In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

University of Alberta

Academic Publishing in a Digital World and the Future of the University Press

by

Susan Ann Adele Hesemeier

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Arts

in

Humanities Computing

Department of Comparative Literature, Religion, and Film/Media Studies

Edmonton, Alberta

Fall 2003

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services
Acquisisitons et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-87573-3
Our file Notre référence
ISBN: 0-612-87573-3

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

University of Alberta

Library Release Form

Name of Author: Susan Ann Adele Hesemeier

Title of Thesis: Academic Publishing in a Digital World and the Future of the University Press

Degree: Master of Arts in Humanities Computing

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

August 28, 2003

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Academic Publishing in a Digital World and the Future of the University Press submitted by Susan Ann Adele Hesemeier in partial fulfillment of the requirements for the degree of Master of Arts in Humanities Computing.

Stéfan Sinclair

Sean Gouglas

Harvey Quamen

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Abstract

The scholarly publishing environment has changed dramatically with the introduction of the Internet and digital publishing and dissemination methods, though the university press has not been as active as the library or academic department in researching electronic publishing endeavors. The credibility of academic electronic publishing will depend quite strongly on the incorporation of the university press into electronic publishing endeavors being conducted through the library and academic departments. As an important entity in the dissemination of academic knowledge, the university press plays an important role in scholarly publishing, but must first study the current state of electronic publication credibility, the technical tools that are available or in development, and long-term issues such as open access and copyright that will affect how academic electronic publications are produced in the digital future.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table of Contents

Introduction 3
Chapter 1: Establishing Credibility: Preparing the Academic Community for Electronic Publications 13
Chapter 2: Digital Scholarship and Academic Experiments with the Electronic Medium 30
The Credibility of Academic Hypertext: Evolution of Peer Review in Scholarly Hypertext Articles 38
Chapter3: Discovering the Tools of the Trade: Technologies and Standards for Academic Electronic Publishing 57
XML and Educational Repositories 57
The Semantic Web and Academic Publishing 66
Standards for Academic Electronic Publishing; or, Why the TEI May Not Be Compatible with University Press Publications 69
An Example of the Difficulties of Using the TEI in Practice with Electronic Publications 73
Using XML Namespaces, XSLT, XML Schema, and RDF with Academic Publications 81
Chapter 4: Bending the Rules of the Game: Access, Archiving, and the Pragmatics of Copyright 91
Open Access 91
BioMed Central 91
Conclusion 97
Works Cited 100

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table of Figures

Figure 1 33
Figure 2 33
Figure 3 30

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Academic Publishing in a Digital World and the Future of the University Press

Introduction

> Let me make a forecast: *In the next 50 years, the entirety of our inherited archive of cultural works will have to be re-edited within a network of digital storage, access, and dissemination. This system, which is already under development, is transnational and transcultural.* Let's say that prophecy is true. Now ask yourself these questions: Who will be carrying out this work? Who will do it? Who should do it? (McGann, "Literary," B7)

Jerome McGann, co-founder of the Institute for Advanced Technology in the Humanities and Professor of English at the University of Virginia, addresses in "Literary Scholarship in the Digital Future" the question that university presses and academic institutions are facing in the twenty-first century: how will traditional dissemination platforms for academic research, such as the publications from a university press, keep up with change in the digital age, and at what point must there be a functioning electronic dissemination and retrieval system for scholarly publications? Coinciding with these issues is how we will make available and archive publications in an age when increasing numbers of undergraduate students will not visit their institution's physical library.[1] These issues, and the changes academics face as electronic access to information is

[1] See Lombardi.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

demanded by an increasing number of students and researchers, must now be investigated by academic institutions and, as a principal creator of academic publications, the university press.

The issues facing the creation of an academic electronic publishing system can be placed broadly into three categories: 1) How is the credibility of electronic publications to be established, through means such as an effective peer-review system, in order to fully exploit the potential of the electronic medium for academic publishing? 2) What technologies do we currently have to work with, and what developments, such as standards, can aid in the implementation of these technologies? 3) What larger-scale issues need to be addressed, such as copyright, access, and coordination between libraries and museums, in order to create a functioning and practical academic electronic publishing, archiving, and retrieval system? The university press, as a vital gatekeeper in the production of academic publications, must be employed to address these issues, though many academic institutions have not recognized the importance of the university press in producing electronic publications.

Involving the university press as a leader in the publication of digital material would be a major step that has until very recently been left out of the development of academic electronic publications. Although university presses have been active in creating electronic journals, non-journal material has been taken on primarily by libraries and university departments. For example, the Electronic Text Centre[2] at the University of New Brunswick holds a collection of digitized works, and is involved in a variety of projects aimed at furthering

