeping with the tradition of stemmatic diagrams, which model processes of textual transmission over time through the spatial orientation of tokens representing textual artifacts, combined with some form of notation that describes the nature of the relationship.

The ways in which graphical representation and modelling are used in textual scholarship take advantage of the added capacity for representation, as opposed to the linearity of written description. Julian Warner, who points out that “[t]he surface offers a resource whose representational capacity exceeds that of the line,” argues that the next step in this evolution is
to “[e]scap[e] the constraints of linearity and surface, through modern information technologies.”\textsuperscript{146} The navigational-viewer prototype is designed to leverage the power of computing not only as a “venue for representation,” as Kirschenbaum notes, but also as an inherently mutable medium.\textsuperscript{147} The navigational viewer is designed to model relationships—using the digital facsimiles as tokens—in ways that cannot be done with the physical artifacts themselves. As will be described presently, the navigational viewer is designed not only to offer the user images of pages in the sequence as bound in the text, but also to discard the fixed orientation of pages in the artifact itself to model the material relationships the text possessed in its pre-bound state, and to visually bring together facsimiles of various copies or editions of the text in a visual orientation that privileges their relationship to the process of the text’s transmission over time.

5.2 The Navigational Viewer

The navigational viewer, alluded to above in the discussion of the proposed digital-facsimile environment, is the focus of the prototyping activity. The navigational viewer provides the user with an overview of a text, in the form of thumbnail images of each page. What is innovative about this overview is that it offers the user three ways of organizing the thumbnail images, each of which privileges a particular way of thinking about and investigating texts: 1) as vehicles for content; 2) as bibliographic objects; and 3) as nodes in the ongoing process of textual transmission. Each viewer functions as a model of one of these three distinct ways of conceptualizing the text, and also as a model for undergoing particular forms of textual investigation.

Before I describe the navigational viewer in detail, a short primer on book-making and bibliographical terms is required. When books are made, multiple pages are printed on a sheet of paper. The pages are oriented on the sheet so that when the sheet is folded and bound, correct textual order (the order maintained by nearly all digital-facsimile environments) is achieved.


\textsuperscript{147} Kirschenbaum, “Introduction,” 3.
The manner in which the pages are arranged on each side of a sheet (or forme) is known as *imposition format*. When four pages are printed on each side of a sheet, the imposition format is called *quarto*. When the number of pages per side is eight, it is an *octavo* format. Traditionally, indicators called *signatures* were printed at the bottom of particular pages to help the binder fold and sew the sheets in correct order, and bibliographers use this same system of notation to refer to individual pages in a text. Typically, letters of the alphabet were used to designate the *quire* (a sheet or group of sheets forming a sewn unit), followed by a number that corresponded to the priority of the page within the quire. Bibliographers have added to this notation a letter at the end that indicates whether the page appears to the right or left in an opening. The present study uses the common “r” for recto (right of the opening) and “v” for verso (left of the opening). Other textual scholars may use “a” and “b” instead.

Now, back to the navigational viewer prototype. This prototype is meant to be played with; at this time I encourage readers to visit the homepage of the prototype, which, once properly archived, should be available at [http://hdl.handle.net/1807/33081](http://hdl.handle.net/1807/33081). Here is where readers can access the navigational viewer prototype, as well as read a general description of this thesis and find out more about the author.

After entering through the *Thresholds of Engagement* homepage and clicking on “Navigational Viewer Prototype” to the left, you will arrive at a page that offers two options, the Quarto Viewer, and the Octavo Viewer.
Figure 5: *Thresholds of Engagement*: Navigational Viewer Prototype page (http://hdl.handle.net/1807/33081).

What you will see first is a jumble of pages, and you may be tempted to try to tidy them up—follow that instinct. Once you draw your cursor over the pages, they will begin to resolve into a pattern. These are the links into the navigational viewer test cases. The newly organized groups of pages give you a preview of what to expect from each link, and also provide visual references for what the words “quarto” and “octavo,” written beneath them, refer to. The grid of four labeled portions illustrates a quarto imposition format, which, once clicked, will bring the user to an example of the navigational viewer at work on a copy of the 1729 quarto of *The Dunciad. With Notes Variorum* printed by Lawton Gilliver, housed at the Thomas Fisher Rare Book Library and digitized by Internet Archive Canada. The collection of pages to the right moves

to form an eight-faceted grid, which illustrates an octavo format and links to the navigational viewer’s representation of *A General History of all Voyages and Travels Throughout the Old and New World*, an octavo text printed in 1708 by the infamous book pirate Edmund Curll, housed at the University of California, and digitized by Internet Archive.149

Figure 6: Navigational Viewer—Default Page View

After clicking on either link, you will see a default view, a column of “page-dummies” with signatures on them that make up the framework of the navigational viewer. To the left is a detail from the Quarto Viewer. This default view demonstrates the relationship between signatures and narrative page-order. The first page in this idealized text is A1r, which is to say, the front of the first leaf in quire A. A clear record for this facsimile indicates that the physical original is housed in Thomas Fisher Rare Book Library, call number B-10 04886.

149 M. l’abbé de Bellegard (Jean Baptiste Morvan), *A general history of all voyages and travels throughout the Old and New world: from the first ages to this present time: illustrating both the ancient and modern geography: containing an accurate description of each country ... an account of [sic] all discoveries hitherto made in the most remote parts* (Richmond, CA: Internet Archive, 2006): http://archive.org/details/generalhistoryof00bellrich; the physical copy resides at one of the libraries at the University of California, however this group of libraries holds more than one copy, and I have not yet narrowed down which copy was used in creating this digitization.
and obvious pattern persists throughout. Clicking on “Load Images” in the top left corner replaces the page-dummies with thumbnail-sized versions of the digital facsimiles of the first five quires of the subject text. Notice that as the viewer switches from the idealized text represented through the iconic page-dummies to the specific pages of the subject text, the signatures also change to accurately label the actual pages pictured.

