

---

## Implementing New Knowledge Environments (INKE)

### Publications

---

Prototyping the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE), Past, Present, and Future Concerns: A Digital Humanities Project Narrative

Ray Siemens, Mike Elkink, Alastair McColl, Karin Armstrong, James Dixon, Angelsea Saby, Brett D. Hirsch and Cara Leitch, with Martin Holmes, Eric Haswell, Chris Gaudet, Paul Girn, Michael Joyce, Rachel Gold, and Gerry Watson, and members of the PKP, Iter, TAPoR, and INKE teams.

2011

© 2011 Siemens et al. This is an open access article distributed under the terms of the Creative Commons 3.0 CC-BY License: <http://creativecommons.org/licenses/by/3.0/>

This article was originally published at:

[http://www.digitalstudies.org/ojs/index.php/digital\\_studies/article/view/182/255](http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/182/255)

---

Citation for this paper:

Siemens, R., Elkink, M., McColl, A., Armstrong, K., Dixon, J., Saby, A., ... INKE. (2011). Prototyping the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE), past, present, and future concerns: A digital humanities project narrative. *Digital Studies/Le champ numérique*, 2(2).

# Digital Studies / Le champ numérique



HOME ABOUT LOG IN REGISTER SEARCH CURRENT  
ARCHIVES ANNOUNCEMENTS

Home > Vol 2, No 2 (2011) > **Siemens (Editor)**

## Prototyping the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE), Past, Present, and Future Concerns: A Digital Humanities Project Narrative

Ray Siemens, University of Victoria: [siemens@uvic.ca](mailto:siemens@uvic.ca)

Mike Elkink, Alastair McColl, Karin Armstrong, James Dixon, Angelsea Saby, Brett D. Hirsch and Cara Leitch, with Martin Holmes, Eric Haswell, Chris Gaudet, Paul Girn, Michael Joyce, Rachel Gold, and Gerry Watson, and members of the PKP, Iter, TAPoR, and INKE teams.

### Abstract / Résumé

The Renaissance English Knowledgebase (REKn) is an electronic knowledgebase consisting of primary and secondary materials (text, image, and audio) related to the Renaissance period. The limitations of existing tools to accurately search, navigate, and read large collections of data in many formats, coupled with the findings of our research into professional reading, led to the development of a Professional Reading Environment (PReE) to meet these needs. Both were conceived as necessary components of a prototype textual environment for an electronic scholarly edition of the Devonshire Manuscript. This article offers an overview of the development of both REKn and PReE at the Electronic Textual Cultures Laboratory (ETCL) at the University of Victoria, from proof of concept through to their current iteration, concluding with a discussion about their future adaptation, implementation, and integration with other projects and partnerships.

### KEYWORDS / MOTS-CLÉS

Renaissance, computer-human interaction, knowledgebase, professional reading, interface design, social annotations, prototyping, social networking / *Renaissance, interaction homme-machine, base de connaissances, lecture professionnelle, conception d'interface, annotations sociales, prototypage, gestion de réseau sociale*

### Contents

- [1. Introduction and Overview](#)
- [2. Conceptual Backgrounds and Critical Contexts](#)
  - [2.1. Conceptual Backgrounds](#)

[Journal Help](#)

USER

Username

Password

Remember me

JOURNAL CONTENT

Search

All

Browse

- [By Issue](#)
- [By Author](#)
- [By Title](#)

FONT SIZE

ARTICLE TOOLS

 [Abstract](#)

 [Print this article](#)

 [Indexing metadata](#)

 [How to cite item](#)

 [Supplementary files](#)

 [Review policy](#)

 [Email this article](#) (Login required)

 [Email the author](#) (Login required)

KEYWORDS

- [2.1.1. New Historicism](#)
    - [2.1.2. The Sociology of Text](#)
    - [2.1.3. Knowledgebases](#)
  - [2.2. Critical Contexts](#)
    - [2.2.1. Knowledge Representation](#)
    - [2.2.2. Professional Reading and Modeling](#)
    - [2.2.3. The Scholarly Edition](#)
    - [2.2.4. Prototyping as a Research Activity](#)
- [3. The Proof of Concept](#)
  - [3.1. Data Structure and Functional Requirements](#)
  - [3.2. Tools and Platforms](#)
  - [3.3. Gathering Primary and Secondary Materials](#)
  - [3.4. Building a Professional Reading Environment](#)
- [4. Research Prototypes: Challenges and Experiments](#)
  - [4.1. Challenge: Scalable Data Storage](#)
  - [4.2. Challenge: Document Harvesting](#)
  - [4.3. Challenge: Standalone vs. Web Application](#)
  - [4.4. Experiment: Shakespeare's Sonnets](#)
  - [4.5. Experiment: The REKn Crawler](#)
    - [4.5.1. Premises](#)
    - [4.5.2 Method](#)
    - [4.5.3. Workflow](#)
    - [4.5.4. Application](#)
- [5. Moving into Full Prototype Development: New Directions](#)
  - [5.1. Rebuilding](#)
  - [5.2. New Directions: Social Networking](#)
    - [5.2.1. Identity and Evaluation](#)
    - [5.2.2. Connections and Communication](#)
    - [5.2.3. User and Content Management](#)
  - [5.3. Designing the PReE Interface](#)
    - [5.3.1. User Needs: Analyzing the Audience](#)
    - [5.3.2. Design Principles, Processes, and Prototypes](#)
- [6. New Insights and Next Steps](#)
  - [6.1. Research Insights and the Humanities Model of Dissemination](#)
  - [6.2. Partnerships and Collaborations](#)
- [Appendix 1: Addresses and Presentations](#)
  - [2003](#) [2004](#) [2005](#) [2006](#) [2007](#) [2008](#) [2009](#)
- [Appendix 2: Prototype Development Platform](#)
  - [Ruby on Rails](#)
  - [Zotero](#)
  - [eXist](#)
  - [Solr](#)
  - [Fedora Commons, Fedora GSearch, and RubyFedora](#)
- [Works cited](#)

[berita terkini](#)  
[Canada Digital Humanities](#)  
[DocumentaryEditing](#)  
[Edition Electronic](#)  
[Editing History](#)  
[Humanities Computing](#)  
[XML computing](#)  
[interface](#)  
[design](#) lateral  
[projects loyalists](#)  
[nineteenth century ontology page](#)  
[Ege. manuscript. leaf. script.](#)  
[Beauvais Missal.](#)  
[digital reconstruction.](#)  
[fragment pedagogy project](#)  
[development technology text](#)  
[analysis text visualization virtual reality](#)

## 1. Introduction and Overview

The Renaissance English Knowledgebase (REKn) is a prototype research knowledgebase consisting of a large dynamic corpus of both primary (15,000 text, image, and audio objects) and secondary (some 100,000 articles, e-books, etc.) materials. Each electronic document is stored in a database along with its associated metadata and, in the case of many text-based materials, a light XML encoding. The data is queried, analyzed and examined through a stand-alone prototype document-centered reading client called the Professional Reading Environment (PReE), written for initial prototyping in .NET and, in a more recent implementation, with key parts modeled in Ruby on Rails.

Recently, both projects have moved into new research developmental contexts, requiring some dramatic changes in direction from our earlier proof of concept.

For the second iteration of PReE, our primary goal continues to be to translate it from a desktop environment to the Internet. By following a web-application paradigm we are able to take advantage of superior flexibility in application deployment and maintenance, the ability to receive and disseminate user-generated content, and multi-platform compatibility. As for REKn, experimentation with the prototype has seen the binary and textual data transferred from the database into the file system, affording gains in manageability and scalability and the ability to deploy third-party index and search tools.

As initial proofs-of-concept, REKn and PReE evoked [James Joyce's](#) apt comment that "a man of genius makes no mistakes;" rather, that "his errors are volitional and are the portals of discovery" (156). In our case, we set out to develop a "project of genius" and found that our errors (volitional or, as was more often the case, accidental) certainly provided the necessary direction to pursue a more usable and useful reading environment for professional readers (on the importance of imperfection and failure, especially as it pertains to a digital humanities audience, see John Unsworth's "[Documenting](#)").

This article offers a brief outline of the development of both REKn and PReE at the Electronic Textual Cultures Laboratory (ETCL) at the University of Victoria, from proof of concept through to their current iterations, concluding with a discussion about their future adaptations, implementations, and integrations with other projects and partnerships. This narrative situates REKn and PReE within the context of prototyping as a research activity, and documents the life cycle of a complex digital humanities research program that is itself part of larger, ongoing, iterative programs of research. Much of the content of the present article has been presented in other forms elsewhere (see [Appendix 1](#) for a list of addresses and presentations from which the present article is drawn); as noted below, the rapidity of developments in the digital humanities is such that oral presentation is usually considered the best method for delivery of new results, with subsequent print publication ensuring breadth of dissemination and archival preservation.

## 2. Conceptual Backgrounds and Critical Contexts

### 2.1. Conceptual Backgrounds

The conceptual origins of REKn may be located in two fundamental shifts in literary studies in the 1980s: first, in the emergence of New Historicism and the rise of the sociology of the text; second, in the proliferation of large-scale text-corpus humanities computing projects in the late 1980s and early 1990s (while it may be useful to give a brief overview of these movements, New Historicism and Social Textual Theory in particular have far too broad a bibliography to be engaged with critically in this article; readers interested in more detailed treatment of both movements can begin with [Erickson](#), [Howard](#), and [Pechter](#) for New Historicism and [Tanselle](#) and [Greetham](#) for the sociology of the text).

#### 2.1.1. New Historicism

New Historicism situated itself in opposition to earlier critical traditions that dismissed historical and cultural context as irrelevant to literary study, and proposed instead that "literature exists not in isolation from social questions but as a dynamic participant in the messy processes of cultural formation" ([Hall](#) vii). Thus, New Historicism eschewed the distinction between text and context, arguing that both "are equal partners in the production of culture" ([Hall](#) vii). In Renaissance studies, as elsewhere, this ideological shift challenged scholars to engage not only with the traditional canon of literary works but also with the whole corpus of primary materials at their disposal. As New Historicism blurred the lines between the literary and non-literary, its proponents were quick to illustrate that all cultural forms—literary and non-literary, textual and visual—could be freely and fruitfully "read" alongside and against one another.

#### 2.1.2. The Sociology of Text

A concurrent paradigm shift in bibliographical circles was the rise of the social

theory of text exemplified in the works of [Jerome J. McGann](#) and [D. F. McKenzie](#). According to Kathryn Sutherland, "[i]f the work is not confined to the historically contingent and the particular," the social theory of text posited, "it is nevertheless only in its expressive textual form that we encounter it, and material conditions determine meanings" ("[Introduction](#)" 5). In addition to being "an argument against the notion that the physical book is the disposable container," as Sutherland has suggested, "it is also an argument in favor of the significance of the text as a situated act or event, and therefore, under the conditions of its reproduction, necessarily multiple" ("[Introduction](#)" 6).

In other words, the social theory of text rejected the notion of individual literary authority in favor of a model where social processes of production disperse that authority. According to this view, the literary "text" is not solely the product of authorial intention, but the result of interventions by many agents (such as copyists, printers, publishers) and material processes (such as revision, adaptation, publication). In practical terms, the social theory of text revised the role of the textual scholar and editor, who, no longer concerned with authorial intention, instead focused on recovering the "social history" of a text—that is, the multiple and variable forms of a text that emerge out of these various and varied processes of mediation, revision, and adaptation.

Developments in New Historicism and the Sociology of the text led in the late 1980s and early 1990s to a proliferation of Renaissance text-corpus humanities computing projects in North America, Europe, and New Zealand (representative examples include: the *Women Writers Project*; the *Century of Prose Corpus* the *Early Modern English Dictionaries Database*; the *Michigan Early Modern English Materials*; the *Oxford Text Archive*; the *Riverside STC Project*; the *Shakespeare Database Project*; and the *Textbase of Early Tudor English*).

In many ways, this development seems inevitable. Spurred on by the project of New Historicism and the rise of interest in the sociology of texts, Renaissance scholars were eager to engage with a vast body of primary and secondary materials in addition to the traditional canon of literary works. Developments computing and the humanities led to the realization that textual analysis, interpretation, and synthesis might be pursued with greater ease and accuracy through the use of an integrated electronic database.

A group of scholars involved in such projects, recognizing the value of collaboration and centralized coordination, engaged in a planning meeting towards the creation of a *Renaissance Knowledge Base* (RKB). Consisting of "the major texts and reference materials [...] recognized as critical to Renaissance scholarship," the RKB hoped to "deliver unedited primary texts" including "old-spelling texts of major authors (Sidney, Marlowe, Spenser, Shakespeare, Jonson, Donne, Milton, etc.), the *Short-Title Catalogue* (1475–1640), the *Dictionary of National Biography*, period dictionaries (Florio, Elyot, Cotgrave, etc.), and the *Oxford English Dictionary*" ([Richardson and Neuman](#) 2). With this collection, the project intended to "allow users to search a variety of primary and secondary materials simultaneously," and to stimulate "interpretations by making connections among many kinds of texts" ([Richardson and Neuman](#) 1-2). Addressing the question of "Who needs RKB?" the application offered the following response:

Lexicographers [need the RKB] in order to revise historical dictionaries (the *Oxford English Dictionary*, for example, is based on citation slips, not on the original texts). Literary critics need it, because the RKB will reveal connections among Renaissance works, new characteristics, and nuances of meaning that only a lifetime of directed reading could hope to provide. Historians need the RKB, because it will let them move easily, for example, from biography to textual information. The same may be said of scholars in linguistics, Reformation theology, humanistic philosophy, rhetoric, and socio-cultural studies, among others. ([Richardson and Neuman](#) 2)

The need for such a knowledgebase was (and is) clear. Since each of its individual components were deemed "critical to Renaissance scholarship," and because the RKB intended to "permit each potentially to shed light on all the others," the group behind the RKB felt that "the whole" was "likely to be far greater than the sum of its already-important parts" ([Richardson and Neuman](#) 2).

Recommendations following the initiative's proposal suggested a positive path, drawing attention to the merit of the approach and suggesting further ways to bring about the creation of this resource to meet the research needs of an even larger group of Renaissance scholars. Many of the scholars involved persevered, organizing an open meeting on the RKB at the 1991 ACH/ALLC Conference in Tempe to determine the next course of action. Also present at that session were Eric Calaluca (Chadwyck-Healey), Mark Rooks (InteLex), and Patricia Murphy, all of whom proposed to digitize large quantities of primary materials from the English Renaissance.