[2] Available online at <www.lib.unb.ca/Texts/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

academic electronic publication for both 'digitized' and 'born-digital' works. Of its many projects, these include a Digital Imaging Centre that "fulfills part of the Electronic Text Centre's mandate to lend technical and educational support to University of New Brunswick faculty, students, and to other institutions for their electronic publishing activities," as well as a "metadata-driven Web management system for indexing and managing UNB Web content," and experimenting with various proprietary and open e-book formats ("ETC Research"). The University of Virginia Library also has an Electronic Text Center[3], which has had "twin missions" to "build and maintain an internet-accessible collection of SGML and XML texts and images," and to "build and maintain user communities adept at the creation and use of these materials" ("Electronic Text Center"). Other library and academic department projects include Indiana University Library's Victorian Women Writers Project[4]—a collection of "transcriptions of works by British women writers of the 19th century, encoded using the Standard Generalized Markup Language (SGML)" (Willett)—and Brown University's Women Writers Project,[5] an English-department-based project that has worked collaboratively with the University of Alberta English Department's Orlando Project[6] on the digitization of women's writing. There has therefore been some substantial work by libraries and university departments on making available and archiving electronic texts. But these are largely archiving projects that have electronically made available publications that were in print and perhaps difficult to find in their

[3] Available online at <http://etext.lib.virginia.edu/>.
[4] Available online at <http://www.indiana.edu/~letrs/vwwp/vwwp-about.html>.
[5] Available online at <http://www.wwp.brown.edu/>.
[6] Available online at <http://www.ualberta.ca/ORLANDO/main_page.htm>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

traditional form. The production of 'born-digital' academic publications in the electronic medium has not been undertaken to any considerable extent by either libraries or university presses, as disseminations of new information in electronic form—such as an online project, electronic book, or a database—have not received the same amount of attention that archiving electronic projects have through university libraries. And this emphasis on archiving projects in libraries might be justified because the responsibility of creating new 'born-digital' academic publications would reside with the university press.

McGann, in "Radiant Textuality," also discusses the position of the library in producing publications and explains why the library has been more involved than the university press in electronic publication endeavors:

> To date, computerization in humanities has been mainly located in work closely associated with the library, and with library-oriented research projects. This has happened for one simple and obvious reason: material demands have driven libraries to study and exploit computerized tools, which allow these research facilities to gain a measure of control over the massive amounts of data they are called upon to manage. (McGann, "Radiant")

The library has had a greater incentive than the university press to invest funds and research in electronic publications, through the form of digitization and archiving projects, as its costs could be minimized substantially by the reduced space that electronic archives and publications occupy. Though the university press would also benefit from a greater move to electronic publications, this has

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

not until recently been demanding enough to incite a larger investment of resources.

A recent article in Publisher's Weekly, titled "Sleepless in St. Louis," explains why it is essential for the university press to get involved in the academic electronic publishing process. John F. Baker reports on the Association of American University Press' (AAUP) annual convention and the currently harsh publishing market due to "diminished sales, rising returns and much disagreement as to whether publishing books with a strong trade flavor is the best way to mend their fortunes or whether it simply leads to higher returns" (Baker). The article outlines many of the current crises faced by university presses in North America, due to "higher financial demands from their parent institutions, the decline of foundation funding, the takeover of professional journals publishing by international conglomerates whose prices strain academic library budgets and the fact that scholarly book production was growing much faster than the market" (Baker). The directors of various university presses at the AAUP annual convention came up with many interesting solutions that ultimately employ the electronic environment. The directors of the Michigan and Minnesota university presses urged that "a public case must be more strongly made... for the role in the culture played by the presses" (Baker). This is one of the most important insights to come out of the gathering, and would remarkably improve the current development of academic electronic publishing by emphasizing to the academic environment the importance of the university press' role over private presses or the department/library in electronic publishing initiatives. Other solutions were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

provided by the Provost of the University of Kansas, David Shulenberger, who described the crisis as a product of state budget cuts that forced a higher concentration on teaching and research, and of serial costs that "had quadrupled in the past 15 years, a rise three times that of scholarly monographs, yet academic libraries had cut out far more monographs from their purchases" (Baker). According to Schulenberger, "libraries do what faculty members want, and it's journals they want," and he stated that the solution to this situation would "see humanities subjects moving away from books as a primary venue, with electronic sources preferred over print in most instances" (Baker). Another identified cause of the crisis was the decline of independent booksellers and distributors. But the solutions to the current problems facing university presses were clear: sustain the distinct mission of the university press to disseminate knowledge that is often "culturally unfashionable, politically unpopular and commercially unprofitable," in the words of Seetha Srinivasan, incoming board president of the AAUP—otherwise, as outgoing board president Peter Milroy states, the university press would "deserve to be replaced by library database managers" (Baker).

There have been some sustained developments with electronic publications by university presses, and the University of Virginia Press has been leading the way for many other presses through its Electronic Imprint (EI).[7] The EI's publications differ from electronic publishing projects of university libraries and departments because it publishes only "original digital scholarship in the humanities and social sciences" ("Electronic Imprint"). By publishing "born-digital" scholarship, or, in other words, "a work intended and constructed as

[7] Available online at <http://www.ei.virginia.edu/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

[digital scholarship] from its origin" ("Digital Scholarship"), the work undertaken by the EI is quite different from any of the digitization and archiving projects of university libraries and departments. Here the university press is actually leading the production of electronic works, but with the mission to "combine the traditional roles of university press publishing with technological innovation in order to disseminate peer-reviewed work comparable in its originality, intellectual rigor, and scholarly value to the books issued by the print side of the Press" ("Electronic Imprint"). This is a major development that some university presses, such as the University of Alberta Press, have begun to follow,[8] but much of the research and development that needs to be invested in this move has yet to be conducted. The University of Virginia has been involved in other electronic publishing endeavors, such as the Bibliographical Society of the University of Virginia's digitization of the entire backlist of its internationally-renowned journal *Studies in Bibliography* in ebook format,[9] allowing "more people [to] now see Studies in a month than have read it over the first half-century of its existence" (Vander Meulen, quoted in Gants). But research on the economic benefits of moving to electronic book-size publications should be emphasized as well because, at this stage, any unnecessary technical costs for university presses that already do not have the financial resources to invest in research and development may only hinder further progression in developing an electronic publishing arm.