Now that the images are loaded, feel free to click on any of them to access a high-resolution version of the page you’ve selected.

Figure 7: Navigational Viewer—Page-Order View of The Dunciad (1729)

The basic page-order view that is now visible displays the thumbnails in the same order that one would encounter them if reading linearly (angelically, as random cloud might have it150), from page to page beginning at the title. Using this orientation, you should be able to easily navigate the digital facsimile in accordance with the direction of the narrative by travelling down the column of images. This view is a model of the artifact as a vehicle for content, in
which the coherence and readability of the written text is of principle concern. This model serves those textual scholars who, like many of the interview participants, sometimes access digital facsimiles simply to view their literary content, rather than to examine them as representations of bibliographic objects. If you are primarily interested in reading the book “from cover to cover,” which is to say page by page in sequence from front to back, conventional reading order, then this is the viewer for you. However, if you are interested in being introduced to one of the book’s “other narratives,” continue reading.

Click on the next button, “Switch to Imposition View,” and the neat column of thumbnails cascades across the screen and resolves into clusters of images. This second view reorganizes the thumbnails to present the user with a navigational view in imposition format.

![Image](imposition-view.png)

Figure 8: Navigational Viewer—Imposition View of The Dunciad (1729)

You may recognize the organization from the Navigational Viewer menu. The *Dunciad*, in quarto (pictured at left), groups into image clusters, each cluster containing two sets of four
The General History in octavo, on the other hand, has two sets of eight page-images per cluster. Of the sets of image groups, the group to the left represents the side of a sheet known as the outer forme, while the group to the right visualizes the inner forme. Because formes, rather than individual pages, are the constitutive units in the construction of a book, many textual scholars, especially those engaged in material analysis, think in terms of formes rather than pages. The imposition view enables these textual scholars to navigate the facsimile according to this rationale. This view models the text as a bibliographic object replete with iterative patterns that together make up the text’s physical structure. This view is particularly well suited to analytical bibliographers, and those interested in historical design principles. Anyone interested in peeking beneath the page to find out what alternative forms the book may have taken before it assumed its bound codex structure might find it helpful to start here. The animated transformation also has pedagogical value, in that it vividly illustrates the relationship between the page order of the text, and the imposition format needed to achieve this order. This aspect was designed with those interview participants in mind who cited the benefits of using digital facsimiles in their teaching.

A third navigational view, not built as part of this prototyping experiment, but which would operate on the same technical principles as those views that the prototype currently facilitates, utilizes the basic framework of the page-order view, but enables the user to view facsimiles of multiple textual witnesses in parallel. Textual scholars, such as Marcus Stock, David Vander Meulen, and Elisa Tersigni, who take advantage of the ability to collate non-collocated texts through digital facsimiles, could use this view not only to compare identical sites in multiple copies, but also to have the ability to model hypotheses about priority by arranging the copies in relation to each other in the navigational view, for instance from first-printed to last-printed. This view can also be used to examine changes in a text from edition to edition. This view presents a model of the text as process rather than product, and is particularly well suited to book historians and editors of critical editions.

5.3 Tool-Building and Textual Scholarship

In his assessment of the field of digital humanities and his call for direction of future developments, McCarty points out the importance of producing digital models that not only provide scholars with tools, but also enable the scholar’s involvement in the creation and
improvement of tools: “it makes less and less sense to be thinking in terms of ‘end-users’ and
to be creating knowledge-jukeboxes for them. It makes more and more sense to be designing for
‘end-makers’ and giving them a scholarly equivalent of Tinker Toys.” This observation may be
especially relevant for textual scholars, insofar as there appears to be a clear trend within this
field, of researchers making their own tools and techniques, and improving upon old ones.
While Charleton Hinman and his optical collation device, the Hinman Collator, may be one of
the most celebrated examples of technical innovation in textual scholarship, one may also turn
to the McLeod Portable Collator, a device that performs the same function, though requiring no
power and fitting into a small carrying case instead of a medium-sized room. The Lindstrand
Comparator and Hailey’s Comet (named after their inventors, Gordon Lindstrand and Carter
Hailey, respectively) are two other products of the inventiveness of textual scholars. Within the
digital realm, many examples of this characteristic exist, such as the design of a digital process
for measuring and recording chain lines for paper identification by David Gants, and the use of
digital images in the comparison of ornaments to determine print-order by Eric Rasmussen.151
Larger-scale innovations of course also exist, such as the extensive research and development
led by Kevin Kiernan into uses of non-visible spectrum imaging, and three-dimensional artifact
modelling, or the use of DNA analysis to glean missing information from medieval manuscripts
by Timothy Stinson in collaboration with C. Michael Stinson.152

The implication of McCarty’s assertion, combined with the observation of the importance of
tool-building and innovation within textual scholarship communities, is that a digital-facsimile
environment designed to facilitate the investigations of textual scholarship should employ
computing principles that are accessible, scalable, and flexible enough that, with a little work
and perseverance, other textual scholars could adjust and add to the environment to suit their


own objectives. This web-based prototype was created with the markup, styling, and scripting languages of XHTML, CSS, and JavaScript, each of which is a language with fairly straightforward syntax, but which can be elaborated on and combined with other languages in powerful ways. Furthermore, the web is replete with resources to help learn these languages. The relative openness and accessibility of an application built with these languages will make its source code more likely to be intelligible to an interested user and consequently result in greater opportunities for alteration and innovation than software applications written using more complex programming languages. In order to encourage users to become “end-makers” in McCarty’s sense, the web page associated with this prototype includes the source code for the prototype (available under a Creative Commons license), as well as links to teaching materials, such as those found at www.w3schools.com.