From here, the RKB project as originally conceived took new (and largely unforeseen) directions. Chadwyck-Healey was to transcribe books from the *Cambridge Bibliography of English Literature* and publish various full-text databases now combined as *Literature Online*. InteLex was to publish its *Past Masters* series of full-text humanities databases, first on floppy disk and CD-ROM and now web-based. Murphy's project to scan and transcribe large numbers of books in the *Short-Title Catalogue* to machine-readable form was taken up by *Early English Books Online* and later the *Text Creation Partnership*. In the decade since the scholars behind the RKB project first identified the need for a knowledgebase of Renaissance materials, its essential components and methodology have been outlined ([Lancashire "Bilingual"](#)). Moreover, considerable related work was soon to follow, some by the principals of the RKB project and much by those beyond it, such as R. S. Bear (*Renascence Editions*), Michael Best (*Internet Shakespeare Editions*), Gregory Crane (*Perseus Digital Library*), Patricia Fumerton (*English Broadside Ballad Archive*), Ian Lancashire (*Lexicons of Early Modern English*), and Greg Waite (*Textbase of Early Tudor English*); by commercial publishers such as Adam Matthew Digital (*Defining Gender, 1450–1910; Empire Online; Leeds Literary Manuscripts; Perdita Manuscripts; Slavery, Abolition and Social Justice, 1490–2007; Virginia Company Archives*), Chadwyck-Healey (*Literature Online*), and Gale (*British Literary Manuscripts Online, c.1660–c.1900; State Papers Online, 1509–1714*); and by consortia such as *Early English Books Online – Text Creation Partnership* (University of Michigan, Oxford University, the Council of Library and Information Resources, and ProQuest) and *Orlando* (Cambridge University Press and University of Alberta).

As part of the shift from print to electronic publication and archiving, work on digitizing necessary secondary research materials has been handled chiefly, but not exclusively, by academic and commercial publishers. Among others, these include Blackwell (*Synergy*), Cambridge University Press, Duke University Press (*eDuke*), eBook Library (*EBL*), EBSCO (*EBSCOhost*), Gale (*Shakespeare Collection*), Google (*Google Book Search*), Ingenta, JSTOR, netLibrary, Oxford University Press, Project MUSE, ProQuest (*Periodicals Archive Online*), Taylor & Francis, and University of California Press (*Caliber*). Secondary research materials are also being provided in the form of (1) open access databases, such as the *Database of Early English Playbooks* (Alan B. Farmer and Zachary Lesser), the *English Short Title Catalogue* (British Library, Bibliographical Society, and the Modern Language Association of America), and the *REED Patrons and Performance Web Site* (Records of Early English Drama and the University of Toronto); (2) open access scholarly journals, such as those involved in the *Public Knowledge Project* or others listed on the *Directory of Open Access Journals*; and, (3) printed books actively digitized by libraries, independently and in collaboration with organizations such as Google (*Google Book Search*) or the Internet Archive (*Open Access Text Archive*).

Even with this sizeable amount of work on primary and secondary materials accomplished or underway, a compendium of such materials is currently unavailable, and, even if it were, there is no system in place to facilitate navigation and dynamic interaction with these materials by the user (much as one might query a database) and by machine (with the query process automated or semi-automated for the user). There are, undoubtedly, benefits in bringing all of these disparate materials together with an integrated knowledgebase approach. Doing so would facilitate more efficient professional engagement with these materials, offering scholars a more convenient, faster, and deeper handling of research resources. For example, a knowledgebase approach would remove the need to search across multiple databases and listings, facilitate searching across primary and secondary materials simultaneously, and allow deeper, full-text searching of all records, rather than simply relying on indexing information alone—which is often not generated by someone with field-specific knowledge. An integrated knowledgebase—whether

the integration were actual (in a single repository) or virtual (via federated searching and/or other means)—would also encourage new insights and allow researchers new ways to consider relations between texts and materials and their professional, analytical contexts. This is accomplished by facilitating conceptual and thematic searches across all pertinent materials, via the incorporation of advanced computing search and analysis tools that assist in capturing connections between the original objects of contemplation (primary materials) and the professional literature about them (secondary materials).

## 2.2. Critical Contexts

### 2.2.1. Knowledge Representation

Other important critical contexts within which REKn is situated arise out of theories and methodologies associated with the emerging field of digital humanities. When considering a definition of the field, Willard McCarty warns that we cannot "rest content with the comfortably simple definition of humanities computing as *the application of the computer to the disciplines of the humanities*," for to do so "fails us by deleting the agent-scholar from the scene" and "by overlooking the mediation of thought that his or her use of the computer implies" ("Humanities Computing" n.p.). After McCarty, Ray Siemens and Christian Vandendorpe suggest that digital humanities or "humanities computing" as a research area "is best defined loosely, as the intersection of computational methods and humanities scholarship" ("Canadian" xii; see also Rockwell).

A foundation for current work in humanities computing is *knowledge representation*, which Unsworth has described as an "interdisciplinary methodology that combines logic and ontology to produce models of human understanding that are tractable to computation" ("Knowledge" n.p.). While fundamentally based on digital algorithms, as Unsworth has noted, *knowledge representation* privileges traditionally held values associated with the liberal arts and humanities, namely: general intelligence about human pursuits and the human social/societal environment; adaptable, creative, analytical thinking; critical reasoning, argument, and logic; and the employment and conveyance of these in and through human communicative processes (verbal and non-verbal communication) and other processes native to the humanities (publication, presentation, dissemination). With respect to the activities of the computing humanist, Siemens and Vandendorpe suggest that *knowledge representation* "manifests itself in issues related to archival representation and textual editing, high-level interpretive theory and criticism, and protocols of knowledge transfer—all as modeled with computational techniques" (xii).

### 2.2.2. Professional Reading and Modeling

A primary protocol of knowledge transfer in the field of the humanities is reading. However, there is a substantial difference between the reading practices of humanists and those readers outside of academe—put simply, humanists are professional readers. As John Guillory suggests, there are four characteristics of professional reading that distinguish it from the practice of lay reading:

First of all, it is a kind of *work*, a labor requiring large amounts of time and resources. This labor is compensated as such, by a salary. Second, it is a *disciplinary* activity, that is, it is governed by conventions of interpretation and protocols of research developed over many decades. These techniques take years to acquire; otherwise we would not award higher degrees to those who succeed in mastering them. Third, professional reading is *vigilant*; it stands back from the experience of pleasure in reading [...] so that the experience of reading does not begin and end in the pleasure of consumption, but gives rise to a certain sustained reflection. And fourth, this reading is a *communal* practice. Even when the scholar reads in privacy, this act of reading is connected in numerous ways to communal scenes; and it is often dedicated to the end of a public and publishable "reading." (31-32)

Much recent work in the digital humanities focuses on modeling professional reading and other activities associated with conducting and disseminating

humanities research (on the importance of reading as an object of interest to humanities computing practitioners see [Warwick](#); professional reading tools are discussed in [Siemens et al.](#) "Iter " and "May Change"). Modeling the activities of the humanist (and the output of humanistic achievement) with the assistance of the computer has identified the exemplary tasks associated with humanities computing: the representation of archival materials; analysis or critical inquiry originating in those materials; and the communication of the results of these tasks (On modeling in the humanities, see [McCarty "Modeling,"](#) and, as it pertains to literary studies in particular, [McCarty "Knowing"](#)). As computing humanists, we assume that all of these elements are inseparable and interrelated, and that all processes can be facilitated electronically.

Each of these tasks will be described in turn. In reverse order, the communication of results involves the electronic dissemination of, and electronically facilitated interaction about the product of, archival representation and critical inquiry, as well as the digitization of materials previously stored in other archival forms (see [Miall](#)). Communication of results takes place via codified professional interaction, and is traditionally held to include all contributions to a discipline-centered body of knowledge—that is, all activities that are captured in the scholarly record associated with the shared pursuits of a particular field. In addition to those academic and commercial publishers and publication amalgamator services delivering content electronically, pertinent examples of projects concerned with the communication of results include the [Open Journal Systems](#), [Open Monograph Press](#) (Public Knowledge Project) and [Collex](#) (NINES), as well as services provided by [Synergies](#) and the [Canadian Research Knowledge Network / Réseau Canadien de Documentation pour la Recherche](#) (CRKN/RCDR).

Critical inquiry involves the application of algorithmically facilitated search, retrieval, and critical processes that, although originating in humanities-based work, have been demonstrated to have application far beyond (Representative examples include [Lancashire "Computer"](#) and [Fortier](#)). Associated with critical theory, this area is typified by interpretive studies that assist in our intellectual and aesthetic understanding of humanistic works, and it involves the application (and applicability) of critical and interpretive tools and analytic algorithms on digitally represented texts and artifacts. Pertinent examples include applications such as [Juxta](#) (NINES), as well as tools developed by the [Text Analysis Portal for Research](#) (TAPoR) project, the [Metadata Offer New Knowledge](#) (MONK) project, the Software Environment for the Advancement of Scholarly Research (SEASR), and by [Many Eyes](#) (IBM).

Archival representation involves the use of computer-assisted means to describe and express print-, visual-, and audio-based material in tagged and searchable electronic form (see [Hockey](#) for a detailed discussion). Associated as it is with the critical methodologies that govern our representation of original artifacts, archival representation is chiefly bibliographical in nature and often involves the reproduction of primary materials such as in the preparation of an electronic edition or digital facsimile either in the context of a scholarly project such as those mentioned above, or in the context of digitization projects undertaken by organizations such as the Internet Archive, Google, libraries, museums, and similar institutions. Key issues in archival representation include considerations of the modeling of objects and processes, the impact of social theories of text on the role and goal of the editor, and the "death of distance" (term coined by [Paul Delany](#)).

Ideally, object modeling for archival representation should simulate the original object-artifact, both in terms of basic representation (e.g. a scanned image of printed page) and functionality (such as the ability to "turn" or otherwise "physically" manipulate the page). However, object modeling need not simply be limited to simulating the original. Although "a play script is a poor substitute for a live performance," [Mueller](#) has shown that "however paltry a surrogate the printed text may be, for some purposes it is superior to the 'original' that it replaces" (61). The next level of simulation beyond the printed surrogate, namely the "digital surrogate," would similarly offer further enhancements to the original. These enhancements might include greater flexibility in the basic representation of the object (such as magnification and otherwise altering its appearance) or its functionality (such as fast and accurate search functions, embedded multimedia, etc).

Archival representation might then involve modeling the process of interaction

between the user and the object-artifact. Simulating the process affords a better understanding of the relationships between the object and the user, particularly as that relationship reveals the user's disciplinary practices—discovering, annotating, comparing, referring, sampling, illustrating, and representing (see [Unsworth "Scholarly"](#)).

### 2.2.3. The Scholarly Edition

The recent convergence of social theories of text and the rise of the electronic medium has had a significant impact on both the function of the scholarly edition and the role of the textual scholar. As [Susan Schreibman](#) argues, "the release from the spatial restrictions of the codex form has profoundly changed the focus of the textual scholar's work," from "publishing a single text with apparatus which has been synthesized and summarized to accommodate to codex's spatial limitations" to creating "large assemblages of textual and non-textual lexia, presented to readers with as little traditional editorial intervention as possible" (284). In addition to acknowledging the value of the electronic medium to editing and the edition, such "assemblages" also recognize the critical practice of "unediting," whereby the reader is exposed to the various layers of editorial mediation of a given text, as well as an increased awareness of the "materiality" of the text-object under consideration (on "unediting" in this sense, see [Marcus](#); on "unediting" as the rejection of critical editions in preference to the unmediated study of originals or facsimiles, see [McLeod "UnEditing."](#) The materiality of the Renaissance text is discussed in [De Grazia and Stallybrass](#) and [Sutherland "Revised"](#)).

Perfectly adaptable to, and properly enabling of, social theories of text and the role of editing, the electronic medium has brought us closer to the textual objects of our contemplation, even though we remain at the same physical distance from them. Like other enabling communicative and representative technologies that came before it, the electronic medium has brought about a "death of distance." This notion of a "death of distance," as discussed by [Delany](#), comes from a world made smaller by travel and communication systems, a world in which we have "the ability to do more things without being physically present at the point of impact" (50). The textual scholar, accumulating an "assemblage" of textual materials, does so for those materials to be, in turn, re-presented to those any who are interested in those materials. More and more, though, it is not only primary materials—textual witnesses for example—that are being accumulated and re-presented. The "death of distance" applies also to objects that have the potential to shape and inform further our contemplation of those physical objects of our initial contemplation, namely, the primary materials (see also [Siemens "Unediting"](#)).

We understand, almost intuitively, the end-product of the traditional scholarly edition in its print codex form: how material is presented, what the scope of that material is, how that material is being related to us and, internally, how the material presented by the edition relates to itself and to materials beyond those directly presented—secondary texts, contextual material, and so forth. Our understanding of these things as they relate to the *electronic* scholarly edition, however, is only just being formed. We are at a critical juncture for the scholarly edition in electronic form, where the "assemblages" and accumulation of textual archival materials associated with social theories of text and the role of editing meet their natural home in the electronic scholarly edition; and, such the large collections of primary materials in electronic form that result from this also meet their equivalent in the world of secondary materials, that ever-growing body of scholarship that informs those materials ([Siemens "Unediting"](#) 426).

To date, two models of the electronic scholarly edition have prevailed. One is the notion of the "dynamic text," which consists of an electronic text and integrated advanced textual analysis software. In essence, the dynamic text presents a text that indexes and concords itself and allows the reader to interact with it in a dynamic fashion, enacting text analysis procedures upon it as it is read ([Lancashire "Working: Bolton"](#) is an exemplary example of three early "dynamic text" Shakespeare editions). The other, often referred to as the "hypertextual edition," exploits the ability of encoded hypertextual organization to facilitate a reader's interaction with the apparatus (textual, critical, contextual, and so forthetc.) that traditionally accompanies scholarly editions, as well as with relevant external textual and graphical resources, critical materials, and so forth (the elements of the hypertextual edition were rightly

anticipated in [Faulhaber](#)).

Advances over the past decade have made it clear that electronic scholarly editions can in fact enjoy the best of both worlds, incorporating elements from the "dynamic text" model—namely, dynamic interaction with the text and its related materials—while at the same time reaping the benefits of the fixed hypertextual links characteristically found in "hypertextual editions." Indeed, scholarly consensus is that the level of dynamic interaction in an electronic edition itself—if facilitated via text analysis in the style of the "dynamic text"—could replace much of the interaction that one typically has with a text and its accompanying materials via explicit hypertextual links in a hypertextual edition. At the same time, there is at present no extant exemplary implementation of this new *dynamic edition*—an edition that transfers the principles of interaction afforded by a dynamic text to the realm of the full edition, comprising of that text and all of its extra- and para-textual materials, textual apparatus, commentary, and beyond.

#### 2.2.4. Prototyping as a Research Activity

In addition to the aforementioned critical contexts, it is equally important to situate the development of REKn and PReE within a methodological context of prototyping as a research activity.

The process of prototyping in the context of our work involves constructing a functional computational model that embodies the results of our research, and, as an object of further study itself, undergoes iterative modification in response to research and testing. A prototype in this context is an interface or visualization that embodies the theoretical foundations our work establishes, so that the theory informing the creation of the prototype can itself be tested by having people use it (see [Sinclair and Rockwell](#) for an example; also in this context the discussion of modeling in [McCarty](#) "Modeling" and "Knowing").

An example of a prototypical tool that performs an integral function in a larger digital reading environment is the Dynamic Table of Contexts, an experimental interface that draws on interpretive document encoding to combine the conventional table of contents with an interactive index (see [Ruecker](#); [Ruecker et al.](#); and [Brown et al.](#)). Readers use the Dynamic Table of Contexts as a tool for browsing the document by selecting an entry from the index and seeing where it is placed in the table of contents. Each item also serves as a link to the appropriate point in the file.