---

[8] The University of Alberta Press has begun work on its first electronic publication, and is currently in the process of developing its own electronic imprint.

[9] Available at <http://etext.lib.virginia.edu/bsuva/sb/sbebooks.html>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

University presses have begun to feel a large enough incentive to tackle the technological hurdle of producing electronic publications, or to at least consider future electronic publication of print projects currently in production. Underemphasized, though, is that the most important aspect of future electronic academic publication will be in the university press' ability to work directly with the library and academic departments. The university press will have a significant competitive advantage over private presses if it works with academics and libraries from a work's conception to its publication, archiving and dissemination. This process is possible, though its advantages have not been understood to this point, and the integration of these separate bodies depends on the incorporation of a few core technologies, which will be discussed in Chapter 3.

The benefits of moving to the electronic medium are evident not only in a closer collaboration between the Press and academic departments, but through the ways in which the electronic medium makes some publications more practical to produce. For example, the production of print reference books that require a highly structured index in order to access the information provided in the book may make these publications more suitable or easy to publish in an electronic format because of the annotations and links to indexes that hypertext links can provide. If considering the library and archiving of publications when the university press produces them, these books would be created in an archival format compatible with Internet dissemination. If XML and non-proprietary databases were used for the creation of entries in a print reference book, for example, presses would also be able to migrate the entries into an electronic

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

format that is deliverable via the Internet, searchable, and linkable with hypertext. The press may even be able to request that the author of the work create the text/entries within the appropriate software or a content management system, eliminating the time-consuming process of converting an author's work into the appropriate publication format.[10] The edited publication could then be sent to the university library's archive for storage and dissemination, with the ability to update information, correct errors, and allow for instant search and retrieval, among many other possibilities provided by the electronic medium.

Along with the technical advancements and collaborations that need to develop in order for academic publishing to keep up with the electronic information demand, academic disciplines and the university press face a variety of challenges, not least of which will involve how the library will facilitate access to information via the Internet and collaborate with the university press in all of the phases of an electronic publication's delivery, from its initial inception and peer-review to the mark-up and tagging of the document for publication, and make it available to the end-user in a user-friendly and effective interface. There have already been initiatives, such as the *Columbia Guide to Digital Publishing*,[11] that address the need to create academic electronic publications that consider from their origin how they will be delivered and archived in an electronic format. To create the academic information infrastructure from which this system would stem, many issues need to be addressed that deal with copyright, access (and

---

[10] See Kasdorf, 2.

[11] Available online at <http://www.digitalpublishingguide.com>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

'open access'), and archival formats, and Chapter 4 will outline some of these issues.

The formation of a functioning academic electronic publishing and retrieval system will depend on the evolution and implementation of a variety of factors, and proposed solutions to current impediments are already being tested. New questions for research in this field will continue to develop that will hopefully not only serve as a guide to university presses in the process of developing electronic publishing, but also individual academics, departments, and libraries that are endeavouring to disseminate knowledge in the evolving electronic frontier.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 1:

Establishing Credibility: Preparing the Academic Community for Electronic Publications

William E. Kasdorf states in *The Columbia Guide to Digital Publishing* that "the real revolutions have already happened" in the technological side of academic electronic publishing (Kasdorf 2). According to Kasdorf, "most fundamental is the realization that the published content is independent of the physical products that convey it to us.... [It] was not until the digital era that we fully appreciated that a book or article is not inextricably bound up in the stacks of paper on which we read it" (Kasdorf 2). This has allowed us to think of more options for publishing that content, since "the digital era also liberates us from the page altogether" (Kasdorf 2). But although the technologies used to publish in a non-printed page format are now available to us, there are other factors, such as the credibility of electronic publications and the creation of an effective system of peer-review through the university press, that have yet to be remedied before we can say that all of the real revolutions in electronic publishing have already occurred.

Credibility for academic electronic publications is inevitable, as the preference for immediate access to library resources and publications is not going to go away. Artificial measures will most likely have to be instilled in order to get students to access their physical library, especially if better peer-review processes and accessibility are established for academic electronic publications.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

An article by Scott Carlson in *The Chronicle of Higher Education*, titled "Web-Loving Students Can Be Prodded to Cite Peer-Reviewed Works in Term Papers, Study Suggests," demonstrates the current attitude of many academics towards electronic sources. Carlson discusses a report released by Cornell University, titled "Effect of the Web on Undergraduate Citation Behavior," which showed that "students in a course at Cornell University have generally used fewer and fewer scholarly materials in their library research in the past six years" (Carlson). Within the article and this study, though, 'non-scholarly' sources are assumed to be synonymous with 'electronic' sources. The "trend" reported in this article of citing fewer scholarly materials was counteracted by an instructor in the study who provided clear guidelines in term paper assignments that students were to use "at least five peer-reviewed materials in the research" (Carlson). Again, 'peer-reviewed' materials are assumed to be non-electronic sources, and the printed book is considered by these instructors to be the most appropriate form.