Without actually implementing the entire environment and running a series of long-term performance tests, it is difficult to determine whether a particular approach meets the criteria set out above, both for creating a digital environment to facilitate textual scholarship, and for relatively easy alteration and innovation. However, the prototyping portion of this study demonstrates, through measuring the success of implementation under conditions of constrained time, resources, and expertise, that the approach developed is within reach of the individual scholar with initially limited technical expertise. Similar to the way that digitization of texts democratizes access, designing digital tools that are open to innovation on the part of end-makers of varying technical expertise democratizes the design and creation of these tools.

Using the elementary principles of professional project management, the prototype was designed based on the identification of a core problem or area for improvement and an analysis of the problem using the data gathered during the publications review and interview sessions. Considerations of schedule (six weeks in total were allotted to the prototyping phase) and scope (a reasonable estimation of the size and complexity of the prototype that could be accomplished in the time allotted) were taken into account during the planning phase, and while cost in the sense usually considered in a professional project-management framework was not relevant to this case, the planning phase did consider a balance between the scope and time constraints of the prototype, and the time and effort needed to acquire the technical skills to implement it.
Two weeks were allotted to developing a functional knowledge of XHTML, CSS, and JavaScript. At the outset, my expertise extended only to a familiarity with XML and HTML, and a beginner’s understanding of CSS. Using the freely available tutorials at www.w3schools.com, within two weeks I acquired the technical expertise necessary to begin implementing the prototype.

Implementation of the prototype occurred over a three-week period, and included weekly progress reports with Professor Alan Galey, which also involved workshopping particularly difficult portions of code. At the conclusion of this period, the prototype had reached the following level of completion: a navigational viewer that displays page-images in both page-order and imposition views, and enables the user to access high-resolution versions of each image, for both the quarto and octavo subject texts.

As McCarty points out, “computational models, however finely perfected, are better understood as temporary states in a process of coming to know rather than fixed structures of knowledge.”153 In this sense, the activity of prototyping (which, particularly in this case, is an act of modelling as well as a means of testing the feasibility and functionality of a larger project) is an ongoing one, subject to further elaboration and innovation not only by its principle creator, but also, ideally, by others. Additions subsequent to the six-week prototyping phase include the webpage framework that houses the navigation viewers, and the animated introductory page, but could theoretically, given enough time and resources, include implementation of the entire digital-facsimile environment proposed in the previous chapter, or beyond. Source code and links to tutorials are included in this project’s website for the very reason of encouraging others to explore and build their own permutations. This prototyping experiment has shown that the ability to create a functional web-based tool for textual scholarship is well within the grasp of an individual with limited time and technical expertise, and therefore it is reasonable to be

optimistic that other textual scholars could adopt such an approach and make their own additions and innovations.

5.4 Limitations of the Prototyping Experiment

The prototype created for this study is intended as a proof-of-concept. While it demonstrates the basic function of the navigational viewer, it is a closed system in that it displays only a small sample of pre-selected images, and applies standard bibliographic formulae, which the texts from which the images are derived adhere to.

In order to achieve the much larger goal of enabling users to work with texts from third-party sources such as Internet Archive, or upload their own texts, a significant amount of additional work would be required, which regrettably falls outside of the scope of this thesis. Assuming that an application-programming interface (API) is available from Internet Archive or a similar source allowing the prototype to navigate its structure, much work with this API would be needed in order to automate the extraction of image libraries for use in the viewer. Scripts would also be needed to automatically resize the extracted images for proper display. Perhaps most critically, an algorithm would need to be written that allowed the program not only to apply standard bibliographic formulae to texts based on user definitions, but also to allow user-made corrections in places where standard formulae do not accurately describe the material reality of the text. Since relatively few texts are composed of a single, consistent formula devoid of idiosyncrasies, improvisations, or interventions of the part of either the printer or an agent emerging further on in the process of transmission, the user will require the ability to correct the computer’s systematic, but assuredly uninsightful, understanding of the structural nature of the text. Teaching the machine to read the virtual text unangelically represents one of the most challenging—and exciting—areas for further research.
6 Conclusion

The design of electronic documents is not “technology,” dictated by necessity like plumbing or aerodynamics. It is design, a system of conscious decisions about the way things should be—like architecture, music or game design. The design of our electronic documents has shaped today’s world. And so far it has been simple-minded, shallow, and darkly limiting.

- Tim Berners-Lee\textsuperscript{154}

Throughout its history, humanities computing has shown a healthy appetite for imagination and innovation while continuing to maintain high scholarly standards. Now that the Internet is such a dominant feature of everyday life, the opportunity exists for humanities computing to reach out much further than has hitherto been possible.

- Susan Hockey\textsuperscript{155}

For, you see, so many out-of-the-way things had happened lately, that Alice had begun to think that very few things indeed were really impossible.