Research prototypes such as those we set out to develop, in other words, are distinct from prototypes designed as part of a production system in that the research prototype focuses chiefly on providing limited but research-pertinent functionality within a larger framework of assumed operation. Production systems, on the other hand, require full functionality and are often derived from multiple prototyping processes.

### 3. The Proof of Concept

REKn was originally conceived as part of a wider research project to develop a prototype textual environment for a *dynamic edition*: an electronic scholarly edition that models disciplinary interaction in the humanities, specifically in the areas of archival representation, critical inquiry, and the communication of results. Centered on a highly encoded electronic text, this environment facilitates interaction with the text, with primary and secondary materials related to it, and with scholars who have a professional engagement with those materials. This ongoing research requires (1) the adaptation of an exemplary, highly-encoded and properly-imaged electronic base text for the edition; (2) the establishment of an extensive knowledgebase to exist in relation to that exemplary base text, composed of primary and secondary materials pertinent to an understanding of the base text and its literary, historical, cultural, and critical contexts; and (3) the development of a system to facilitate navigation and dynamic interaction with and between materials in the edition and in the knowledgebase, incorporating professional reading and analytical tools; to allow those materials to be updated; and to implement communicative tools to facilitate computer-assisted interaction between users engaging with the materials.

This second point in particular represents an important distinction between REKn and the earlier RKB project: while RKB set out to include "old-spelling texts of major authors (Sidney, Marlowe, Spenser, Shakespeare, Jonson, Donne, Milton, etc.), the *Short-Title Catalogue* (1475–1640), the *Dictionary of National Biography*, period dictionaries (Florio, Elyot, Cotgrave, etc.), and the *Oxford English Dictionary*" (Richardson and Neuman 2), REKn is not limited to "major authors" but seeks to include all canonical works (in print and manuscript) and most extra-canonical works (in print) of the period.

The electronic base-text selected to act as the initial focal point for the prototype was drawn from Ray Siemens's Social Sciences and Humanities Research Council (SSHRC)-funded electronic scholarly edition of the Devonshire Manuscript (BL MS Add. 17492). Characterized as a "courtly anthology" (Southall "Devonshire" 15; *Courtly*) and as an "informal volume" (Remley 48), the Devonshire Manuscript is a poetic miscellany consisting of 114 original leaves, housing some 185 items of verse (complete poems, fragments, extracts from larger extant works, and scribal annotations). Historically privileged in literary history as a key witness of Thomas Wyatt's poetry, the manuscript has received new and significant attention of late, in large part because of the way in which its contents reflect the interactions of poetry and power in early Renaissance England and, more significantly, because it offers one of the earliest examples of the explicit and direct participation of women in the type of literary and political-poetic discourses found in the document (on the editing of the Devonshire Manuscript in terms of modeling and knowledge representation, see Siemens and Leitch and Siemens et al. "Drawing").

While editing the Devonshire Manuscript as the base text was underway, work on REKn began by mapping the data structure in relation to the functional requirements of the project, selecting appropriate tools and platforms, and outlining three objectives: to gather and assemble a corpus of primary and secondary texts to make up the knowledgebase; to develop automated methods for data collection; and to develop software tools to facilitate dynamic interaction between the user(s) and the knowledgebase.

### 3.1. Data Structure and Functional Requirements

We felt that the database should include tables to store relations between documents; that is, if a document includes a reference to another document, whether explicitly (such as in a reference or citation) or implicitly (such as in keywords and metadata), the fact of that reference or relation should be stored. Thus, the document-to-document relationship will be a many-to-many relationship.

In addition to a web service for public access to the database, it was proposed that there should be a standalone data entry and maintenance application to allow the user(s) to create, update, and delete database records manually. This application should include tools for filtering markup tags and other formatting characters from documents; allow for automating the data entry of groups of documents; and allow for automating the data entry of documents where they are available from web services, or by querying electronic academic publication amalgamator services (such as *EBSCOhost*).

Finally, a scholarly research application to query the database in read-only mode and display documents—along with metadata where available (such as author, title, publisher)—was to be developed. The appearance and operation of the application should model the processes of scholarly research, with many related documents visible at the same time, easily moved and grouped by the researcher. The application should display the document in as many different forms as are available—plain text, marked up text, scanned images, audio streams, and so forth. Users should also be able to navigate easily between related documents; to search easily for documents that have similar words, phrases or word patterns; and to perform text analysis on the document(s)—word list, word frequency, word collocation, word concordance—and display the results.

### 3.2. Tools and Platforms

The database management system chosen for the REKn prototype was PostgreSQL. As a standard system commonly used by the academic community, PostgreSQL allows for future collaboration with other researchers

and integration with other projects. PostgreSQL's open source status caters to the possibility of writing custom functions and indexes that cannot be supplied by other means. Moreover, PostgreSQL offers scaling and clustering of database systems and the data in the systems. Redundancy is also possible with PostgreSQL—that is, if one server in a cluster crashes, the others will continue processing queries and data uninterrupted.

A similar rationale dictated writing the web service in PHP, since PHP is a commonly used and well-understood framework for database access via the Internet, in addition to being open source. The data-entry application is likewise based on Perl scripts to use the web service as a database access proxy, since in addition to being open source software, Perl is well suited for string processing.

### 3.3. Gathering Primary and Secondary Materials

The gathering of primary materials for the knowledgebase was initially accomplished by pulling down content from open-access archives of Renaissance texts, and by requesting materials from various partnerships (researchers, publishers, scholarly centers) interested in the project. These materials included a total of some 12,830 texts in the public domain or otherwise generously donated by EEBO-TCP (9,533), Chadwyck-Healey (1,820), *Text Analysis Computing Tools* (311), the *Early and Middle English Collections* from the University of Virginia Electronic Text Centre (273 and 27 respectively), the Brown *Women Writers Project* (241), the *Oxford Text Archive* (241), the *Early Tudor Textbase* (180), *Renascence Editions* (162), the *Christian Classics Ethereal Library* (65), *Elizabethan Authors* (21), the Norwegian University of Science and Technology (8), the Richard III Society (5), the University of Nebraska School of Music (4), *Project Bartleby* (2), and *Project Gutenberg* (2) (see "[Subsidium: Master List of REKn Primary Sources](#)" for a master list of the primary text titles and their sources). The harvesting and initial integration of these materials took a year, during which time various formats of almost 4 gigabytes of files were standardized into a basic TEI-compliant XML format. Roughly a dozen different implementations of XML, SGML, COCOA, HTML, plain text, and more eclectic encoding systems were accommodated.

For example, accommodating the XML TEI P4 conforming documents obtained from the University of Virginia Electronic Text Centre's *Early English Collection* required the following three-step process:

- `EarlyUVaStepOne.xsl`: Application of an XSL transformation to remove the unnecessary XML tags and to restructure the document using our internal-use tags. This step also derived a minimal set of metadata necessary for identifying the document with bibliographic MARC records.
- `EarlyUVaStepTwo.xsl`: Cleaning, stripping, and possible restructuring of documents from step one. This step also transformed the XML list of our metadata into an HTML list, built links to the HTML and XML files, and provided some rudimentary navigation and statistics.
- `EarlyUVaToHTML.xsl`: Simple transformation (applied to either the source document or to the result of the `EarlyUVaStepOne.xsl` transformation) intended produced HTML suitable for web browsers. These transformations introduce minimal HTML tagging; when we wish to serve more polished products to web browsers, this XSLT will serve as a starting point.

The bulk of the primary material was so substantial that harvesting the secondary materials manually would be too onerous a task—clearly, automated methods were desirable and would allow for continual and ongoing harvesting of new materials as they became available. Ideally, these methods should be general enough in nature so that they can be applied to other types of literature, requiring minimal modification for reuse in other fields. This emphasis on transportability and scalability would ensure that the form and structure of the knowledgebase could be used in other fields of scholarly research.

Initially, the strategy was to assemble a sample database of secondary materials in partnership with the University of Victoria Libraries, gathering materials harvested automatically from electronic academic publication amalgamator services (such as *EBSCOhost*). An automated process was developed to retrieve relevant documents and store them in a purpose-built

database. This process would query remote databases with numerous search strings, weed out erroneous and duplicate entries, separate metadata from text, and store both in a relational database. The utility of our harvesting methods would then be demonstrated to the amalgamators and other publishers with the intent of fostering partnerships with them.

### 3.4. Building a Professional Reading Environment

At this stage REKn contained roughly 80 gigabytes of text data, consisting of some 12,830 primary text documents and an ongoing collection of secondary texts in excess of 80,000 documents; together with associated image data, the complete collection was estimated to be in the 2 to 3 terabyte range. Given its immense scale, development of a document viewer with analytical and communicative functionality to interact with REKn was a pressing issue. The inability of existing tools to search, navigate, and read large collections of data accurately and in many formats, later coupled with the findings of our research into professional reading, led to the development of a Professional Reading Environment (PRE).

Initially designed as a desktop GUI to the PostgreSQL database containing REKn, the PRE proof of concept was developed as a .NET Windows Form application. Very little consideration was given to further use of the code at this stage—the focus was solely on testing whether it all could work. Using .NET Framework was justified on the grounds that it is the standard development platform for Microsoft Windows machines, presumably used by a large portion of our potential users. Developing the proof of concept in .NET Framework meant that the application could use the resources of the client's machine to a greater extent than if the application were housed in a browser. Local processing would be necessary if, for example, users were to use image-processing tools on scanned manuscript pages.

As demonstrated in the video below ([Video 1](#)), the proof of concept built in .NET sported a number of useful features. Individual users were able to log in, opening as many separate document-centred instances of the GUI as they desired simultaneously, and perform search, reading, analytical, and composition and communication functions. These functions, in turn, were drawn on our modeling of professional reading and other activities associated with conducting and disseminating humanities research. Searches could be conducted on document metadata and citations (by author, title, and keyword) for both primary and secondary materials ([Figure 1](#)). A selected word or phrase could also spawn a search of documents within the knowledgebase, as well as a search of other Internet resources (such as the [Oxford English Dictionary Online](#) and [Lexicons of Early Modern English](#)) from within PRE. Similarly, the user could use TAPoR Tools to perform analyses on the current text or selected words and phrases in PRE ([Figure 2](#)).

Figure 1: Metadata Search and Search Results.

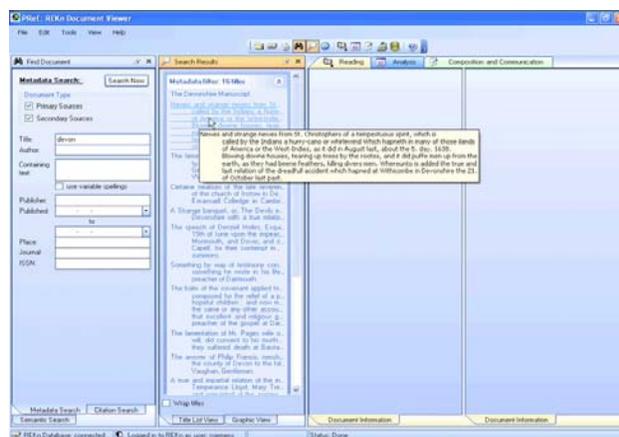
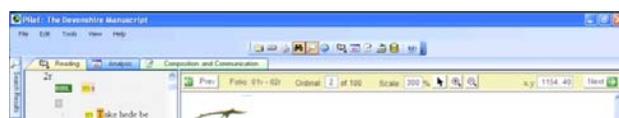
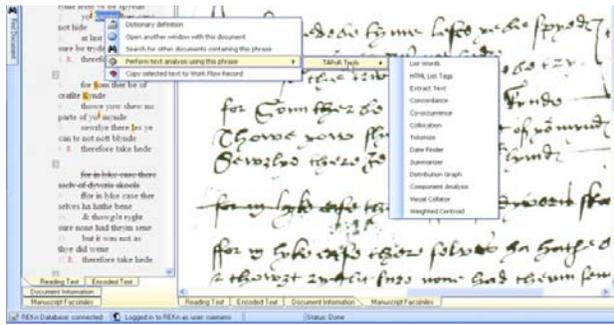


Figure 2: Spawned Search and Analytical Functions.





The proof-of-concept build could display text data in a variety of forms (plain-text, HTML, and PDF) and display images of various formats (Figure 3 and Figure 4). Users could zoom in and out when viewing images, and scale the display when viewing texts (Figure 5). If REKn contained different versions of an object—such as images, transcriptions, translations—they were linked together in PReE, allowing users to view an image and corresponding text data side-by-side (Figure 6).

Figure 3: Reading Text Data.

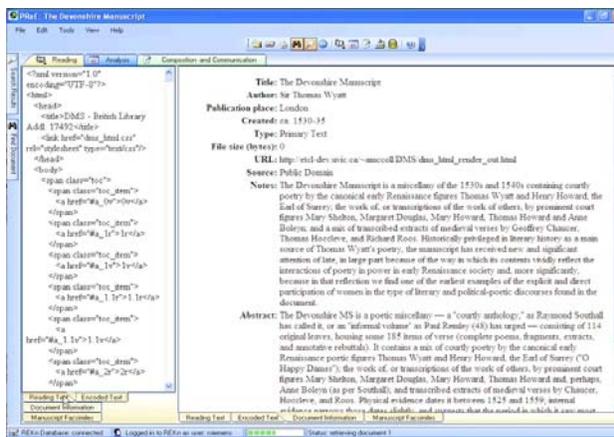


Figure 4: PDF Display.

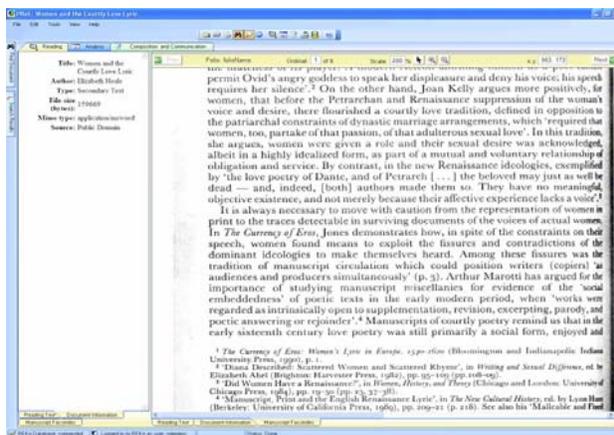


Figure 5: Zoom and Pan Images.