In the years before the guidelines described in Carlson's article were established, "researchers had seen a precipitous drop in the use of books and an equally steep rise in the use of Web sites" (Carlson). But it would have been useful in the above study to also include an examination of graduate citation behaviour, as graduate students may be equipped with better research methodology skills and able to discern whether an electronic source is as useful as or better than a source they may find in print. Yet the fact that the undergraduate students in Carlson's article must be 'prodded' in order to cite print sources only enforces the permanence of this 'problem': students and researchers will use what

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is most easily available to them, which means accessible, fast, and free, and not signed-out; in the case of graduate students and researchers, if an electronic resource is available that is as credible and useful as what they would find in print, the more accessible version would be preferred.

Carlson's article shows a good example of the current attitude towards electronic publications within academe, and the trend to use electronic resources within humanities research is not isolated to undergraduate work. In fact, the necessity to review and develop electronic publication acceptance within academe has been addressed by many scholarly associations, including the Modern Languages Association Ad Hoc Committee on the Future of Scholarly Publishing. According to findings by the committee, "scholars in the humanities have become regular users of electronic resources such as bibliographies, encyclopedias, concordances, and databases available through university libraries," and online publications and journals are regarded as "presenting exciting new possibilities for our profession" (MLA 180). They also state that many issues still need to be addressed in the development of electronic academic publications:

> Most urgently, we need to address the issue of peer review for electronic publication in the humanities, whether of monographs and specialized books or of articles in online journals. It is crucial that electronic publications—including book-length studies, periodicals, editions, and scholarly Web sites—contain a statement about the form of review used to evaluate the quality of the work published and that such review be comparable in type and standard with that employed by university presses

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> and reputable print journals. Electronic publications included in tenure and promotion dossiers will likely be viewed as suspicious unless a widely accepted system of quality control is in place. (MLA 182)

As with citations of online sources in undergraduate papers, online publications on academic curriculum vitaes are not likely to be as reputable as their print counterparts. There is therefore a disparity between the acceptance of electronic publications by students, and their acceptance by tenure promotion and hiring committees as well as the academic community as a whole. For example, it is not uncommon for a student to search a library system's database of its print collection, come across a book that would be absolutely integral to the term paper he/she is writing, and realise that the one print copy is located in another library and is currently signed out. The idealizing academic would believe that students encountering this problem would have started the research for their paper with enough time to recall books—normally at least two weeks, if the book is returned in time. But the more common solution is that the student will move on to the next available book, sign it out of the library, and get by with the resources that are immediately available. Thus, the student does not use a resource that may be integral to their paper. And this situation might only occur in the first place if the student will actually enter their institution's physical library, rather than using only immediately-available electronic resources.

Students have made do with the resources immediately available to them, and although citing from the first five articles that appear in the library's database of electronic journal articles is the most efficient solution for many students that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

may be busily writing their papers at midnight the night before the due-date, some students will still go through the drudge of trekking to the library, searching through the library's database, taking the elevator or stairs to the appropriate floor, and meandering through the stacks of books that seem to be organized by librarians as a fun way to watch undergraduates wander through a rat-maze. After the student finally finds the appropriate call number, this may be repeated for every single book that is needed—that is, if the book is available. Our current library system does have flaws, yet many academics still wonder why there have been 'difficulties' with students increasingly citing electronic sources ever since the Web appeared.

Quality control of print academic publications has already been established through the university press, but in most scholarly discussions of the current state of academic electronic publishing and the need for credibility and peer-review, little is mentioned about the university press and its role in legitimizing publications. Ray Siemens, in "The Credibility of Electronic Publications," for example, discusses the current process of academic publication in relation to its history and the electronic environment. Siemens discusses the "argument *ad fonts*," which is used to "suggest that we might turn to earlier models of scholarly exchange" (Siemens)—models such as were used before the establishment of scientific journals and the more institutionalized exchange of academic research. This often involved passing around manuscripts published through subsidies from the author or their sponsor:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> In earlier times, when formal methods of inquiry and dissemination as we now know them were being shaped, the exchange of ideas and knowledge and the advancement of scholarship was facilitated as much by private exchanges and the circulation of private manuscripts and correspondence as by other means. (Siemens)

The development of scientific journals "saw the beginnings of academic publication's formalized dissemination," which also heralded "the development of the role of the scholarly editor—the scholarly editor, in this case, being someone who took on the task of re-circulating materials of interest to a community of scholarly readers" (Siemens). Of course, in our present context, the role of the scholarly editor has been assigned through the press, though the importance of the university press in this process has still tended to be underemphasized. Even when Siemens discusses the goals of the credibility report, they are concerned mainly with the current producers of academic electronic publications, which would involve other academics rather than the university press:

> working with those who are responsible for the review, publication and preservation of academic materials in electronic form, to ensure that the best qualitative standards, reliable dissemination and ongoing availability and archiving are wholly integrated into their publication and preservation processes. (Siemens)

What is missing here is an emphasis on working with the sector responsible for the *print* peer-review, publication, and preservation of academic materials, and to involve the university press in a process that should not be developing in its own

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

vacuum. But Siemens does recognize that "like peer-review, a publisher's imprimatur, or imprint, is seen to be a very important indicator of qualitative assurance in academic culture" (Siemens), and is one of the most important factors in laying the groundwork for academic electronic publications; the psychological groundwork within the minds of many academics that needs to be established:

> In short, the academic climate is, at present, one in which pragmatic, professional concerns appear to act as barriers to the acceptance and use of venues in the electronic medium that might, quite clearly, better facilitate work towards goals common to all academic and scholarly activities. (Siemens)

Although the Social Sciences and Humanities Federation of Canada (SSHFC/CFHSS) commissioned the report, it is a pitch that also needs to be aimed at the university press in order to encourage its movement from traditional print publication models.

The importance of the university press is also left out of another section of this report on the future of academic electronic publishing, even within articles that are directly on the topic of peer review and imprint, such as "Peer Review and Imprint," by Jean-Claude Guédon. The university press is mentioned only within the context of respondents to the survey upon which the article is based, which quantitatively studies the opinions of questionnaire respondents when asked whether they felt "a publishers imprimatur, or imprint, was important"

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Guédon). Guédon provides a summary of the process of peer-review and how it is important though also imperfect:

> Whenever a journal became more than the particular vehicle for an individual or a tightly-defined school, it resorted to having a peer-review process in place. As much a way to protect the editor-in-chief and his (or her) decisions as an attempt to guarantee some objective and serene judgment in the midst of intense competition, the peer-review process nevertheless rested on procedures that put the editor-in-chief in absolute control, albeit in an acceptable way. Receiving a submission, the editor can immediately reject especially if it can be argued that the manuscript does not fit with the journal coverage, or its style, etc. If the manuscript is accepted for review, it is then dispatched by editors to various reviewers. The choice of these reviewers is his or her own and one can easily imagine that a suitable choice ensures a good deal of the outcome. This arbitrariness is facilitated by the fact that, more often than not, the correspondence between editors and reviewers is done privately and cannot be consulted in a public manner. In short, the peer-review process, while often used absolutely honestly and to good effects, can nevertheless harbour enormous zones of occult arbitrariness. (Guédon)

Guédon describes the process of peer-review for print journals that functions similarly for books or for publications specifically produced by the university press, but, again, the possibilities of an imprint specifically associated with the academic institution have not been emphasized

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Guédon discusses the importance of the imprint, but in the context of how it has been built up "over the peer review process and has contributed to warping the whole evaluation results in a not insignificant way" (Guédon). The notion of 'core journals' in one's field has created a "fuzzy notion" of the "Impact factor" of an article, most notably within the sciences with the Science Citation Index (SCI), but Guédon also forewarns of a greater use of a list of core journals with a mechanism to rate articles by citation counts within the social sciences and humanities:

> Let us not forget that a Social Science Citation Index also exists and although its present importance is far more limited than its natural science counterpart, it is nevertheless pointing to some future which will be difficult to avoid. (Guédon)

Guédon hypothesizes about how this move to a more important SSCI would affect academia, although he believes that this "does not concern us directly as it relates to commercial publisher strategies to buy or engineer core journals" (Guédon), and that there would likely be a less elastic journal market if a citation index was given more importance. There is no use hypothesizing about this possible event, though, as especially in the Web medium these developments may move naturally in the direction of the greater good for academic electronic publications, and, as one could not foresee before the development of the Web, new technologies sometimes come along that may cause a completely different shift in development.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

There seems to be a general move against the imprint while keeping the process of peer-review within the academic literature on the future of scholarly publishing, and this would point to a move away from the university press. But difficulties that may be inherent in some journal imprints and peer-review processes, such as "commercial publisher strategies to buy or engineer core journals," would not apply to book publications of an imprint, or even as Guédon points out, to "social science and humanities journals (SSHJ) [as they] have largely escaped this fate so far because the notion of core journals is more difficult to establish in fields where several paradigms may coexist at the same time" (Guédon), as opposed to the sciences where core journals such as *Science* and *Nature* are well-defined. Guédon warns that it is only a matter of time, though, before the social sciences and humanities have the same monopolizing journals in place as in the sciences, especially if the scholarly monograph begins to be replaced more frequently by the journal article:

> the market for SSHJ has proved far more elastic than its natural science counterpart, and this has prevented the creation of de facto cognitive monopolies that have translated into enormous subscription or licensing costs in the science, technology and medical science (STM) publications. However, the writing is on the wall and it is only a question [of] time before this area follows the example already set up in STM. A quick glance at the evolution of certain fields such as economics and quantitative social sciences reveals similar trends, albeit with a delay. (Guédon)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Alternative methods of ranking journal quality have been in development, such as the "approach based on a forensic expert witness testimony and computer litigation support" (Rushinek) proposed in the article "What Makes a Publisher Important" by Avi Rushinek and Sara Rushinek, but this would not be more effective than peer-review and imprint in electronic publications. Their approach "uses the relative number of citations of a search engine as the evaluation criteria" (Rushinek) and would create a separate electronic system of peer review than has been used with print publications, whereas the incorporation of the same peer-review and imprint mechanisms used with print publications would be more effective at establishing credibility for publications in the electronic medium.