- Lewis Carroll\textsuperscript{156}

As Susan Hockey points out, in her chapter for the Blackwell Companion to Digital Humanities, “In the early days of digital imaging there was much discussion about file formats, pixel depth, \ldots

\textsuperscript{154} Quoted in Transliterature, A Humanist Design (http://transliterature.org/).

\textsuperscript{155} Hockey, “The History of Humanities Computing.”

and other technical aspects of the imaging process and much less about what people can actually do with these images other than view them.”

Indeed, so far the relationship between digital imaging and textual scholarship has been largely limited to merely creating digital facsimiles of textual artifacts and making them available for viewing. The vast majority of digital facsimiles of texts have been delivered to us through an infrastructure that is limited by the relatively narrow conception of the text as a container for verbal content, a container which itself is of no particular consequence. The textual scholar, who is interested in the many material forms a text has taken over time, is not well served by this limited approach.

I introduced this thesis by explaining its title, *Thresholds of Engagement*. To conclude this thesis, I will return of the metaphor of the threshold. In *Alice’s Adventures in Wonderland*, Alice enters a world in which she feels “few things indeed were really impossible.” It is a feeling that many people share when they are introduced to the design possibilities of the digital medium, in which the ability to access, store, and manipulate information seems limited only by one’s imagination. However, Alice’s travels through Wonderland are checked when she encounters a miniscule door. Peeking through, she sees a beautiful garden awaiting her on the other side, but the way in is simply too small to let her through. If one considers how the interface of a digital facsimile repository operates as a portal to the material housed in the repository, it is easy to imagine how an interface limited in the ways it provides access to digital textual materials is like the miniscule door that Alice confronts in her Adventures in Wonderland.

Another author interested in doors is Bruno Latour, who provides a useful examination of technology design by demonstrating the social implications of the design of doors, and door-closers. He asserts that the door, as a technology, facilitates the organization of people and

materials, enabling environments for specialized activities. That same mode of ingress and egress also privileges certain uses, and certain users, by nature of its design.\textsuperscript{159}

As I discussed at length in chapter 2, digital-facsimile environments such as EEBO, Google Books, Internet Archive, or even the British Library’s Turning the Pages viewer, seek to facilitate the specialized activity of engaging with texts, and the interface of each of these resources prescribes a certain type of use. Each offers a navigational model that leads the user through the text in conventional reading order, presupposing the use of the resource as a means of accessing the textual content—the work in its allographic form—more so than as a means of investigating the material artifact—the autographic text (see chapter 1). The ability to magnify images admittedly goes a long way towards facilitating the use of these resources to investigate physical details of the artifacts represented, the degree to which this is the case being more or less directly related to the available resolution and quality of the images. However, these environments still dictate a passive investigation. This study, and in particular the discussion of the responses of those textual scholars who participated in in-depth interviews, indicates that textual scholars tend rather to engage actively with their materials, making comparisons among multiple texts, contrasting phenomena within individual texts, and using a wide range of techniques and resources to work through their hypotheses using modelling and representation. It seems that, compared to the needs of textual scholars for a door to the virtual text that provides richness, versatility, and active engagement, most digital-facsimile environments provide a narrow entrance indeed.

When Alice encounters an impassible threshold, she avails herself of a trick she has only just begun to believe was possible—the trick of changing her size—which she uses to fit through the tiny door. However, Alice discovers that changing herself to fit the proportions of the door would only end in tears. If the portals that digital facsimile providers offer to access and use their materials are too limited to support the needs of textual scholars interested in the material nature of their targets of study, thresholds that enable engagement with digital facsimiles must

\textsuperscript{159} Ibid., 301.
be designed with the needs of textual scholars at heart. We need bigger doors, not “drink-me” potions to shrink us down to size.


This thesis has sought to examine, in detail, what the methodological needs and interests of textual scholars are, and to deploy knowledge gained from past and current projects and tools relating to image-based textual artifact representation along with theories and practices in interface design and digital-image handling. What this thesis proposes in its description of a digital-facsimile environment particularly attuned to the needs of textual scholars is that a suitable threshold for engaging in digital representations of textual artifacts is one that not only
provides models of the ways that textual scholars think about books, but also models for engaging with artifacts that suit their particular methods.

The digital-facsimile environment I have proposed, and the prototype I have presented, provide models of thinking about texts from different perspectives within textual scholarship, as I have interpreted them from the responses from the twelve textual scholars I had the privilege of interviewing in depth. This environment starts by offering the conventional content-centric view of texts, by providing the digital facsimiles of a book in narrative/reading order. It then extends to provide a view of the book from the perspective of a bibliographer, who engages with the text as a physical object made up of its component parts, and who works to reveal stages in the process of the artifact’s creation. I also hypothesize an interface that enables textual scholars concerned with a diachronic understanding of the text to view multiple copies or editions in relation to each other, researchers such as book historians and textual editors who view each book as a node in an ongoing process of transmission. In this way the proposed digital-facsimile environment creates portals to the digital text that specifically facilitate the specialized activities of textual scholarship.