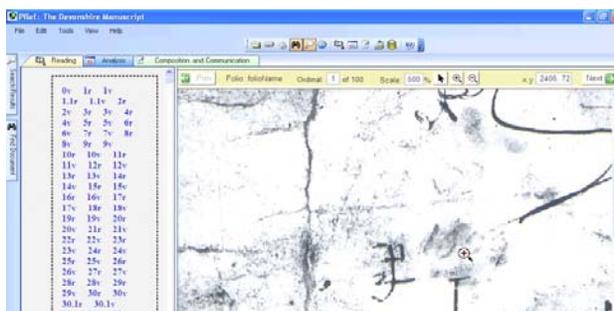
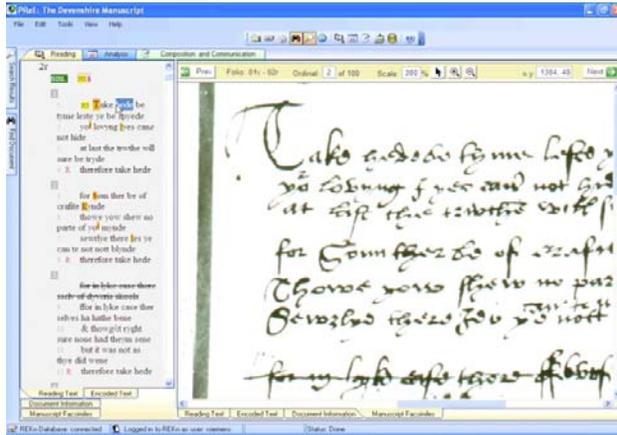


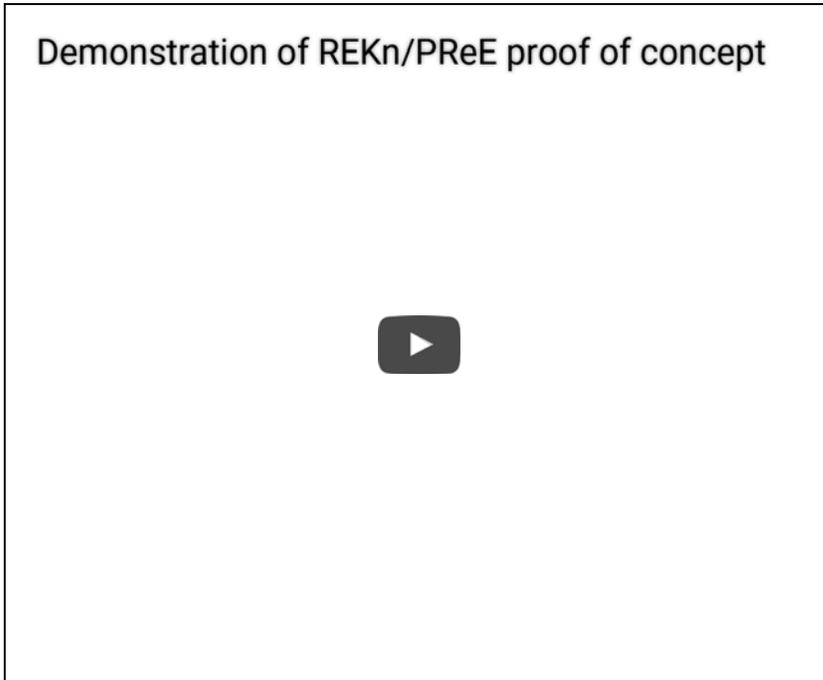


Figure 6: Side-by-Side Display of Texts and Images.



This initial version of PReE also offered composition and communication functions, such as the ability for a user to select a portion of an image or text and to save this to a workflow, or the capacity to create and store notes for later use. Users were also able to track their own usage and document views, which could then be saved to the workflow for later use. Similarly, administrators were able to track user access and use of the knowledgebase materials, which might be of interest to content partners (such as academic and commercial publishers) wishing to use the data for statistical analysis.

Video 1: Demonstration of REKn/PReE proof of concept.



## 4. Research Prototypes: Challenges and Experiments

After the success of our proof of concept, we set out to imagine the next steps of modeling as part of our research program. Indeed, growing interest amongst knowledge providers in applying the concept of a professional reading environment to their databases and similar resources led us to consider how to expand PReE beyond the confines of REKn. After evaluating our progress to

date, we realized that we needed to take what we had learned from the proof of concept and apply that knowledge to new challenges and requirements. Our key focus would be on issues of scalability, functionality, and maintainability.

#### **4.1. Challenge: Scalable Data Storage**

In the proof-of-concept build, all REKn data was stored in binary fields in a database. While this approach had the benefit of keeping all of the data in one easily accessible place, it raised a number of concerns—most pressingly, the issue of scalability. Dealing with several hundred gigabytes is manageable with local infrastructure and ordinary tools; however, we realized that we had to reconsider the tools when dealing in the range of several terabytes. Careful consideration would also be necessary for indexing and other operations that might require exponentially longer processing times as the database increased in size.

Even with a good infrastructure, practical limitations on database content are still an important consideration, especially were we to include large corpora (the larger datasets of the Canadian Research Knowledge Network were discussed, for example) or significant sections of the Internet (via thin-slicing across knowledge-domain-specific data). Setting practical limitations required us to consider what was essential and what needed to be stored—for example, did we have to store an entire document, or could it be simply a URL? Storing all REKn data in binary fields in a database during the proof-of-concept stage posed additional concerns. Incremental backups, for example, required more complicated scripts to look through the database to identify new rows added. Full backups would require a server-intensive process of exporting all of the data in the database. This, of course, could present performance issues should the total database size reach the terabyte range. Equally, to distribute the database in its current state amongst multiple servers would pose no mean feat.

Indexing full-text in a relational database does not give optimum performance or results: in fact, the performance degradation could be described as exponential in relation to the size of the database. Keeping both advantages and disadvantages in mind, it was proposed that all REKn binary data be stored in a file system rather than in the database. File systems are designed to store files, whereas the PostgreSQL database is designed to store relational data. To mix the two defeats the separate advantages of each. Moreover, in testing the proof of concept, users found speed to be a significant issue, with many unwilling to wait five minutes between operations. In its proof-of-concept iteration, the computing interaction simply could not keep pace with the cognitive functions it was intended to augment and assist. We recognized that this issue could be resolved in the future by recourse to high-performance computing techniques—in the meantime, however, we decided to reduce the REKn data to a subset, which would allow us to imagine and work on functionality at a smaller scale.

Having decided to store all binary data in a file system, we had to develop a standardized method of storing and linking the data, one that accounted for both linking the relational data to the file system data as well as keeping the data mobile (such as would allow migrating the data to a new server or distributing the files over multiple servers). Flexibility was also flagged as an important design consideration, since the storage solution might eventually be shared with many different organizations, each with their own particular needs. This method would also require the implementation of a search technology capable of performing fast searches over millions of documents. In addition to the problem posed by the sheer volume of documents, the variety of file types stored would require the employ of an indexing engine capable of extracting text out of encoded files. After a survey of the existing software tools, Lucene presented the perfect fit for our project requirements: it is an open source full-text indexing engine capable of handling millions of files of various types without any major degradation in performance, and it is extensible with plug-ins to handle additional file types should the need arise.

#### **4.2. Challenge: Document Harvesting**

The question of how to go about harvesting data for REKn, or indeed any content-specific knowledgebase, turned out to be a question of negotiating with the suppliers of document collections for permission to copy the documents.

Since each of these suppliers (such as the academic and commercial publishers and the publication amalgamator service providers) has structured access to the documents differently, scripts to allow for harvesting their documents had to be tailored individually for each supplier. For example, some suppliers provide an API to their database, others use HTTP, and still others distribute their documents via tapes or CDs of files. Designing an automated process for harvesting documents from suppliers could be accomplished by combining all of these different scripts together with a mechanism for automatically detecting the various custom access requirements and selecting the correct script to use.

Inserting documents into REKn offered technical challenges as well. Documents from different sources often had different XML structures. Even TEI-standard documents from various sources had different markup tags and elements, depending on the goals of the projects supplying the documents and the particular TEI DTDs used.

### 4.3. Challenge: Standalone vs. Web Application

Developed as a down-and-dirty solution to the original project requirements, PReE at the proof-of-concept stage was built as an installable standalone Windows application; for the second version of PReE, we considered whether to translate it from a desktop environment to the Internet.

The main advantages of following a web-application (or rich Internet application) paradigm are its superior flexibility in application deployment and maintenance, and its ability to receive and disseminate user-generated content and multi-platform compatibility. The main disadvantage is that browsers impose limitations on the design of applications and usually restrict access to the resources (file system and processing) of the local machine.

A major advantage that standalone applications have over web applications is that performance and functionality are not dependent on the speed or availability of an Internet connection. Further, standalone desktop applications are able to use all of the resources of the local machine with very few design restrictions other than those imposed by the target hardware and software tools. However, standalone applications must be installed by each individual user and, as a result, involve a level of training, familiarization, and support, which may discourage some users. Perhaps most importantly, given the goals of the project, standalone applications simply do not offer the same level of multi-platform compatibility or flexibility in application deployment and maintenance.

Essentially the question came down to identifying the features or services users would require, and whether those could be accommodated in the client application. For example, if users required the ability to create files and store them locally on their own machines, it may not have been feasible for the client application to be a web-browser. After weighing the pros and cons, we decided that PReE would be further developed as a web application. This decision was followed by a survey of the relevant applications, platforms, and technologies in terms of their applicability, functionality, and limitations ([Appendix 2](#)).

### 4.4. Experiment: Shakespeare's Sonnets

As outlined above, to facilitate faster prototyping and development of both REKn and PReE it was proposed that REKn should be reduced to a limited dataset. Work was already underway on an electronic edition of Shakespeare's Sonnets, so limiting REKn data to materials related to the Sonnets would offer a more manageable dataset.

Modern print editions of the Sonnets admirably serve the needs of lay readers. For professional readers, however, print editions simply cannot hope to offer an exhaustive and authoritative engagement with the critical literature surrounding the Sonnets, a body of scholarship that is continually growing. Even with the considerable assistance provided by such tools as the *World Shakespeare Bibliography* and the *MLA International Bibliography*, the sheer volume of scholarship published on Shakespeare and his works is difficult to navigate. Indeed, existing databases such as these only allow the user to search for criticism related to the Sonnets through a limited set of metadata, selected and presented in each database according to different editorial priorities, and often by those without domain-specific expertise. Moreover,

while select bibliographies such as these have often helped to organize specific areas of inquiry, the last attempt to compile a comprehensive bibliography of scholarly material on Shakespeare's Sonnets was produced by [Tetsumaro Hayashi](#) in 1972. Although it remains an invaluable resource in indicating the volume and broad outlines of Sonnet criticism, Hayashi's bibliography is unable to provide the particularity and responsiveness of a tool that accesses the entire text of the critical materials it seeks to organize.

Without the restrictions of print, an electronic edition of Shakespeare's Sonnets could be both responsive to the evolution of the field, updating itself periodically to incorporate new research, and more flexible in the ways in which it allows users to navigate and explore this accumulated knowledge. Incorporating the research already undertaken toward an edition of Shakespeare's Sonnets, we sought to create a prototype knowledgebase of critical materials reflecting the scholarly engagement with Shakespeare's Sonnets from 1972 to the present day.

The first step required the acquirement of materials to add to the knowledgebase. A master list of materials was compiled through consultation with existing electronic bibliographies (such as the *MLA International Bibliography* and the *World Shakespeare Bibliography*) and standard print resources (such as the *Year's Work in English Studies*). Criteria were established to dictate which materials were to be included in the knowledgebase. To limit the scope of the experiment, materials published before 1972 (and thus considered already in Hayashi's bibliography) were excluded. It was also decided to exclude works pertaining to translations of the Sonnets, performances of the Sonnets, and non-academic discussions of the Sonnets. Monograph-length discussions of the Sonnets were also excluded on the basis that they were too unwieldy for the purposes of an experiment.

The next step was to gather the materials itemized on the master list. Although a large number of these materials were available in electronic form, and therefore much easier to collect, the various academic and commercial publishers and publication amalgamator service providers delivered the materials in different file formats. A workable standard would be required, and it was decided that regularizing all of the data into Rich Text format would preserve text formatting and relative location, and allow for any illustrations included to be embedded. Articles available only in image formats were fed through an Optical Character Recognition (OCR) application and saved in Rich Text format.

Materials unavailable in electronic form were collected, photocopied, and scanned as grayscale TIFF images. A resolution of 400 dpi was agreed upon as maintaining a balance between image clarity and file size. As a batch, the scanned images were enhanced with a negative brightness and a slightly high contrast in order to throw the type characters into relief against the page background. In addition to being stored in this format, the images were then processed through an OCR application and saved in Rich Text format.

The next step will involve applying a light common encoding structure on all of the Rich Text files and importing them into REKn. The resulting knowledgebase will be responsive to full-text electronic searches, allowing the user to uncover swiftly, for example, all references to a particular sonnet. License agreements and copyright restrictions will not allow us to make access to the knowledgebase public. However, we will be exploring a number of possible output formats that could be shared with the larger research community. Possibilities might include the use of the Sonnet knowledgebase to generate indices, concordances, or even an exhaustive annotated bibliography. For example, a dynamic index could be developed to query the full-text database and return results in the form of bibliographical citations. Since many users will come from institutions with online access to some or most of the journals, and with library access to others, these indices will serve as a valuable resource for further research.

Ideally, such endeavors will mean the reassessment of the initial exclusion criteria for knowledgebase materials. The increasing number of books published and republished in electronic format, for example, means that the inclusion of monograph-length studies of the Sonnets is no longer a task so onerous as to be prohibitive. Indeed, large-scale digitization projects such as Google Books and the Internet Archive are also making a growing number of books, both old and new, available in digital form.

## 4.5. Experiment: The REKn Crawler

We recognized that the next stages of our work would be predicated on the ability to create topic- or domain-specific knowledgebases from electronic materials. The work, then, pointed to the need for a better Internet resource discovery system, one that allowed topic-specific harvesting of Internet-based data, returning results pertinent to targeted knowledge domains, and that integrated with existing collections of materials (such as REKn) operating in existing reading systems (such as PReE), in order to take advantage of the functionality of existing tools in relation to the results. To investigate this further, we collaborated with [Iter](#), a not-for-profit partnership created to develop and support electronic resources to assist scholars studying European culture from 400 to 1700 CE (on the mandate, history, and development of [Iter](#), see [Bowen "Path"](#) and ["Building:"](#) for a more detailed report on this collaborative experiment, see [Siemens et al. "Iter"](#)).

### 4.5.1. Premises

We thought we could use technologies like Nutch and models from other more complex harvesters (such as [DataFountains](#) and the [Nalanda iVia Focused Crawler](#); see also [Mitchell](#)) to create something that would suit our purposes and be freely distributable and transportable among our several partners and their work. In using such technologies, we hoped also to explore how best to exploit representations of ontological structures found in bibliographic databases to ensure that the material returned via Internet searches was reliably on-topic.

### 4.5.2 Method

The underlying method for the prototype REKn Crawler is quite straightforward. An [Iter](#) search returns bibliographic (MARC) records, which in turn provide the metadata (such as author, title, subject) to seed a web search, the results of which are returned to the knowledgebase. In the end, the original corpus is complemented by a collection of pages from the web that are related to the same subject. While all of these web materials may not always be directly relevant, they may still be useful.

The method ensures accuracy, scalability, and utility. Accuracy is ensured insofar as the results are disambiguated by comparison against [Iter](#)'s bibliographic records—that is, via a process of domain-specific ontological structures. Scalability is ensured in that individual searches can be automatically sequenced, drawing bibliographic records from [Iter](#) one at a time to ensure that the harvester covers all parts of an identified knowledge domain. Utility is ensured because the resultant materials are drawn into the reading system and bibliographic records are created (via the original records, or using [Lemon8-XML](#)).