Guédon emphasizes the importance of peer-review as "it allows the selection of the best metarials," and because it is "widely used to manage personal careers," but imprint is given attention as a dangerous "conflation of individual value with journal imprint factors" (Guédon). Guédon sees electronic publishing's potential rivalry with imprint as a possible solution to some problems with imprint:

> No wonder that electronic publishing is viewed with caution as it obviously is bound to modify a number of rules that will affect power structures within scientific communities and economic revenues for a number of large, rich and powerful commercial publishers. It is important to remember this in reviewing the literature on the topic as much of it is embedded in this latent struggle for preeminence or revenue streams. (Guédon)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Though some academics may see technology as a 'saviour', the electronic medium will not solve some of the problems that may be inherent with imprint: in fact, the electronic medium has emphasized these problems, which can happen when new technology is brought in to fix something. The introduction of the electronic medium has, though, shown many scholars the importance of providing an imprint with its implied peer-review in order to attain credibility for publications.

Although the move to greater production of electronic publications through the university press would imply that the process of peer-review would benefit these publications, it may be argued that acceptance rates of publications by some of the most prestigious presses are also a strong factor in producing credibility for electronic publications. For example, the amount of articles that are accepted for Oxford University Press publication, as compared to a smaller university press, and the subsequent acceptance rate, would be much lower for Oxford University Press. This adds to the prestige and credibility of the publication, although acceptance rate cannot be said to function separately from peer-review and imprint. The publications from a reputable press are credible because of the quality they are known to have, which is not disconnected from the processes of peer-review used by the press that ensure this quality, and its reputable imprint is a consequence of its high quality publications *and* its low acceptance rate. In the context of electronic versus print publishing, though, this functions quite similarly—anything can be placed on the Web, so the 'acceptance rate' for Web publications in general is quite high; when it comes to having a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Web publication produced through an imprint, though, the acceptance rate is lowered quite a bit, thus adding a degree of implied quality and prestige to these publications.

The electronic medium does add many benefits to electronic publications, but, as Alan Burk, James Kerr, and Andy Pope note in "Archiving and Text Fluidity/Version Control," there are some issues that arise from the ability to constantly change a publication. They define text fluidity as such:

> Texts can be adapted, abstracted, translated, edited, condensed, corrected, marked up, transcribed, annotated, amended, paraphrased, transliterated, illustrated, indexed, or abridged. They can be commented upon or referenced by other texts. They can be analysed or synthesized. These modifications, and any others that might be considered, produce distinct versions, all related to the original text. Modifications to a text do not produce a text that is fluid. Every version of a given text could be uniquely named and its relation to the first version, indeed its relation to every other version, could be specified if such relational specificity made sense. All of this holds whether a text is hand written, typewritten, printed, in braille, on punched cards, in some analog auditory medium, in electronic form, or produced in some form yet to be developed. (Burk et. al.)

What happens with the ability to replace an original version in the electronic form, though, is that it becomes more difficult to control and keep track of these different versions. As will be discussed in chapter 3, though, developments with metadata promise to be able to automatically encode information about each

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

version of a text, enabling different versions of a text to be managed as they are updated or replaced.

Peer review for academic electronic publications has been addressed largely on the scientific journal front, as these publications tended to accept the electronic format for its speed and accessibility long before the humanities and social sciences.[12] But the current peer-review process in academic journals has come into question by various proponents, most notably in an article by Frank J. Tipler in the International Society for Complexity, Information, and Design. Tipler approaches the process of peer review as a scholar involved in a 'theistic evolutionary studies', and he criticizes the current process of peer-review in refereed journals as restricting innovation and new ideas. The current scientific peer-review process is different from its historical origins, and like Siemens, Tipler tends to favour a more traditional approach for our contemporary publishing environment:

> prior to the Second World War the refereeing process, even where it existed, had very little effect on the publication of novel ideas, at least in the field of physics. But in the last several decades, many outstanding scientists have complained that their best ideas— the very ideas that brought them fame—were rejected by the refereed journals. Thus, prior to the Second World War, the refereeing process worked primarily to eliminate crackpot papers. Today, the refereeing process works primarily to enforce orthodoxy. (Tipler)

[12] See Kasdorf.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Tipler argues that the current process of peer-review does not perform the same function as when it was first established, and that "'peer' review is *not* peer review: the referee is quite often not as intellectually able as the author whose work he judges" (Tipler). The current system of peer-review may not be, as Tipler states, "pygmies standing in judgment on giants," (Tipler) but he does make some good points about the effect of contemporary peer-review practices on the publication of ideas that may be less fashionable or have not yet gained a large audience. Tipler cites the publication of three important papers by Einstein in *Annalen der Physik* as examples. The papers, which presented Einstein's theory of relativity, photons, and the 'Brownian motion' of dust particles, were published in this major physics journal without any of the papers having been sent to referees:

> Instead the editors—either the editor in chief, Max Planck, or the editor for theoretical physics, Wilhelm Wien—made the decision to publish. It is unlikely that whoever made the decision spent much time on whether to publish. Almost every paper submitted was published. So few people wanted to publish in *any* physics journal that editors rarely rejected submitted papers. (Tipler)

What Tipler seems to find as the solution to the "orthodox" process of peer-review in scientific journals is the ability for anybody to publish, which was possible due to the low supply of article submissions and high demand for content for *Annalen der Physik*. But this is possible in our present context with the ability to publish anything on the Web. Though the process of peer-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

review may be conservative, this is precisely what gives it value, and especially in a medium such as the Web in which the publication of anything *is* possible.

What is special about the electronic medium is that it could allow both of these factions to co-exist: the push for the publication and acceptance of new ideas that Tipler proposes, and the conservative, filtering or restraining process of peer-review that has evolved over the last century to what has been in use primarily with print publications. The evolution that has to be taken one step further is the full implementation of the peer-review process into the Web. Web users have been publishing anything on anyone and everything in the electronic medium, through personal Web pages, weblogs ('blogs')[13], and other files and Web pages produced through easily available desktop publishing tools. What we do not have in the electronic environment is a co-existing standard for publications and a credible source for most electronic-based publications produced by academic institutions. *Academic Search Premiere* and other library subscription databases are available that hold scholarly journal articles, but the monograph and the university press have continued to be left out of this process, and its incorporation is a necessary step in the future of academic electronic publishing, for both the university press and scholars.

---

[13] A weblog, "also known as a blog, is a frequently updated website consisting of dated entries arranged in reverse chronological order so that the reader sees the most recent post first. The style is typically personal and informal. Freely available tools on the World Wide Web make it easy for anybody to publish their own weblog, so there is a lot of variety in the quality, content and ambition of weblogs, and a weblog may have anywhere from a handful to tens of thousands of daily readers." From Jill Walker in jill/txt weblog <http://huminf.uib.no/~jill/archives/blog_theorising/final_version_of_weblog_definition.html>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 2: Digital Scholarship and Academic Experiments with the Electronic Medium

Definitions of 'digital scholarship' have started to become defined by presses and academic institutions, and the Electronic Imprint of the University of Virgina Press provides an excellent definition of what 'digital scholarship' is:

> Digital scholarship is publication that (1) exists in digital format, i.e. as an electronic file or set of files that can be stored, transported, and displayed on general-purpose computers or other devices that manipulate digital files; (2) is incapable of being translated without loss of information or value into a non-digital format, such as that of a printed book, because it makes use of media, tools, structuring, or other features of computer presentation that cannot be conveyed in any other medium; and (3) is subject in all other respects to the demands of traditional print scholarship for originality, value, and selection via a process of peer review. ("Digital Scholarship")

Thus, the digital files used to publish a hard-cover book are not 'digital scholarship', because the final publication "is not in the digital format" ("Digital Scholarship"). The definition expands to include "born-digital scholarship," and even electronic publications that may appear to fit into this definition may not be "digital scholarship" ("Digital Scholarship"). For example, a "PDF file representing the precise appearance of the printed book is not DS, because it can be translated into the printed book without loss of information" ("Digital

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Scholarship"). Even Web-based editions of an originally-print book that may expand upon the text, are not "born-digital scholarship", and the University of Virginia Press is clear on making this disctinction:

> But suppose I create a Web-based edition of *Anatomy of Criticism* that expands upon the text in many ways. I add detailed structural tagging to identify authors, genres, languages, and dates of literary quotations, making it possible to search the text using each of those categories. I add hyperlinks to online sources of all quotations, both to primary and to secondary sources; I even add multimedia links to video files or audio recordings of scenes from Shakespeare that Frye refers to. I provide an annotation function that lets readers add their own links and commentary on the text. I use tagging to create "paths" in the text that allows a reader to trace particular themes from point to point in the book, creating alternate views or "maps" of Frye's work. ("Digital Scholarship")

The resulting work fits within the definition of digital scholarship. Its content can never be captured in its entirety by a printed book, no matter how long or heavily illustrated. And if all of our technology had been around when Frye first planned his work, and if *Anatomy of Criticism* had first been published online, the result would have been "born-digital" scholarship: a work intended and constructed as digital scholarship from its origin.[14]

Digital scholarship and an effective future peer-review system might also include guidelines for evaluating software and the technical tools behind

[14] See "Digital Scholarship."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

electronic publications, such as in the Text Analysis Portal for Research (TAPoR)[15]. Susan Hockey, in *Electronic Texts in the Humanities*, emphasizes that "more effort needs to be directed towards the development of guidelines for evaluating computer-based work" (Hockey 170). This process is very different from evaluating material in print, according to Hockey, as "it takes time to understand all the possibilities in a piece of software" (Hockey 170). To Hockey, the ideal electronic publication would be an "electronic text": a term "used specifically to mean a transcription of a text, rather than an electronic or digital image of it" (Hockey 1). The importance of this distinction between an 'image text', such as a PDF, versus an electronic text is essentially the difference between the possibilities of the print publication versus the electronic publication: the static image/print text is not able to be searched by automated search programs or embedded with metadata beyond card-catalogue information, along with many of the other functions that can be performed on electronic texts. As such, we should also be exploring the software tools that may be packaged along with these publications: concordances, lexical databases, descriptive and prescriptive mark-up, metadata and intelligent search agents. These require skills and time that may be beyond the scope of the traditional university publishing house, which enforces the need for collaboration between departments, the university press, the library, and its community of users.