Simultaneously, the environment I propose in this thesis takes into account the types of common behaviours, or scholarly primitives, of textual scholars’ investigations of physical artifacts, and provides affordances for these activities. In addition to providing a facsimile viewer that models various artifact-centric views of the text (in other words, the type of engagement that de Grazia and Stallybrass advocate when they insist that early editions should be “looked at, not seen through”), the proposed environment includes the ability to sample, manipulate, compare, and organize. On the one hand, these affordances prescribe the sort of active engagement with digital facsimiles that those textual scholars interviewed described performing through other means in the course of their studies. On the other, the digital-facsimile environment itself executes a sort of work similar to what Latour calls translation. In Latour’s discussion of the miraculous invention known as the door, he describes the door’s ability to facilitate the metamorphosis of a wall into a hole, and back again, thereby replacing the need to destroy and rebuild a wall each time one wishes to cross a threshold. Latour’s example depicts a technology that introduces the ability to perform simple and reversible state changes into a scenario where the pre-existing physical conditions allow such changes only at great cost. Such is the case with
the features of the digital-facsimile environment I propose. Like Latour’s door, a digital-
facsimile environment that enables organization, reorganization, sampling, manipulation, and
annotation creates the ability to augment the examination of physical texts, which themselves
could only be reorganized, sampled, manipulated, and annotated at grave, irreversible cost—the
destruction of the artifact. Therefore, the digital-facsimile environment designed with such
activities in mind opens doors not just to the study of the facsimile, but enables active
engagement with the threshold of the physical text itself.

Finally, in order to surmount both the barriers of platform-specific computing devices, and to
enable more fluid traversal of the threshold of the library itself, the digital-facsimile
environment I have proposed, and tested the viability of through prototyping, is designed to be a
web-based application, which can run on nearly any computing device, of any size or brand, so
long as it is equipped with a modern web browser. As Susan Hockey has pointed out, scholars in
the humanities realized early on that the web is a powerful tool for research.\(^{160}\) The added
benefit of a web-based approach is that it is an environment that is conducive to tinkering.
Textual scholars tend, as I have discussed, to be inclined towards innovating their techniques
and tools, both through the invention of new devices or methods, or through the refinement or
repurposing of those of others. A web-based platform, as Hockey notes, allows for incremental
creation and ongoing emendation.\(^ {161}\) The accessibility and scalability of web languages such as
XHTML, CSS, and JavaScript (tutorials for which are freely accessible at sites such as
www.w3schools.com), combined with making the source code for the environment freely
available and easy to access, encourages the sort of innovative behaviour that this study has
shown to be such a common practice within textual scholarship.

The three quotes at the beginning of this conclusion hint at a trajectory for the design of digital
texts and the ways they are delivered to users. The quote by Tim Berners-Lee points out that
modes of digital textuality are determined by the design choices that its creators make, far more
so than the technology used, and that so far the potential for rich, imaginative design has not yet

\(^{160}\) Hockey, “The History of Humanities Computing.”

\(^{161}\) Ibid.
been exercised. Hockey’s quote highlights with optimism the benefits that the imagination and innovation of the humanities computing community can bring to the larger world, and notes that current technology enables more involvement from this community than has before been possible. What remains to be seen is how the innovative nature of that class of scholar most well-equipped to intelligently inform the digital existence of textual artifacts—textual scholars—will continue to influence the digital afterlives of material books, both as remnants of our literary heritage, and as precursors to the communication technology that shapes the digital landscape today. With this community continuing to extend the thresholds of engagement with textual artifacts through the digital medium, we will all have the opportunity to see, like Alice, that very few things indeed are really impossible.

*June 29th, 2012. Washington, DC.*
References

“About EEBO.” EEBO: Early English Books Online.
http://eebo.chadwyck.com.myaccess.library.utoronto.ca/about/about.htm.


de Bellegard, M. l’abbé (Jean Baptiste Morvan). *A general history of all voyages and travels throughout the Old and New world : from the first ages to this present time : illustrating both the ancient and modern geography : containing an accurate description of each country ... an accounof [sic] all discoveries hitherto made in the most remote parts.* Richmond, CA: Internet Archive, 2006:

http://archive.org/details/generalhistoryof00bellrich.

“Digital Athenium Project.” *Electronic Facsimiles & Texts.*

http://www.eppt.org/digitalatheneum/.


http://www.eppt.org/eppt/.


http://www.textcreationpartnership.org/tcp-eebo/.

“EFT HomePage.” *Electronic Facsimiles & Texts.*

http://www.eppt.org/.


“Goals of the Project.” The UVic Image Markup Tool Project.


http://journals1.scholarsportal.info.myaccess.library.utoronto.ca/details.xqy?uri=/00104817/v36i0001/3_eiihc.


Dear Dr. Galey and Ms. Rebecca Niles,

Re: Your research protocol entitled, "Thresholds of Engagement: The Integration of Image-Based Digital Resources into Textual Scholarship"

**ETHICS APPROVAL**

<table>
<thead>
<tr>
<th>Original Approval Date:</th>
<th>October 17, 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expiry Date:</td>
<td>October 16, 2012</td>
</tr>
<tr>
<td>Continuing Review Level:</td>
<td>1</td>
</tr>
</tbody>
</table>

We are writing to advise you that the Social Sciences and Humanities Research Ethics Board (REB) has granted approval to the above-named research protocol under the REB's delegated review process. Your protocol has been approved for a period of **one year** and ongoing research under this protocol must be renewed prior to the expiry date.

Any changes to the approved protocol or consent materials must be reviewed and approved through the amendment process prior to its implementation. Any adverse or unanticipated events in the research should be reported to the Office of Research Ethics as soon as possible.

Please ensure that you submit an Annual Renewal Form or a Study Completion Report 15 to 30 days prior to the expiry date of your current ethics approval. Note that annual renewals for studies cannot be accepted more than 30 days prior to the date of expiry.

If your research is funded by a third party, please contact the assigned Research Funding Officer in Research Services to ensure that your funds are released.

Best wishes for the successful completion of your research.