### 4.5.3. Workflow

From a given corpus or record set, the basic workflow for the REKn Crawler is as follows:

- *Extract keywords* from every document in a given corpus. For the prototype, we used a large MARC file from [Iter](#) as our record set and used PHP-MARC, an open source software package built in PHP that allows for manipulation and extraction of MARC records.
- *Build search strings* from the keywords extracted earlier. The following combinations were used in our experimentation: author; author and title; title; author and subject; subject.
- *Query the web* using each constructed search string. Up to fifty web page results per search are then collected and stored in a site list. Search engines that follow the OpenSearch standard can be queried from the back-end of a software application—the REKn Crawler employs this technique. OpenSearch-compatible search engines provide access to a variety of materials.
- *Harvest web pages from the site list* generated in step 3 using a web crawler. We are currently exploring implementation strategies for this stage of the project. Nutch is currently the best candidate because it is an open source web-search software package that builds on Lucene Java.

Consider the following example. A user views a document in PReE; for instance, [Edelgard E. DuBruck](#), "Changes of Taste and Audience Expectation in Fifteenth-Century Religious Drama." Viewing this document triggers the crawler, which begins crawling via the document's Iiter MARC record (record number, keywords, author, title, subject headings). Search strings are then generated from the Iiter MARC record data (in this particular instance the search strings will include: DuBruck, Edelgard E.; DuBruck, Edelgard E. Changes of Taste and Audience Expectation in Fifteenth-Century Religious Drama; DuBruck, Edelgard E. Religious drama, French; DuBruck, Edelgard E. Religious drama, French, History and criticism; Changes of Taste and Audience Expectation in Fifteenth-Century Religious Drama; Religious drama, French; Religious drama, French, History and criticism). The Crawler conducts searches with these strings and stores them for the later process of weeding out erroneous returns.

In the example given above, which took under an hour, the Crawler generated 291 unique results to add to the knowledgebase relating to the article and its subject matter. In our current development environment, the Crawler is able to harvest approximately 35,000 unique web pages in a day. We are currently experimenting with a larger seed set of 10,000 MARC records, which still amounts to a 1% subset of Iiter's bibliographical data.

#### 4.5.4. Application

The use of the REKn Crawler in conjunction with both REKn and PReE suggests some interesting applications, such as increasing the scope and size of the knowledgebase; being able to analyze the results of the Crawler's harvesting to discover document metadata and document ontology; and harvesting blogs and wikis for community knowledge on any given topic, and well beyond.

## 5. Moving into Full Prototype Development: New Directions

### 5.1. Rebuilding

Our rebuilding process was primarily driven by the questions generated from our earlier proof of concept. The proof-of-concept pointed us toward a web-based user interface to meet the needs of the research community. Building human knowledge into our application also becomes more feasible with a web environment, since we can depend on a centralized storage system and an ability to share information easily. The proof-of-concept also suggested that we rethink our document storage framework, since exponential slow-downs in full-text searching speed quickly render the tool dysfunctional in environments with millions of documents. For long-term scalability a new approach was necessary.

In order to move into full prototype development, we were first required to rebuild the foundation of both REKn and PReE applications, as outlined in detail in the previous section. To summarize:

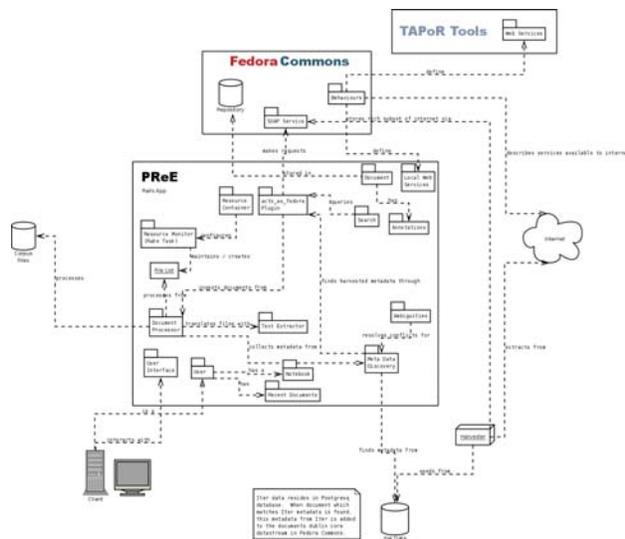
- We are rebuilding the PReE user interface. A web-based environment allows us to be agile in our development practices and to incorporate emerging ideas and visions quickly.
- The Ruby programming language has been selected as the new development platform. While it can be considered the "new kid on the block" of web-scripting languages, the benefits it offers (such as the Ruby on Rails application framework) make it an enticing choice to say the least. The use of Ruby on Rails offers a rapid prototyping environment, which cuts huge chunks of development time out of our overhead. Ruby on Rails also provides us with the ability to add "Web 2.0" user interface features to our project simply and easily.
- We are working on developing a "one-stop" administrative interface for harvesting and processing new documents. Rather than having bits and pieces scattered around, we propose to use an extensible model for adding processing abilities to our application. Once the model has been built, the processing of a new type of document will simply require the addition of a new plug-in to bring the document into the application.
- We decided to keep the relational database for application-specific data

needs (such as user info and user created content) in addition to implementing a dedicated full-text indexing engine to search both the text and the associated metadata. An application that offers time-efficient full-text searchability of documents is greatly valued by its users. To this end we decided to enlist the use of *Lucene*, the "granddaddy" of open-source full-text indexing engines. Lucene gives us fast, robust and scalable full-text searching. The *Solr* layer on top of Lucene allows us to "talk" to Lucene from any programming language we choose and give it powerful additions such as basic text analysis and the ability to identify a document uniquely. While *Fedora Commons* might prove to be a better alternative to Solr, the switch will have to wait until such time as the Fedora GSearch tool has been built into the *RubyFedora* library.

- We are working toward centralizing document processing. Until now, a different stand-alone tool processed each style of document. We are planning to pull all of these tools together in one place and to allow new tools to be added easily, with the facility for administrators to go through the process of adding new documents into the knowledgebase attached to PReE.
- We are rebuilding the interconnections between PReE and other related community tools. From metadata lookup tools to applications providing data analysis, the next development of PReE will be designed with flexibility and long-term scalability in mind.

With new development paths come new questions and concerns. For example, how would we provide consistent metadata for widely disparate sources? To address this, we are investigating the possibility of using natural language processing tools (NLP) to discover key information points within the document, and using this information to do a lookup within a robust metadata database. At the time of writing, metadata for our documents is stored inside the database structures. The documents are transformed into HTML or plain-text equivalents, which are then fed into Solr through its REST web interface. PReE uses Solr's REST API to provide full-text searching, handing off each search request to Solr and converting the search results into HTML for the browser.

Figure 7: High Level Architectural Diagram of REKn/PReE.



A high level architectural diagram (Figure 7) was created that situated the Crawler (marked 'Harvester') within the intended rebuild of REKn and PReE. As suggested by the diagram, we maintained the belief that integration with Fedora Commons was the ideal solution (see Appendix 2), but that we would have to wait until the technology allowed.

## 5.2. New Directions: Social Networking

Users are beginning to expect more from web applications than ever before. Social networking tools and the "Web 2.0" pattern of design has given web application developers many new ways of building knowledge into their applications. By adopting a web-application model for PReE, we could tie into existing social networking tools and begin to innovate with the creation of new

tools designed specifically for the professional reader. The decision to include social networking capabilities in the PReE design was based on research conducted by the Public Knowledge Project (PKP) into the reading strategies of domain-expert readers, a subset of professional readers (see [Siemens et al. "Iter"](#) and ["May Change"](#)). Like PReE, the goal for the reading tools developed by PKP was to provide access to research and scholarship and to support critical engagement with those materials. During interviews conducted by PKP and ETCL researchers, expert readers identified the ability to communicate with other researchers as an important benefit of an online reading environment. These readers also expressed interest in contextual information that would help them judge the value of an author's work. From these observations, researchers concluded that future online reading environments would need to provide the kind of communication and profile-management features currently offered by social networking tools.

Before adding social networking components to the PReE features list, we researched existing social networking tools and their use by expert readers ([Leitch et al.](#)). Based on evidence gathered during the PKP study we determined that as expert readers became adept at using online tools, they would demand a higher level of sophistication from an online reading environment. In order to respond to this increasing awareness of the potential of social networking tools for scholarly research, a successful online reading environment should integrate social networking tools in such a way that it extends the readers' existing research strategies. We identified three key strategies that readers used as part of their research: *evaluating*, *communicating*, and *managing*. Our survey found that no single social networking tool supported all three of these strategies. An environment able to facilitate all three strategies would be of immense value to the expert reader, who would not be forced to use a variety of disjointed social networking tools. Instead, he or she would be able to perform the same tasks from within the reading environment.

How could we incorporate these findings into PReE? In answering that question we were effectively reconceptualizing PReE as *social software*, "loosely defined" by [Tom Coates](#) as software that "supports, extends, or derives added value from, human social behaviour" (n.p.). If we could outline the common elements of the social networking tools we wished to incorporate, the task of combining them could be more streamlined. For [Ralph Gross and Alessandro Acquisti](#), the feature common to all social networking applications is the ability to create a user-generated *identity* (or "profile") for other users to peruse "with the intention of contacting or being contacted by others" (71). Acknowledging the importance of identity, [Judith Donath and danah boyd](#) have proposed that "a core set of assumptions" underlie all social networking applications, all of which emphasize the notion of making connections, that "there is a need for people to make more connections, that using a network of existing connections is the best way to do so, and that making this easy to do is a great benefit" (71).

### 5.2.1. Identity and Evaluation

The "Digital Footprints" report prepared by the Pew Internet and American Life Project found that "one in ten internet users have a job that requires them to self-promote or market their name online," and that "voluntarily posted text, images, audio, and video has become a cornerstone of engagement with Web 2.0 applications" to the point that "being 'findable and knowable' online is often considered an asset in participatory culture where one's personal reputation is increasingly influenced by information others encounter online" ([Madden et al. iii, 4](#)). Similar assertions have been made by other scholars: [Andreas Gergen](#) and [Alison Lee](#) suggest that one of the benefits of creating and maintaining a profile on a social networking site is the opportunity to create a "persistent and verifiable identity" (137), whereas [boyd and Nicole Ellison](#) note that "what makes social network sites unique is not that they allow individuals to meet strangers, but rather that they enable users to articulate and make visible their social networks" (n.p.).

Given the importance expert readers place on markers of authority such as credentials and past publications, it is in the individual's best interest to exert some control over his or her online identity. The ability to create and maintain an online profile as part of PReE allows users to include the kind of information expert readers look for when evaluating the value of research material.

### 5.2.2. Connections and Communication

Expert readers learn about new ideas and develop existing ones by engaging in scholarly communication with their peers and colleagues. Online, these readers participate in discussion forums, mailing lists, and use commenting tools on blogs and other social networking sites. As [Kathleen Fitzpatrick](#) observes:

Scholars operate in a range of conversations, from classroom conversations with students to conference conversations with colleagues; scholars need to have available to them not simply the library model of texts circulating amongst individual readers but also the coffee house model of public reading and debate. This interconnection of individual nodes into a collective fabric is, of course, the strength of the network, which not only physically binds individual machines but also has the ability to bring together the users of those machines, at their separate workstations, into one communal whole. (n.p.)

Likewise, [Christopher Hoadley and Peter Kilner](#) have asserted that *conversation* is the method by which information becomes knowledge; they suggest that "knowledge-building communities are a particular kind of community of practice focused on learning," where the "explicit goal [is] the development of individual and collective understanding" (32). Adopting this definition, PReE models a knowledge-building community of practice by combining content with communication through the use of social networking tools.

### 5.2.3. User and Content Management

Searching, retrieving, classifying, and organizing research material is a primary activity of professional readers. Expert readers employ a variety of strategies ranging from simple filing systems to elaborate systems of classification and storage. Reference management tools allow users to find, store, and organize research materials online. The use of *folksonomy* tagging in reference management tools can improve on a reader's existing research strategies by providing him or her with a flexible and easily accessible way of organizing research according to his or her own criteria (for the origin of the term *folksonomy* and its use to describe the practice of socially-derived content tagging, see [Vander Wal](#)). These tools also allow users to share research collections with colleagues and find material relevant to their interests in other collections. Moreover, as [Bryan Alexander](#) has observed, social bookmarking functions in a higher education context as a tool for "collaborative information discovery" (36). Alexander suggests that "finding people with related interests" through social bookmarking "can magnify one's work by learning from others or by leading to new collaborations," and that "the practice of user-created tagging can offer new perspectives on one's research, as clusters of tags reveal patterns (or absences) not immediately visible" (36). User incentives for tagging include the ability to quickly retrieve research material, to share relevant material with colleagues, and to express an opinion or make a public statement about one's interests ([Marlow et al.](#) 34-5). The planned inclusion of similar tools in PReE extends expert readers' existing management strategies by simplifying the organization process and creating new opportunities for collaborative categorization.

## 5.3. Designing the PReE Interface

When the original interface was designed for the proof of concept of REKn in .NET, very little consideration was given to further use of the code. The focus was solely on producing a down-and-dirty prototype. The decision to translate PReE from a desktop application to a web application promised a whole host of new benefits: superior flexibility in application deployment and maintenance, the ability to receive and disseminate user-generated content, and multi-platform compatibility. These new benefits, however, came with new challenges.

Migrating the application from desktop to Internet also offered us an opportunity to rethink completely the appearance and functionality of the interface. This gave us the chance to consult with prominent researchers working in the field of professional reading and designing such interfaces, as well as the opportunity to conduct our own usability surveys in order to improve accommodation for professional readers of various disciplinary

backgrounds and levels of expertise.

### 5.3.1. User Needs: Analyzing the Audience

Before embarking on a new interface design, it was pertinent to identify the features and functions that users would expect and desire from PReE. Surveys and interviews were conducted, and the results led to our distinguishing between users of PReE in terms of their backgrounds, goals, and needs. Of course, it was recognized that the usefulness of these user profiles was limited, particularly with respect to the needs of interdisciplinary users and users from less text-centric disciplines (such as Fine Arts). These limitations notwithstanding, this initial discussion allowed us to identify three general user profiles: graduate students ("students"), teaching professors ("teachers"), and research professors ("researchers").

"Student" users were characterized as coming from potentially broad disciplinary backgrounds. Their goals were to conduct self-directed research for the purposes of acquiring a thorough knowledge of a particular field; to complete their doctoral or masters' theses; and to build their scholarly reputations. Needs and desires dictated by these goals included access to citations and bibliographies; a way of assessing the impact-factor of a given article, topic, or researcher in a particular field; and a system to facilitate both formal and informal peer review of their research.

"Teacher" users were characterized as potentially belonging to broad disciplinary backgrounds (such as history) and/or specific fields (such as late medieval English military history). Their goals included recommending readings to students, undertaking self-directed research for the purpose of compiling knowledge-area bibliographies (often annotated), and writing and delivering lectures. These goals required access to citations and surveys of new and recent research in their particular field(s).

"Researcher" users were similarly characterized as potentially coming from a broad field and/or a more specific field of research expertise. Their goals included self-directed research for the purpose of building knowledge-area bibliographies (often annotated), writing and presenting conference papers, writing and delivering lectures, engaging in scholarly publication, and building and maintaining their scholarly reputations.