Yours sincerely,

Margaret Schneider, Ph.D.,
C.Psych
REB Chair

Dean Sharpe, Ph.D.
REB Manager
Dear Madam/Sir,

My name is Rebecca L. Niles. I am a Master of Information student at the Faculty of Information, University of Toronto. I am undertaking a Master's thesis on the intersection of textual scholarship and digital environments, and am currently seeking both practicing and aspiring textual scholars to participate as interview subjects in this research. This letter is to invite you to participate.

The premise of my study is that bibliographers, textual critics, and book historians, who perform investigations of documents as physical artifacts, have a long history of innovation in the tools and techniques they use in the rare book room, archive, and other sites. In recent history the field of digital humanities has developed, which also engages substantially in the development of new tools and techniques to perform scholarship in the traditional humanities. In particular, the use of digital facsimiles in humanities research has grown significantly with the advances in digital technology increasingly enabling high-quality, low-cost digitizations of textual artifacts as well as the increased ability to deliver these digitized artifacts to users. Unfortunately, in the past, physical textual artifacts and their digital facsimiles have been dichotomized, with the facsimile being figured as a substitution or replacement for the physical artifact. This study eschews that line of argument, and instead proposes to explore the ways in which image-based digital technology, along with mobile computing devices, can be used to aid and augment the on-site exploration of textual artifacts, placing the material and digital texts in a complementary, rather than adversarial, relationship.

This study seeks to develop a concrete profile or profiles of the methods and practices that both new and established textual scholars (including textual critics, bibliographers and book/print historians) use in the course of their investigations. The ultimate goal of this study is to discover potential areas of growth in textual scholars' use of digital tools in the field, and in particular what qualities would contribute to the successful design of digital tools that employ digital facsimiles of artifacts.

You are being invited to participate in this study because you are a textual scholar whose published work demonstrates that you engage in on-site examination of textual artifacts in the course of your studies. Your perspective on bibliographic methods and techniques would therefore be a great asset to this study.

The nature of your involvement in this study would be to act as an informant about the methods, techniques, and tools you currently employ in your investigations in the field. Participation would consist of engaging in a brief interview session, in person if possible, or otherwise via telephone or a voice over internet protocol (VoIP) service, such as Skype. You would be asked questions on the subject of the methodology you employ in your examination of textual artifacts. General questions about your work and working practices will be asked, as well as specific questions about the types of tools you employ in your investigations, and the ways in which you employ them. The interview is expected to last for approximately one hour, with the potential that you are asked to participate in a shorter, follow-up interview. Select participants may be invited to participate in short on-site observation sessions later in the research process, however, participation in the interview portion in no way necessitates involvement in any observation session.
I encourage you to consider participating in my research. While monetary compensation is not available for participants, you may be interested in having the opportunity to discuss your scholarly methods, techniques, and tools, while contributing to research that has the long-term potential to benefit your field of expertise. If you are interested in participating in this study, or would like more information, please contact me at rebecca.niles@utoronto.ca, or call me at 647-895-3767. Alternatively, you may also contact my Faculty Supervisor, Dr. Alan Galey, at alan.galey@utoronto.ca. Please note that in no way will you at any point be obligated to participate, and if you choose to participate you are completely free at any time during the research process to withdraw from the study.

Thank you very much for your time and attention.

Sincerely,

Rebecca L. Niles
Master of Information Candidate
Faculty of Information, University of Toronto

[If sent as email, text to include researcher’s header information as footer]
Dear Madam/Sir

My name is Rebecca L. Niles. I am a Master of Information student at the Faculty of Information, University of Toronto. I am currently undertaking a Master’s thesis on the intersection of textual scholarship and digital environments, and am currently seeking both practicing and aspiring textual scholars to participate as interview subjects in this research. This letter is to invite you to participate.

The premise of my study is that in the rare book room, archive, and other sites where bibliographers, textual critics and book historians perform investigations of documents as physical artifacts, there is a long history of innovation in the tools and techniques used. In recent history the field of digital humanities has developed, which also engages substantially in the development of new tools and techniques to perform scholarship in the traditional humanities. In particular, the use of digital facsimiles in humanities research has increased significantly with the advances in digital technology increasingly enabling high-quality, low-cost digitizations of textual artifacts as well as the increased ability to deliver these digitized artifacts to users. Unfortunately, in the past, physical textual artifacts and their digital facsimiles have been dichotomized, with the facsimile being figured as a substitution or replacement for the physical artifact. This study eschews that line of argument, and instead proposes to explore the ways in which image-based digital technology, along with mobile computing devices, can be used to aid and augment the on-site exploration of textual artifacts, placing the material and digital texts in a complementary, rather than adversarial, relationship.

This study seeks to develop a concrete profile or profiles of the methods and practices that both new and established textual scholars (including textual critics, bibliographers and book/print historians) use in the course of their investigations. The ultimate goal of this study is to discover potential areas of growth in textual scholars’ use of digital tools in the field, and in particular what qualities would contribute to the successful design of digital tools that employ digital facsimiles of artifacts.

As a student of textual scholarship, the methods, techniques, and tools that you use when examining textual artifacts on-site are of interest to this study.

The nature of your involvement in this study would be to act as an informant about the methods, techniques, and tools you currently employ in your investigations in the field. Participation would consist of engaging in a brief interview session, in person if possible, or otherwise via telephone or a voice over internet protocol (VoIP) service, such as Skype. You would be asked questions on the subject of the methodology you employ in your examination of textual artifacts. General questions about the work and working practices will be asked, as well as specific questions about the types of tools you employ in your investigations, and the ways in which you employ them. The interview is expected to last for approximately one hour, with the potential that you are asked to participate in a shorter, follow-up interview. Select participants may be invited to participate in short on-site observation sessions later in the research process, however, participation in the interview portion in no way necessitates involvement in any observation session.