As a whole, these results suggested three key user requirements: the facilitation of high-level research, the facilitation of collaboration, and the achievement of recognition in their field of study. Although additional features were suggested, meeting these key requirements would be the driving force behind the design of the new PReE interface.

### 5.3.2. Design Principles, Processes, and Prototypes

A series of design principles were also agreed upon, which dictated that the interface design should focus on providing efficient ways to complete tasks (*efficiency*), on managing higher and lower priority objects (*visual balance*), on testing usability (*prototyping*), and on the ability to execute tasks rapidly in an agile work environment (*flexibility*). These principles suggested a design process of four steps. The first step was to conduct environmental scans in order to survey successful features offered by other web applications and assess their applicability for our present needs. The next step was to construct workflow sketches. The third step was to develop simple prototypes, and the fourth, to develop initial designs.

*Video 2: Design Processes of the PReE User Interface*



Environmental scans focusing on the search and display functions of existing web applications highlighted a number of useful user features. A useful feature of some applications is the suggestion of search terms to the user, either by way of a drop-down list or by auto-completion of the search string. Other applications offer "bookshelves" of saved search items, allowing their users to group items together and to tag, rate, and comment on them (Figure 8). The survey of reader and display functions similarly suggested useful features that we could implement in the PReE user interface. As outlined in more detail above (see 5.2), there is growing interest in the research application of social annotations and annotation tools as in Figure 9 (for a useful survey and assessment of existing annotation tools and their implementation in electronic editions of literary texts, see [Boot](#)). Other web applications enrich their content through the inclusion of user-contributed data, such as comments, tags, links, ratings, and other media (Figure 10). As in the original proof-of-concept, the capacity for viewing images and texts side-by-side was also expected to be included (Figure 11). As indicated in [Video 2](#) above, all of these features were included in the PReE workflow sketches, simple prototypes, and initial designs of the user interface.

Figure 8: Interface design: bookshelves.

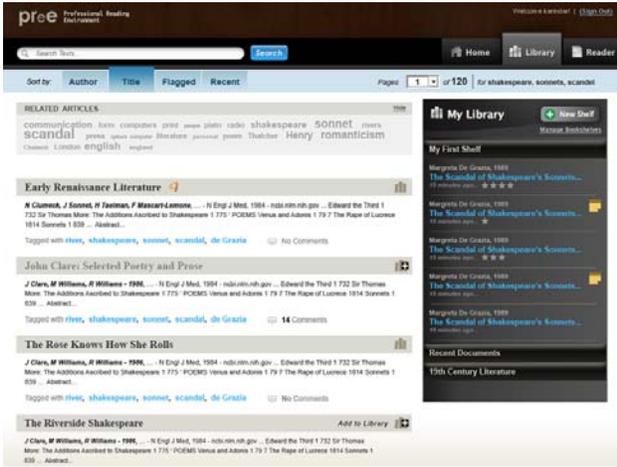


Figure 9: Interface design: annotations and bookmarks.

Interface design: annotations and bookmarks.

Figure 10: Interface design: annotations, bookmarks, and user comments.

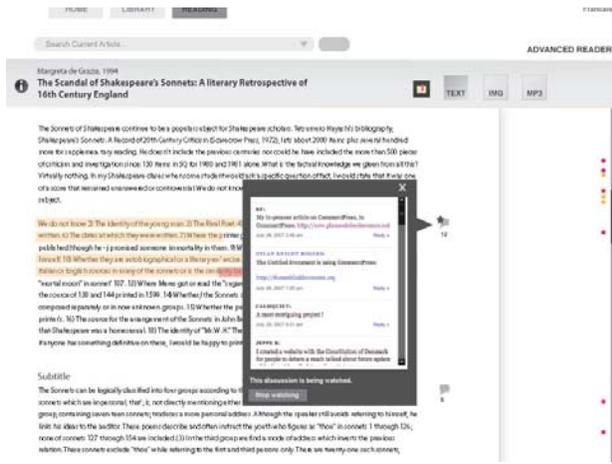


Figure 11: Interface design: side-by-side text and image display.

Interface design: side-by-side text and image display.

## 6. New Insights and Next Steps

### 6.1. Research Insights and the Humanities Model of Dissemination

While we have learned much about humanistic engagement with the technologies under consideration, we recognize also that we have gained significant experience and understanding about the nature of the work itself from a disciplinary perspective.

One unexpected insight involved the nature of where the research lies in our endeavor. Our original approach to the project was to work toward a reading environment that suited the needs of professional readers, with the belief that we understood our own needs best and could therefore contribute to the development of professional reading tools through our active participation in pertinent research processes. Conceptualizing and theorizing the foundations of and rationales for humanist tools and their features was an important part of our role, as was modeling the features and functions computationally so that it was clear that what we wished to do could be done. Indeed, we had particular success in amalgamating previously unconnected (but research-pertinent) database contents so that a researcher could speed workflow by not having to enter search terms across several unconnected databases and interfaces. By modeling these processes we were better able to understand the problems and to suggest possible solutions. From our perspective as researchers, developing the prototype that proved the concept was our primary goal—anything beyond this was more production- than research-oriented, and it was unclear to us whether production was part of our endeavor.

In the second instance, we found that the most valuable point of impact for our research work manifested in ways that our humanities disciplines could not readily understand, evaluate, and appreciate. Our research-related successes often involved (1) the identification of a key area of intervention pertaining to our larger program of research; (2) understanding this area and modeling it with the computer; (3) testing and refining the model until we achieved acceptable functionality in proof of concept; (4) delivering a conference paper on this as quickly as possible (because computational fields, their tools, and the possibilities they enable advance rapidly) and engaging in further discussions with those who were interested in carrying this work further; and either (5a) working with a partner who was interested in putting our research into production within their own work; (5b) watching others involved in adjacent programs of research implement similar features in their own work and advancing our own research in that way; or (5c) noting the adoption of our procedures without our involvement by other area stakeholders. As a progression from idea to point of impact, this is ideal in every way except one: our home disciplines in the humanities find it difficult to document this impact in professional terms. It simply does not fit the article- and book-focused publication and dissemination model favoured by humanities scholarship, and

most digital humanities venues do not integrate conference presentation and publication in a way that provides immediate publication on presentation (as is common in the sciences). As a result, work related to this project has, for the most part, been disseminated without publication, and is therefore largely unquantifiable in humanities disciplinary terms.

## 6.2. Partnerships and Collaborations

The second phase of our development of both REKn and PReE is at a crossroads. Over the course of some five years, we have been working on REKn and PReE in various ways. During this time we have presented our findings at conferences and discussed our methodology of modeling and prototyping with other research groups. The professional and pedagogical exercise of this work has been immense, driven at its core by a consistent aim to explore document-centered reading environments, and to work toward the production of a functional tool for a variety of professional readers. As with any project of this nature, our research experience has been (and continues to be) attended by successes and fraught with apparent dead-ends. However, as the preceding project narrative has made clear, even these seemingly inconclusive pursuits are in fact evidence of an active pedagogical process and a professional evolution in design and implementation—something privileged in all academic pursuit—where each step has led to a better understanding of how our overall research goals could be accomplished.

In light of the insights gained and lessons learned, our next steps are firmer and more secure, and we bring our experience to a series of very fruitful partnerships in which elements of our research are being extended in ways not initially considered. Moreover, we are incorporating our research experience into a large collaborative initiative, Implementing New Knowledge Environments (INKE), sponsored by the SSHRC Major Collaborative Research Initiatives program, as well as contributing to further developments associated with TAPoR.

Our research on interfaces, annotation, social interaction, and document-centered reading environments has also been incorporated into more focused research partnerships with groups like PKP and Synergies. Our collaboration with PKP has seen work toward the integration of professional reading tools into the PKP Open Journal Systems (OJS). As outlined briefly above, our partnership began with conducting user experience surveys to identify and assess elements of users' engagement with texts and the OJS interface (in [Siemens et al. "Iter" and "May Change"](#)). Work was then undertaken towards the identification of basic principles for an OJS interface redesign to respond to needs identified by the study; the carrying out of more precise user analysis and profiling; the design of wireframes (sketch prototypes) to emulate workflows; and consultation about technological facilitation for interaction that was imagined (including the integration of social networking technologies). These processes led to iterative computational modeling and testing, aimed at the creation of a proof-of-concept prototype. This prototype was presented to PKP in early 2008, in order that they might consider integrating it into their current development cycle—and also in more traditional research dissemination (see the list of presentations delivered in 2008 in [Appendix 1](#), in particular those presented in June). The next step of this conjoint research program is to build on earlier work carried out toward provision of a knowledgebase approach to speed professional readers' workflow through better access to pertinent critical textual resources. In turn, this new work draws on earlier and ongoing work with Iter, another of our research partners, to further develop the concept of enriched domain-specific knowledgebases, as well as ongoing research as part of a collaboration with the Transliterations and BlueSky working groups at the University of California, Santa Barbara, towards the prototyping of an interface with document-centered professional reading tools and advanced social networking capabilities.

To return to the words of [James Joyce](#) with which this article began, our experience in developing REKn and PReE thus far has shown that the errors we encountered on the way truly were "portals of discovery" (9.229). As we embark on new directions and build new partnerships and collaborations, we expect many more portals in the immediate future, and beyond.

## Appendix 1: Addresses and Presentations

This article cumulates and builds upon a series of addresses and presentations given during the developmental steps and stages, outlined below. We wish to thank the organizers of the various conferences and lectures for the valuable opportunity to present on our ongoing research, and all present for their feedback.

### 2003

Siemens, Ray. "The Dynamic Textual Edition: Underpinnings and Above." Distinguished Speaker Series. Maryland Institute for the Humanities, U of Maryland. 20 Feb. 2003. Address.

———. "Humanities Computing and the Scholarship of Integration: Modelling Disciplinary Interaction in Literary Studies through Humanities Computing." Research Showcase. Malaspina U-C, Nanaimo. 17 Apr. 2003. Address.

———. "Toward a Computing Environment for the Literary Studies Reader." Invited Lecture. Sheffield Hallam U, Sheffield. 17 Oct. 2003. Address.

———. "Imagining the Printed Book in an Electronic Age." Lansdowne Lecture in Humanities Computing. U of Victoria, Victoria. 16 Nov. 2003. Address.

———. "Algorithm and Interface in the Electronic Scholarly Edition." Theorizing the Interface. MLA Annual Convention. Manchester Grand Hyatt, San Diego. 29 Dec. 2003. Address.

———, and William R. Bowen. "The Role of Text Analysis in the Creation of a Knowledge Base: Preliminary Thoughts on the Future of Iter: Gateway to the Middle Ages and Renaissance." CaSTA: The Canadian Symposium on Text Analysis Research. U of Victoria, Victoria. 14 Nov. 2003. Address.

### 2004

Siemens, Ray. "Pragmatic Notes Toward a Dynamic Scholarly Edition." Seminar Series. Centre for Computational Studies, U of Kentucky. 16 Sep. 2004. Address.

———. "Modelling Humanistic Activity in the Electronic Scholarly Edition." (The Face of Text, CaSTA: The Third Canadian Symposium on Text Analysis Research, McMaster U. 21 Nov. 2004. Address.

———. "Access to Knowledge." Technology, Culture, Aesthetics: Hypermedia and the Changing Nature of Knowledge. New Media and Culture Network Workshop Series. U of British Columbia, Vancouver. 6 May 2004. Address.

———, Elaine Toms, Geoffrey Rockwell, Stéfan Sinclair, and Lynne Siemens. "The Humanities Scholar in the Twenty-First Century: How Research is Done and What Support is Needed." ALLC/ACH Joint International Conference. Göteborgs U, Göteborg. 16 Jun. 2004. Address.

———, Elaine Toms, Geoffrey Rockwell, Stéfan Sinclair, and Lynne Siemens. "Modelling the Humanities Scholar at Work." The Face of Text. CaSTA: Canadian Symposium on Text Analysis Research. McMaster U, Hamilton. 19 Nov. 2004. Address.

### 2005

Siemens, Ray. "Imagining the Printed Book and Manuscript in an Electronic Age." Form and Functionality: Human-Computer Interface and Interaction Issues for the Electronic Book. Annual Meeting of the Consortium for Computers in the Humanities (COCH-COSH), Congress in the Canadian Federation of Humanities and Social Sciences. U of Western Ontario, London. 30 May 2005. Address.

———. "Humanities Computing and the Modeling of Humanistic Activity." Invited Lecture. Sheffield Hallam U, Sheffield. 9 Sep. 2005. Address.

———. "Electronic Scholarly Editions and Models of Humanistic Activity."

Summit on Digital Tools for the Humanities. U of Virginia, Charlottesville. 28 Sep. 2005. Address.

———, and Chris Gaudet. "A Knowledgebase Toward an Electronic Edition of Shakespeare's Sonnets." CaSTA 2005: The Fourth Canadian Symposium on Text Analysis Research. U of Alberta, Edmonton. 4 Oct. 2005. Address.

## 2006

Siemens, Ray. "Modelling the <Textual> Activity of the Humanist." The Computer: The Once and Future Medium for the Humanities and Social Sciences (CFHSS, CCHC, SDH-SEMI). York U, Toronto. 30 May 2006. Address.

———. "The Renaissance English Knowledgebase (REKn) and its Professional Reading Environment (PRE): An Overview." Teaching Digital Texts, English Subject Centre/Methods Network. King's College, London. 16 Jul. 2006. Address.

———. "Knowledge Management and Textual Cultures? Work toward the Renaissance English Knowledgebase (REKn) and its Professional Reading Environment (PRE)." CASTA 2006: The Breadth of Text. U of New Brunswick, Fredericton. 14 Oct. 2006. Address.

———. "Modelling Scholarly Practices with a Renaissance English Knowledgebase." Contexts for Electronic Editing, MLA Annual Meeting. Philadelphia Marriott, Philadelphia. 30 Dec. 2006. Address.

———, Eric Haswell, Gerry Watson, Alastair McColl, and Karin Armstrong. "Integrating Tools into Professional Academic Processes: A First Look at the Renaissance English Knowledgebase (REKn)." Bringing Text Alive: The Future of Scholarship, Pedagogy, and Electronic Publication. Text Creation Partnership Conference. U of Michigan, Ann Arbor. 15 Sep. 2006. Address.

———, and Alastair McColl. "Learning Curves and Tempered Results: Toward a Renaissance English Knowledgebase (REKn)." What To Do with a Million Books Chicago Colloquium on Digital Humanities and Computer Science. U of Chicago, Chicago. 8 Nov. 2006. Address.

———, John Willinsky, and Analisa Blake. "Giving Them a Reason to Read Online: Reading Tools for Humanities Scholars." Digital Humanities 2006. U Paris-Sorbonne, Paris. 8 Jul. 2006. Address.

## 2007

Bowen, William R., and Ray Siemens. "Iter as Knowledgebase." New Technologies and Renaissance Studies III: Catalogues of Knowledge, RSA Annual Meeting. The New Radisson Hotel, Miami. 23 Mar. 2007. Address.

Siemens, Ray "A Renaissance English Knowledgebase (REKn) in a Professional Reading Environment (PRE)." New Technologies and Renaissance Studies I: The Early Modern Codex in Contemporary Electronic Context, RSA Annual Meeting. The New Radisson Hotel, Miami. 23 Mar. 2007. Address.

———. "Prototyping a 'Knowledgebase' Approach to Online Materials, to Meet the Demands of Professional Readers in the Humanities." Digital Humanities: Practice, Methodology, Pedagogy. Centre for Studies in Print and Media Cultures Symposium. Simon Fraser U, Vancouver. 3 May 2007. Address.

———. "Modeling and Knowledge [Re]presentation: As a Context for the Contemporary Editor of Earlier Textual Materials." Round Table: Accessing, Organizing, and Analyzing Digital Evidence. Congress of the Canadian Federation of Humanities and Social Sciences. U of Saskatchewan, Saskatoon. 30 May 2007. Address.

———. "A Professional Reading Environment, Modeled for a Renaissance English Knowledgebase." Public Knowledge Project Scholarly Publication Conference. Simon Fraser U, Vancouver. 12 Jul. 2007. Address.

———. "Working with a Renaissance English Knowledgebase (REKn) in a Professional Reading Environment (PRE)." Early Renaissance Literature. IAUPE Triennial Conference. Lunds U, Lund. 6 Aug. 2007. Address.

———. "A Knowledgebase Approach to Professional Reading." Models of Partnership in Digital Research Colloquium. Sheffield Hallam U, Sheffield. 3 Sep. 2007. Address.

———. "A Scholarly Reading Interface for a Renaissance English Knowledgebase." ACCESS 2007: TechTonic OnTologies. U of Victoria, Victoria. 11 Oct. 2007. Address.

———. "Prototyping a 'Knowledgebase' Approach to Texts and Secondary Resources in Renaissance Studies: The Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE)." Renaissance English Text Society Josephine A. Roberts Forum, MLA Annual Meeting. Hyatt Regency, Chicago. 29 Dec. 2007. Address.

———, Mike Elkind, and Karin Armstrong. "Building One to Throw Away, Toward the One We'll Keep: Next Steps for the Renaissance English Knowledgebase and Professional Reading Environment." Chicago Colloquium on Digital Humanities. Northwestern U, Chicago. 22 Oct. 2007. Poster presentation.

## 2008

Leitch, Cara, Ray Siemens, James Dixon, Mike Elkind, Angelsea Saby, and Karin Armstrong. "Social Networking and Online Collaborative Research with REKn and PReE." Society for Digital Humanities/Société pour l'étude des médias interactifs, Congress of the Canadian Federation of Humanities and Social Sciences. U of British Columbia, Vancouver. 3 Jun. 2008. Poster presentation.

———, Ray Siemens, James Dixon, Mike Elkind, Angelsea Saby, and Karin Armstrong. "Social Networking and Online Collaborative Research with REKn and PReE." Digital Humanities 2008. U of Oulu, Oulu. 27 Jun. 2008. Poster presentation.

Siemens, Ray. "Consolidated Knowledge-bases and the Promise of Text Analysis in the Short Term, and Beyond: TAPoR, Synergies, CRKN." Building Cyberinfrastructure for the Humanities. Society for Digital Humanities/Société pour l'étude des médias interactifs, Congress of the Canadian Federation of Humanities and Social Sciences. U of British Columbia, Vancouver. 2 Jun. 2008. Address.

———. "The Renaissance English Knowledgebase (REKn) Crawler for a Professional Reading Environment (PReE)." Society for Digital Humanities/Société pour l'étude des médias interactifs, Congress of the Canadian Federation of Humanities and Social Sciences. U of British Columbia, Vancouver. 3 Jun. 2008. Poster presentation.

———, Cara Leitch, Angelsea Saby, James Dixon, Mike Elkind, and Karin Armstrong. "Interface Design Principles for a Professional Reading Environment (PReE)." Society for Digital Humanities/Société pour l'étude des médias interactifs, Congress of the Canadian Federation of Humanities and Social Sciences. U of British Columbia, Vancouver. 2 Jun. 2008. Poster presentation.

———, Rachel Gold, James Dixon, and Karin Armstrong. "Generating Topic-Specific, Individual Knowledge-bases from Internet Resources: REKn/PReE Crawler for Professional Reading Environments." CaSTA: Canadian Symposium on Text Analysis. U of Saskatchewan, Saskatoon. 29 Aug. 2008. Address.

## 2009

Hirsch, Brett D., and Ray Siemens. "Prototyping the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE): Past, Present, Futures." Society for Digital Humanities/Société pour l'étude des médias interactifs, Congress of the Canadian Federation of Humanities and Social Sciences. Carleton U, Ottawa. 27 May 2009. Address.

## Appendix 2: Prototype Development Platform

### Ruby on Rails

Ruby on Rails is a development framework that was created for use with the Ruby language. Its code generation tools and scaffolding make it ideal for fast and flexible prototype development. Developing simple features for a web application follows a very specific process in Ruby on Rails, and, since Rails handles a lot of the steps along the way for you, you save time. For example, forms that allow you to modify data in a database can be auto-generated by Rails. In PHP, by contrast, there are similar tools and frameworks that aid in rapid development, but there are too many to choose from and often time is cannibalized by the hunt for the most appropriate tools.

Ruby on Rails was chosen as our development platform because it offered a rapid prototyping environment, with a structured framework for documentation. This would save us time and provide us with a solid structure for our code base. In addition, because Ruby is a very object-orientated language, it would encourage the use of good software design principles. Ruby on Rails also provides an excellent basis for a Representational State Transfer (REST) API. REST API calls are special cases of regular HTTP URL requests, with an XML data payload defining the contents of the request. By documenting and publishing our REST API any partners that we wished to integrate and collaborate with would be able to make direct use of our database, using tools able to understand RESTful communications protocols.

## Zotero

Zotero is an open source tool that automatically extracts bibliographic information from websites and organizes this data with the click of a button. As a Mozilla Firefox plug-in, Zotero is limited to users of that browser that have installed the plug-in. Even so, Zotero offers a quick and easy way to demonstrate some flashy reading tools for the next phase of development. For example, if PReE formatted bibliographic data for documents in a way that was easy for Zotero to parse, we could demonstrate this functionality.

## eXist

eXist is an open source database management system built on Java and XML technology. It stores XML data according to the XML data model and features efficient, index-based XQuery processing. Initially, an eXist database seemed like the best choice for storing and indexing XML data for PReE, so we developed an API to assist with interaction of the eXist database with Ruby and released it as open source (eXist XML-RPC API). Since then, we decided to move towards Fedora Commons and Fedora GSearch for storing and indexing our data. The advantages of using Fedora Commons (outlined below) seem to outweigh the advantages of using eXist.

## Solr

Solr is an open source enterprise search server based on the Lucene Java search library, with XML/HTTP and JSON APIs, hit highlighting, faceted search, caching, replication, and a web administration interface. It runs in a Java servlet container such as Tomcat. Because of its search capabilities, Solr was integrated into the Ruby on Rails development of PReE. In the current version of PReE, metadata from the Shakespeare Sonnets REKn subset is injected into Solr's index. When a user searches for a document, Solr looks through the metadata for each document in PReE and identifies documents that seem to match the users query. Solr does a very fast job of this, as it is based on the quick Lucene Java search library. Although we were keen to investigate using Fedora GSearch for PReE, this tool has not yet been easily incorporated into the Ruby on Rails development framework.

## Fedora Commons, Fedora GSearch, and RubyFedora

Fedora Commons provides sustainable technologies to create, manage, publish, share and preserve digital content as a basis for intellectual, organizational, scientific, and cultural heritage by bringing two communities together. Fedora Commons has number of advantages over using a Solr server for REKn/PReE. Essentially, documents in Fedora Commons are represented by objects in the system, and these objects have many unique abilities that will be of significant benefit to PReE:

- Fedora objects have a special data stream reserved for Dublin Core metadata. When full-text documents from REKn/PReE are fed into Fedora Commons, the document metadata can be attached to the same objects as the full-text.
- A Fedora object can have any number of data streams of many different types. If we wished to attach a series of images to a documents object, for example, this is easily accomplished in Fedora. We could also, for example, attach the original RTF file of a sonnet to a documents object. The possibilities here are endless.
- Fedora objects data streams can be versioned. This ability would allow us to display different versions of the same document should we wish to implement this feature.

Fedora GSearch allows for searching these objects using Solr. We made contact with the developers of RubyFedora, a code library to make integration between Ruby on Rails and Fedora Commons much easier. RubyFedora would assist in incorporating Fedora Commons into PReE, but progress was cut short when it was realized that the library had not yet included the Fedora GSearch tool. Including the feature to search for documents inside a Fedora Commons repository would take a lot of time if we were forced to code the interaction with Fedora Commons and Fedora GSearch ourselves. Until such times as the Fedora GSearch tool has been built into the RubyFedora library, Fedora Commons would not be used as a replacement for Solr as a document index for PReE.

---

## Works Cited

Alexander, Bryan. "Web 2.0: A New Wave of Innovation for Teaching and Learning?" *Educause Review* 41.2 (2006): 32-44. Print.

Austin, David. "How Google Finds Your Needle in the Web's Haystack." *Feature Column*. American Mathematical Society. Dec. 2006. Web. 24 Apr. 2009. <http://www.ams.org/featurecolumn/archive/pagerank.html>.

Bolton, Whitney. "The Bard in Bits: Electronic Editions of Shakespeare and Programs to Analyze Them." *Computers and the Humanities* 24.4 (1990): 275-87. Print.

Boot, Peter. "Mesotext: Digitised Emblems, Modelled Annotations and Humanities Scholarship." Diss. U Utrecht, 2009. Print.

boyd, danah. "The Significance of Social Software." *BlogTalks Reloaded: Social Software Research and Cases*. Ed. Thomas N. Burg and Jan Schmidt. Norderstedt: Books on Demand, 2007. 15-30. Print.

———, and Nicole B. Ellison. "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication* 13.1 (2007): n. p. Web. 24 Apr. 2009.

Bowen, William R. "Iter: Building an Effective Knowledge Base." *New Technologies and Renaissance Studies*. Ed. William R. Bowen and Ray Siemens. *New Technologies in Medieval and Renaissance Studies* 1. Toronto and Tempe: Iter and Arizona Center for Medieval and Renaissance Studies, 2008. 101-9. Print.

———. "Iter: Where Does the Path Lead?" *Early Modern Literary Studies* 5.3 (2000): 2.1-26. Web. 24 Apr. 2009. <http://extra.shu.ac.uk/emls/Q5-3/bowiter.html>.

Brown, Susan, Stan Ruecker, Jeffrey Antoniuk, Sharon Balasz, Patricia Clements, and Isobel Grundy. "Designing Rich-Prospect Access to a Feminist Literary History." *Women Writing and Reading* 2.1 (2007): 12-17. Print.

Canadian Research Knowledge Network / Réseau Canadien de Documentation pour la Recherche (CRKN/RCDR). Web. 24 Apr. 2009. <http://researchknowledge.ca/>.

Coates, Tom. "An Addendum to a Definition of Social Software." *Plasticbag.org*. 5 Jan. 2005. Web. 24 Apr. 2009. <http://www.plasticbag.org/archives/2005/01>

[/an addendum to a definition of social software/](#).

Collex. NINES. Web. 24 Apr. 2009. <http://www.collex.org/>.

Data Fountains. iVia Project. U of California, Riverside. Web. 24 Apr. 2009. <http://datafountains.ucr.edu/>.

De Grazia, Margreta, and Peter Stallybrass. "The Materiality of the Shakespearean Text." *Shakespeare Quarterly* 44 (1993): 255-83. Print.

Delany, Paul. "Virtual Universities and the Death of Distance." *TEXT t Technology* 7 (1997): 49-64. Print.

Donath, Judith, and danah boyd. "Public Displays of Connection." *BT Technology Journal* 22.4 (2004): 71-82. Print.

Drucker, Johanna, and Geoffrey Rockwell. "Introduction: Reflections on the Ivanhoe Game." *TEXT Technology* 12.2 (2003): vii-xviii. Print.

DuBruck, Edelgard E. "Changes of Taste and Audience Expectation in Fifteenth-Century Religious Drama." *Fifteenth-Century Studies* 6 (1983): 59-91. Print.

Early Modern English Dictionaries Database. Ed. Ian Lancashire. U of Toronto. Web. 24 Apr. 2009. <http://www.chass.utoronto.ca/~ian/emedd.html>.

Erickson, Peter. "Rewriting the Renaissance, Rewriting Ourselves." *Shakespeare Quarterly* 38 (1987): 327-37. Print.

eXist. Wolfgang Meier *et al.* Web. 24 Apr. 2009. <http://exist.sourceforge.net>.

eXist XML-RPC API. Mike Elkind and James Dixon. Web. 24 Apr. 2009. <http://www.rubyforge.org/projects/exist-xml-rpc/>.

Faulhaber, Charles B. "Textual Criticism in the 21st Century." *Romance Philology* 45 (1991): 123-48. Print.

Fedora. Fedora Commons. Web. 24 Apr. 2009. <http://www.fedora-commons.org/>.

Fitzpatrick, Kathleen. "CommentPress: New (Social) Structures for New (Networked) Texts." *Journal of Electronic Publishing* 10.3 (2007): n. p. Web. 24 Apr. 2009. <http://dx.doi.org/10.3998/3336451.0010.305>.

Fortier, Paul. "Babies, Bathwater, and the Study of Literature." *Computers and the Humanities* 27 (1993-94): 375-85. Print.

Girgensohn, Andreas, and Alison Lee. "Making Web Sites Be Places for Social Interaction." *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*. New York: ACM, 2002. 136-45. Print.

Google Book Search. Google. Web. 24 Apr. 2009. <http://books.google.com/>.

Greetham, D. C. *Theories of the Text*. Oxford: Oxford UP, 1999. Print.

Gross, Ralph, and Alessandro Acquisti. "Information Revelation and Privacy in Online Social Networks." *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*. New York: ACM, 2005. 71-80. Print.

Guillory, John. "The Ethical Practice of Modernity: The Example of Reading." *The Turn to Ethics*. Ed. Marjorie Garber, Beatrice Hanssen, and Rebecca L. Walkowitz. New York: Routledge, 2000. 29-46. Print.

Hall, Kim F. "About This Volume." *Othello: Texts and Contexts*. Ed. Kim F. Hall. New York: Bedford/St. Martin's, 2007. vii-xii. Print.

Hayashi, Tetsumaro. *Shakespeare's Sonnets: A Record of Twentieth-Century Criticism*. Metuchen: Scarecrow P, 1972. Print.

Hoadley, Christopher M., and Peter G. Kilner. "Using Technology to Transform Communities of Practice into Knowledge-Building Communities." *SIGGROUP Bulletin* 25.1 (2005): 31-40. Print.

Hockey, Susan. *Electronic Texts in the Humanities: Principles and Practice*. Oxford: Oxford UP, 2000. Print.

Howard, Jean E. "The New Historicism in Renaissance Studies." *English Literary Renaissance* 16 (1986): 13-43. Print.