I encourage you to consider participating in my research. While monetary compensation is not available for participants, you may be interested in having the opportunity to discuss your scholarly methods, techniques, and tools, while contributing to research that has the long-term potential to benefit your field of expertise.
you are interested in participating in this study, or would like more information, please contact me at rebecca.niles@utoronto.ca, or call me at 647-895-3767. Alternatively, you may also contact my Faculty Supervisor, Dr. Alan Galey, at alan.galey@utoronto.ca. Please note that in no way will you at any point be obligated to participate, regardless of whether you express interest in participating, and if you choose to participate you are completely free at any time during the research process to withdraw from the study.

Thank you very much for your time and attention.

Sincerely,

Rebecca L. Niles
Master of Information Candidate
Faculty of Information, University of Toronto

[if sent as email, text to include researcher’s header information as a footer]
Appendix D—Letter of Informed Consent

Thresholds of Engagement: Integrating Image-Based Digital Resources into Textual Scholarship
Informed Consent for Participation in Interview Session

You have been asked to participate in a study by Rebecca L. Niles, Master of Information student at the Faculty of Information, University of Toronto. The study, undertaken as part of thesis research, seeks to develop a concrete profile or profiles of the methods and practices that both new and established textual scholars (including textual critics, bibliographers and book/print historians) use in the course of their investigations. The ultimate goal of this study is to discover potential areas of growth in textual scholars’ use of digital tools in the field, and in particular what qualities would contribute to the successful design of digital tools that employ digital facsimiles of artifacts.

Eligibility: You are being invited to participate either because you have demonstrated yourself to be a practicing textual scholar by publishing work in this field or by being a member of a professional group pertaining to textual scholarship, or because you have demonstrated yourself to be an aspiring textual scholar by attending a course relating to textual scholarship. From this information it is reasonable to suppose that part of your research involves working directly with textual artifacts; if this is the case, then you are an appropriate participant in the present study. If your work does not involve direct investigation of textual artifacts then you will need to be excluded from this study.

Voluntary Participation: Participation in the present study is entirely voluntary; you are free to decline to participate, with absolutely no consequences to you. If you choose to participate in this study, you are free to subsequently withdraw at any time up to the point that the written report associated with this study has been submitted for evaluation, and all information pertaining to you and your responses during interviewing will be removed from the data set. If you agree to participate you retain the right to decline to answer any particular questions in the interview. You also have the right to decline to have your responses audio-recorded.

Procedure, Duration, Time Involvement: The present study consists of a literature review, a series of interviews and on-site observation sessions, and a prototyping phase, culminating in a written report in satisfaction of the requirements of a Master’s thesis. The duration of this study is from September 2011 to approximately June 2012. During this study, participants will be interviewed, ideally in person or else remotely (i.e., via telephone, Skype, etc.), on the subject of the methodology they employ in their examination of textual artifacts. General questions about the participants’ work and working practices will be asked, as well as specific questions about the types of tools they employ in their investigations, and the ways in which they employ them. If the participant consents to being audio-recorded, then the interview will be recorded and transcribed at a later time. The interview is expected to last for approximately one hour, with the potential that the participant is asked for a shorter, follow-up interview. Later in the course of the study, select participants may be invited to participate in short on-site observation sessions, however, participation in the interview portion in no way necessitates involvement in any observation session.

Foreseeable Harms and Benefits: Aside from the potential time inconvenience, there are no significant foreseeable risks, harms or inconveniences associated with the present study. Below, the protocol to be used to ensure security of participant information will be outlined, and the participant will be able to select the degree of confidentiality she/he would like to maintain. On the other hand, while no immediate direct benefits to the participant are expected to arise from involvement in the current study, the study’s primary goal is to develop a framework for the informed, intelligent development of new digital tools for textual scholarship, which is a benefit to textual scholarship generally, albeit an abstract one in the short term.

Compensation/Reimbursement: Unfortunately, the Faculty of Information at the University of Toronto does not allocate funding for Master’s thesis research, and therefore monetary compensation is unavailable for participants.

Claude T. Bisell Building, 140 St. George Street, Toronto, Ontario, M5S 3G6 Canada
www.ischool.utoronto.ca
Privacy/Confidentiality: All participant information, including personal information such as name and contact information, will be kept secure and confidential to the extent chosen by the participant. By default, all participant information will be kept confidential; participant information will available only to the researcher and her supervisor, Dr. Alan Galey of the Faculty of Information, University of Toronto, and the participant’s identity will not be revealed to any other party though any stage of the study, including the final written report. Any information used in the final report will be reported in an anonymous fashion. However, since the nature of the interview content is professional methodology, some participants may wish to be credited for the statements they make about their techniques, analogous to how such information would be credited if it derived from a publication. If you would prefer to be credited in relation to statements made in interview sessions, please sign the appropriate line below.

Use of Interview Data: The results of the participant interview sessions will be used in two main ways. Responses will be analyzed with the aim of identifying common activities or methods among participants and developing methodology profiles, which will be used a) to inform the tool prototyping phase and b) to serve as the basis of the final report. A summary of this study’s findings will be provided to all participants, and a copy of the final report may be requested by the participant.

Conflicts of Interest: No apparent, actual or potential conflicts of interest are expected to result from the present study. Participants are encouraged to raise any concerns about conflict of interest (or any other aspect of the study) with the researcher, her faculty supervisor, or the Office of Research Ethics (contact information below).

Potential participants are encouraged to contact the researcher, Rebecca L. Niles, by email (rebecca.niles@utoronto.ca) or by telephone (647-895-3767) to discuss or raise concerns about any aspect of their participation in this study, or to withdraw from this study, which they are free to do at any time during the research process. Participants may also contact the researcher’s Faculty Supervisor, Dr. Alan Galey, at alan.galey@utoronto.ca or by telephone at 416-946-5361. Participants can contact the Office of Research Ethics at ethics.review@utoronto.ca or by telephone at 416-946-3273, if they have questions about their rights as participants.

DECLARATION OF INFORMED CONSENT:

I, ___________________________________________________________________, have read and understand the information contained in this form, and have been afforded enough time and access to additional information to make an informed decision about my participation, and on this basis consent to participate in the present study:

________________________________________________________________________

While by default participation in this study is confidential and the results of the interview between myself and the researcher will be reported anonymously, I wish rather to waive this default confidentiality and be credited in the researcher’s written report for the information I provide: ___________________________________________________________________

I consent to having my interview session audio-recorded, with the understanding that recordings will be used by the researcher solely for later reference and transcription, and not shared with other individuals or groups: __________________________________________________________________
Appendix E—Interview Guide

1) General short discussion of the form of textual scholarship practiced by the interviewee
2) How long have you been practicing textual scholarship?
3) In what ways do you feel that the way you practice textual scholarship has changed in the period in which you have been practicing?
   a. Have new methods appeared?
   b. Have new tools appeared?
   c. How did new methods and tools (if any) become known to you, and how did you learn to use them?
   d. In what ways did new methods and tools affect your work?
4) How important is having direct access to the textual artifact you are studying? What kind of activities can only be performed in proximity to the artifact, and what kinds of activities can be performed elsewhere?
5) Do you ever use facsimiles, either digital or paper-based, in your studies? If so, then for what purposes?
6) To what extent do archives and records (such as contracts, correspondence, and bibliographies including ESTC) figure in your research? Is your research primarily based on information gleaned from these sources, initiated and supported by such information (i.e., is investigation of textual artefacts initiated by discoveries made in archives and records), only tangentially supported by such information (i.e., records and archives are only consulted for individual problems that arise from textual artifact investigation), or does this type of information not figure in your research at all?
7) My research so far indicates that there are two major categories of textual scholarship investigation, collation, or the comparison of multiple copies to each other, and structural analysis, examination of the relations among the various parts of a single text.
   a. To what extent do you feel that you engage in comparison among multiple copies? [provide examples if clarification is needed] Please describe the methods and tools that you use when making comparisons among documents (if you engage in this practice) and if possible, provide a few examples of when you have used these techniques and/or tools.
   b. To what extent do you feel that you engage in analyzing the relationships among the different parts of a single text? [provide examples if clarification is needed] Please describe the methods and tools that you use when analyzing the relationships among the different parts of a single text (if you engage in this practice) and if possible, provide a few examples of when you have used these techniques and/or tools.
   c. Laying aside the activities you have just described, what parts of your investigation have you not described so far? In other words, what activities do you regularly perform that you feel do not fall within these categories?
8) My review of current publications in the field indicate a distinct emphasis on pattern recognition within a set of information, whether this set be multiple texts, or particular phenomena within an individual text. This pattern recognition appears to require sampling, or an isolation of the occurrence that may display a pattern, and comparison among occurrences to establish a pattern and then judge whether additional occurrences adhere to, or depart from, the pattern. This process sometimes includes quantification of phenomena, mathematical calculations, and modelling of observed patterns.
   a. To what extent do you feel that you engage in sampling and comparison when investigating a text or set of texts? [provide examples if clarification is needed]
      i. If so, what methods do you use to record and track this research? i.e., when sampling, how do you represent/index samples, and when comparing, by what measures are the comparisons made?
   b. Do you engage in quantification of your data, and if so, what types of mathematical transformations to you subject your data to? This includes anything from calculating enumerative difference (i.e., instance A occurs # times more/less than instance B), calculating percentage, etc.
   c. In what ways do you represent the patterns you observe in your subjects, either in the investigation stage, as a means of revealing patterns, or in the presentation stage, as a means of expressing patterns to others?
9) Now, going through the examples of investigation techniques and tools you have described to me, could you please reflect on the benefits and drawbacks of your current approach? Are your current techniques and tools:
   a. Precise enough?
   b. Easy to apply regardless of the site (i.e., if you travel to a distant library)?

10) Turning now to your use of computing in your work, please describe to me generally to what extent you use computers in:
   a. The discovery stage (i.e., when finding material to make an interesting subject for a study)
   b. The investigation stage (i.e., when examining the textual artifact(s) either generally or in relation to a specific research question)
   c. The analysis stage (i.e., when drawing up theories and structuring arguments to support them—this would include any modelling activities such as drawing)
   d. The presentation stage (i.e., when writing up and/or disseminating your findings)

11) To what extent do you currently use computing on site, in the rare book room, archive, or other location? This would include your own devices as well as the facilities’ devices.

12) How do you feel computing devices currently help or hinder your work?

13) What kind of computing technology do you currently own? This category would include desktops, laptops, tablets, and smart phones; please describe all devices, regardless of whether you currently use them in your studies.

14) In broad strokes what do you think is the purpose of textual scholarship?