*Internet Shakespeare Editions*. Coordinating Ed. Michael Best. U of Victoria. Web. 24 Apr. 2009. <http://internetsakespeare.uvic.ca/>.

*Iter*. Renaissance Society of America, U of Toronto Centre for Reformation and Renaissance Studies, Arizona Center for Medieval and Renaissance Studies. Web. 24 Apr. 2009. <http://www.itergateway.org/>.

Joyce, James. *Ulysses*. Ed. Hans Walter Gabler. New York: Random House, 1986. Print.

*Juxta*. NINES. Web. 24 Apr. 2009. <http://www.juxta.org/>.

Lancashire, Ian. "Bilingual Dictionaries in an English Renaissance Knowledge Base." *Historical Dictionary Databases*. Ed. T. R. Wooldridge. *CCH Working Papers* 2 (1992): 69-88. Print.

———. "Computer Tools for Cognitive Stylistics." *From Information to Knowledge: Conceptual and Content Analysis by Computer*. Ed. Ephraim Nissan and Klaus M. Schmidt. Oxford: Intellect, 1995. 28-47. Print.

———. "Working with Texts." IBM Academic Computing Conference. Anaheim, California. June 1989. Address.

Leitch, Cara, Ray Siemens, James Dixon, Mike Elkink, Angelsea Saby, and Karin Armstrong. "Social Networking and Online Collaborative Research with REKn and PReE." Society for Digital Humanities/Société pour l'étude des médias interactifs, Congress of the Canadian Federation of Humanities and Social Sciences. U of British Columbia, Vancouver. 3 Jun. 2008. Poster presentation.

*Lemon8-XML*. Public Knowledge Project. U of British Columbia, Stanford U, and Simon Fraser U. Web. 24 Apr. 2009. <http://pkp.sfu.ca/lemon8>.

*Lexicons of Early Modern English*. Ed. Ian Lancashire. U of Toronto Library and U of Toronto P. Web. 24 Apr. 2009. <http://leme.library.utoronto.ca/>.

*Literature Online*. Chadwyck-Healey Literature Online. ProQuest. Web. 24 Apr. 2009. <http://lion.chadwyck.com/>.

*Lucene*. Apache Software Foundation. Web. 24 Apr. 2009. <http://lucene.apache.org/>.

Machan, Tim William. "Late Middle English Texts and the Higher and Lower Criticisms." *Medieval Literature: Texts and Interpretation*. Ed. Tim William Machan. *Medieval and Renaissance Texts and Studies* 79. Binghamton: Center for Medieval and Renaissance Studies, 1991. 3-16. Print.

Madden, Mary, Susannah Fox, Aaron Smith, and Jessica Vitak. "Digital Footprints: Online Identity Management and Search in the Age of Transparency." *Pew Internet and American Life Project*. 16 Dec. 2007. Web. 24 Apr. 2009. <http://pewinternet.org/Reports/2007/Digital-Footprints.aspx>.

*Many Eyes*. IBM Collaborative User Experience Research Group, Visual Communication Lab. Web. 24 Apr. 2009. <http://manyeyes.alphaworks.ibm.com/manyeyes/>.

Marcus, Leah S. *Unediting the Renaissance: Shakespeare, Marlowe, Milton*. New York: Routledge, 1996. Print.

Marlow, Cameron, Mor Naaman, danah boyd, and Marc Davis. "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read." *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*. New York: ACM, 2006. 31-40. Print.

McCarty, Willard. "Modeling: A Study in Words and Meanings." *A Companion to Digital Humanities*. Ed. Susan Schreibman, Ray Siemens, and John Unsworth. Malden: Blackwell, 2004. 257-70. Print.

———. "Knowing . . . : Modeling in Literary Studies." *A Companion to Digital Literary Studies*. Ed. Ray Siemens and Susan Schreibman. Malden: Blackwell,

2008. 391-401. Print.

———. "What is Humanities Computing? Toward a Definition of the Field." Address. Reed College, Portland. Mar. 1998. Web. 24 Apr. 2009. [http://staff.cch.kcl.ac.uk/~wmccarty/essays/McCarty\\_What\\_is\\_humanities\\_computing.pdf](http://staff.cch.kcl.ac.uk/~wmccarty/essays/McCarty_What_is_humanities_computing.pdf).

McLeod, Randall. "Information on Intformation." *Text* 5 (1991): 240-81. Print.

———. "UnEditing Shakespeare." *Sub-Stance* 33-34 (1982): 26-55. Print.

McGann, Jerome J. *A Critique of Modern Textual Criticism*. Chicago: U of Chicago P, 1983. Print.

———, and Johanna Drucker. "The Ivanhoe Game: An Introduction." 2000-1. Web. 24 Apr. 2009. <http://jefferson.village.virginia.edu/~jjm2f/old/IGamehtm.html>.

———, and Lisa Samuels. "Deformance and Interpretation." *New Literary History* 30 (1999): 25-56. Print.

McKenzie, D. F. *Bibliography and the Sociology of Texts*. London: British Library, 1986. Print.

Miall, David S. "The Library versus the Internet: Literary Studies Under Siege?" *PMLA* 116 (2001): 1405-14. Print.

*Michigan Early Modern English Materials*. Eds. Richard W. Bailey, Jay L. Robinson, James W. Downer, and Patricia V. Lehman. U of Michigan. Web. 24 Apr. 2009. <http://quod.lib.umich.edu/m/memem/>.

Mitchell, Steve. "Machine-Assisted Metadata Generation and New Resource Discovery: Software and Services." *First Monday* 11.8 (2006): n.p. Web. 24 Apr. 2009. [http://firstmonday.org/issues/issue11\\_8/mitchell/](http://firstmonday.org/issues/issue11_8/mitchell/).

Metadata Offer New Knowledge (MONK) Project. Web. 24 Apr. 2009. <http://monkproject.org/>.

Mueller, Martin. "The Nameless Shakespeare." *TEXT Technology* 14.1 (2005): 61-70. Print.

*Nalanda iVia Focused Crawler*. iVia Project. U of California, Riverside. Web. 24 Apr. 2009. <http://ivia.ucr.edu/projects/Nalanda>.

*Nutch*. Apache Software Foundation. Web. 24 Apr. 2009. <http://lucene.apache.org/nutch/>.

*Open Access Text Archive*. Internet Archive. Web. 24 Apr. 2009. <http://www.archive.org/details/texts>.

*Open Journal Systems*. Public Knowledge Project. U of British Columbia, Stanford U, and Simon Fraser U. Web. 24 Apr. 2009. <http://pkp.sfu.ca/ojs>.

*Open Monograph Press*. Public Knowledge Project. U of British Columbia, Stanford U, and Simon Fraser U. Web. 24 Apr. 2009. <http://pkp.sfu.ca/omp>.

*Oxford English Dictionary Online*. Oxford UP. Web. 24 Apr. 2009. <http://dictionary.oed.com/>.

*Oxford Text Archive*. Oxford University Computing Services, Oxford U. Web. 24 Apr. 2009. <http://www.ota.ox.ac.uk/>.

Pechter, Edward. "The New Historicism and Its Discontents: Politicizing Renaissance Drama." *PMLA* 102 (1987): 292-302. Print.

*PostgreSQL*. PostgreSQL Global Development Group. Web. 24 Apr. 2009. <http://www.postgresql.org/>.

*Public Knowledge Project*. U of British Columbia, Stanford U, and Simon Fraser U. Web. 24 Apr. 2009. <http://pkp.sfu.ca/>.

Remley, Paul. "Mary Shelton and Her Tudor Literary Milieu." *Rethinking the Henrician Era: Essays on Early Tudor Texts and Contexts*. Ed. Peter C. Herman. Urbana: U of Illinois P, 1994. 40-77. Print.

Richardson, David A., and Michael Neuman [with David A. Bank, Jonquil Bevan, Lou Burnard, Thomas N. Corns, Michael Crump, R. J. Fehrenback, Alistair Fox, Roy Flannagan, S. K. Heniger Jr., Arthur F. Kinney, Ian Lancashire, George M. Logan, Willard McCarty, Louis T. Milic, Barbara Mowat, Joachim Neuhaus, Michael Neuman, Henry Snyder, Frank Tompa, and Greg Waite]. "Application for NEH Funding: A Planning Conference for a Renaissance Knowledge Base." Funding Application, 1990. Print.

Rockwell, Geoffrey. "Is Humanities Computing an Academic Discipline?" Humanities Computing Seminar. U of Virginia, Charlottesville. Address. 19 Nov. 1999. Web. 24 Apr. 2009. <http://www.iath.virginia.edu/hcs/rockwell.html>.

*Ruby on Rails*. David Heinemeier Hansson. Web. 24 Apr. 2009. <http://www.rubyonrails.org/>.

*RubyFedora*. MediaShelf. Web. 24 Apr. 2009. <http://yourmediashelf.com/rubyfedora/>.

Ruecker, Stan. "The Electronic Book Table of Contents as a Research Tool." Congress of the Humanities and Social Sciences: Consortium for Computers in the Humanities / Consortium pour Ordinateurs en Sciences Humaines (COCH/COSH) Annual Conference. U of Western Ontario, London. 30 May 2005. Address.

———, Milena Radzikowska, Susan Brown, Thomas M. Nelson, Isobel Grundy, Patricia Clements, Sharon Balasz, Jeff Antoniuk, and Stéfan Sinclair. "The Dynamic Table of Contents: Extending a Venerable List in a Digital Context." The Potential and Limitations of a List: An International Transdisciplinary Workshop. Prague, Czech Republic. Nov. 2007. Address.

Schreibman, Susan. "Computer-Mediated Texts and Textuality: Theory and Practice." *Computers and the Humanities* 36 (2002): 283-93. Print.

*Shakespeare Database Project*. Dir. H. Joachim Neuhaus. Westfälische Wilhelms-U, Münster. Web. 24 Apr. 2009. <http://www.shkspr.uni-muenster.de/>.

Siemens, Ray. "Text Analysis and the Dynamic Edition? Some Concerns with an Algorithmic Approach in the Electronic Scholarly Edition." *TEXT Technology* 14.1 (2005): 91-98. Print.

———. "Unediting and Non-Editions: The Death of Distance, the Notion of Navigation, and New Acts of Editing in the Electronic Medium." *Anglia* 119.3 (2001): 423-55. Print.

———, and Cara Leitch. "Editing the Early Modern Miscellany: Modeling and Knowledge [Re]Presentation as a Context for the Contemporary Editor." *New Ways of Looking at Old Texts IV*. Ed. Michael Denbo. Tempe: Arizona Center for Medieval and Renaissance Studies, 2008. 115-30. Print.

———, and Christian Vandendorpe. "Canadian Humanities Computing and Emerging Mind Technologies." *Mind Technologies: Humanities Computing and the Canadian Academic Community*. Ed. Ray Siemens and David Moorman. Calgary: U of Calgary P, 2006. xi-xxiii. Print.

———, William R. Bowen, Jessica Natale, Karin Armstrong, Alastair McColl, and Greg Newton. "Iter Database: Research Report on the Inclusion of Electronic Resources." Whitepaper. Electronic Textual Cultures Laboratory, University of Victoria. 2006. Web. 24 Apr. 2009. <http://etcl-dev.uvic.ca/public/iter-report/>.

———, Johanne Paquette, Karin Armstrong, Cara Leitch, Brett D. Hirsch, and Eric Haswell. "Drawing Networks in the Devonshire Manuscript (BL Add MS 17492): Toward Visualizing a Writing Community's Shared Apprenticeship, Social Valuation, and Self-Validation." *Digital Studies/Le Champ Numérique*. In Press.

———, John Willinsky, Analisa Blake, Karin Armstrong, Lindsay Colahan, and Greg Newton. "A Study of Professional Reading Tools for Computing Humanists." Report. Electronic Textual Cultures Laboratory, U of Victoria. May 2006. Web. 24 Apr. 2009. [http://etcl-dev.uvic.ca/public/pkp\\_report/](http://etcl-dev.uvic.ca/public/pkp_report/).

———, John Willinsky, Cara Leitch, and Analisa Blake. "It May Change My Understanding of the Field: Understanding Reader Tools for Scholars and

Professional Readers." *Digital Humanities Quarterly*. In Press.

Sinclair, Stéfan, and Geoffrey Rockwell. "Reading Tools, or Text Analysis Tools as Objects of Interpretation." *Digital Humanities 2007*. U of Illinois at Urbana-Champaign, Illinois. June 2007. Address.

*Solr*. Apache Software Foundation. Web. 24 Apr. 2009. <http://lucene.apache.org/solr/>.

Southall, Raymond. "The Devonshire Manuscript Collection of Early Tudor Poetry, 1532–41." *Review of English Studies*, new series 15 (1964): 142-50. Print.

———. *The Courtly Maker: An Essay on the Poetry of Wyatt and His Contemporaries*. Oxford: Blackwell, 1964. Print.

Sutherland, Kathryn. "Introduction." *Electronic Text: Investigations in Method and Theory*. Ed. Kathryn Sutherland. Oxford: Oxford UP, 1997. 1-18. Print.

———. "Revised Relations? Material Text, Immaterial Text, and the Electronic Environment." *Text* 11 (1998): 17-30. Print.

*Synergies*. Web. 24 Apr. 2009. <http://www.synergiescanada.org/>.

Tanselle, G. Thomas. "Textual Criticism and Literary Sociology." *Studies in Bibliography* 44 (1991): 83-143. Print.

*TAPoR Tools*. Text Analysis Portal for Research (TAPoR) Project. Web. 24 Apr. 2009. <http://portal.tapor.ca/>.

*Textbase of Early Tudor English*. Eds. Alistair Fox and Greg Waite. U of Otago. Web. 24 Apr. 2009. <http://www.hlm.co.nz/tudortexts/>.

Unsworth, John. "Knowledge Representation in Humanities Computing." eHumanities NEH Lecture Series on Technology and the Humanities. Washington. Address. Apr. 2001. Web. 24 Apr. 2009. <http://www.iath.virginia.edu/~jmu2m/KR/KRinHC.html>.

———. "Documenting the Reinvention of Text: The Importance of Failure." *Journal of Electronic Publishing* 3.2 (1997): n. p. Web. 24 Apr. 2009. <http://dx.doi.org/10.3998/3336451.0003.201>.

———. "Scholarly Primitives." *Humanities Computing: Formal Methods, Experimental Practice*. King's College, London. Address. May 2000. Web. 24 Apr. 2009. <http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html>.

Vander Wal, Thomas. "Folksonomy Coinage and Definition." *Off the Top*. 2 Feb. 2007. Web. 24 Apr. 2009. <http://www.vanderwal.net/folksonomy.html>.

Warwick, Claire. "Print Scholarship and Digital Resources." *A Companion to Digital Humanities*. Ed. Susan Schreibman, Ray Siemens, and John Unsworth. Malden: Blackwell, 2004. 366-82. Print.

*Women Writers Project*. Women Writers Project. Brown U. Web. 24 Apr. 2009. <http://www.wwp.brown.edu/>.

*Zotero*. Center for History and New Media, George Mason U. Web. 24 Apr. 2009. <http://www.zotero.org/>.



This work is licensed under a [Creative Commons Attribution 3.0 License](http://creativecommons.org/licenses/by/3.0/).