There are other factors that require collaboration on a very large scale, such as the need to train students (and future scholars) in digital scholarship.

[15] Available online at <http://huco.ualberta.ca/Tapor/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Scholars have noted that the proliferation of editorial and archival projects in digital form requires students trained in both the practical and theoretical components of digital publishing.[16] But these programs have not, to this point, been carried out by academic departments or in collaboration with university presses:

> Departments of literary study have perhaps the greatest stake in these momentous events, and yet they are—in this country—probably the least involved.... Here, the work is being carried by librarians and systems engineers. Many, perhaps most, of those people are smart, hardworking, and literate. Their digital skills and scholarship are outstanding. Few, however, have a strong grasp of the theory of texts. (McGann, "Literary," B8)

A new model of the publishing process needs to be developed that will incorporate all of these entities, and the scientific community's progress in electronic publishing can be followed by the humanities as well. Figures 1 and 2, from a presentation by BioMed Central, sum up the current publishing process and provide a new model for academic electronic publication.

[16] See McGann, "Literary," B8.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1:

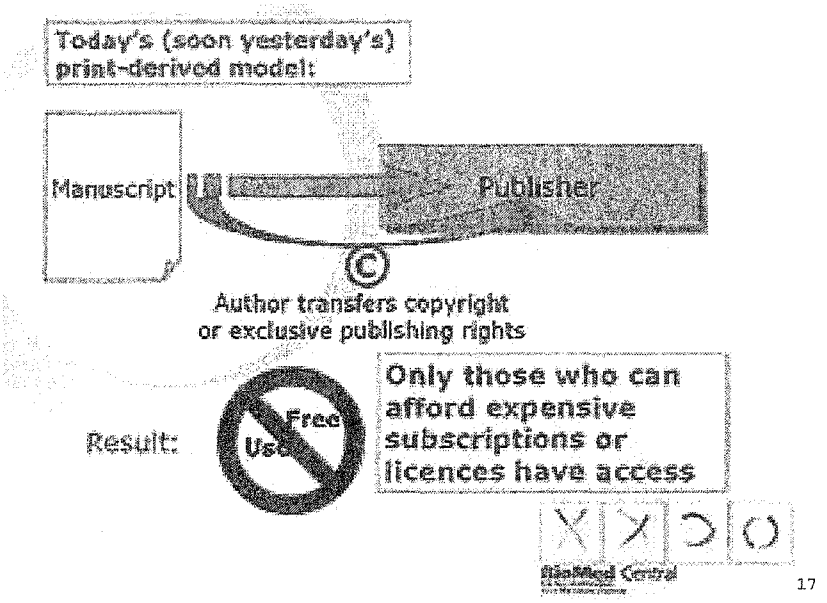
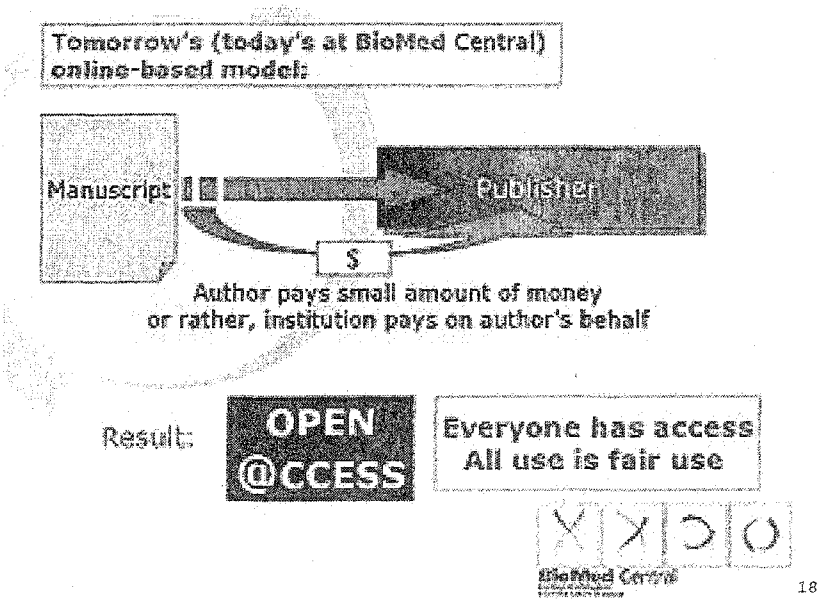


Figure 2:



[17] Available online at <http://www.biomedcentral.com/>.
[18] Available online at <http://www.biomedcentral.com/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

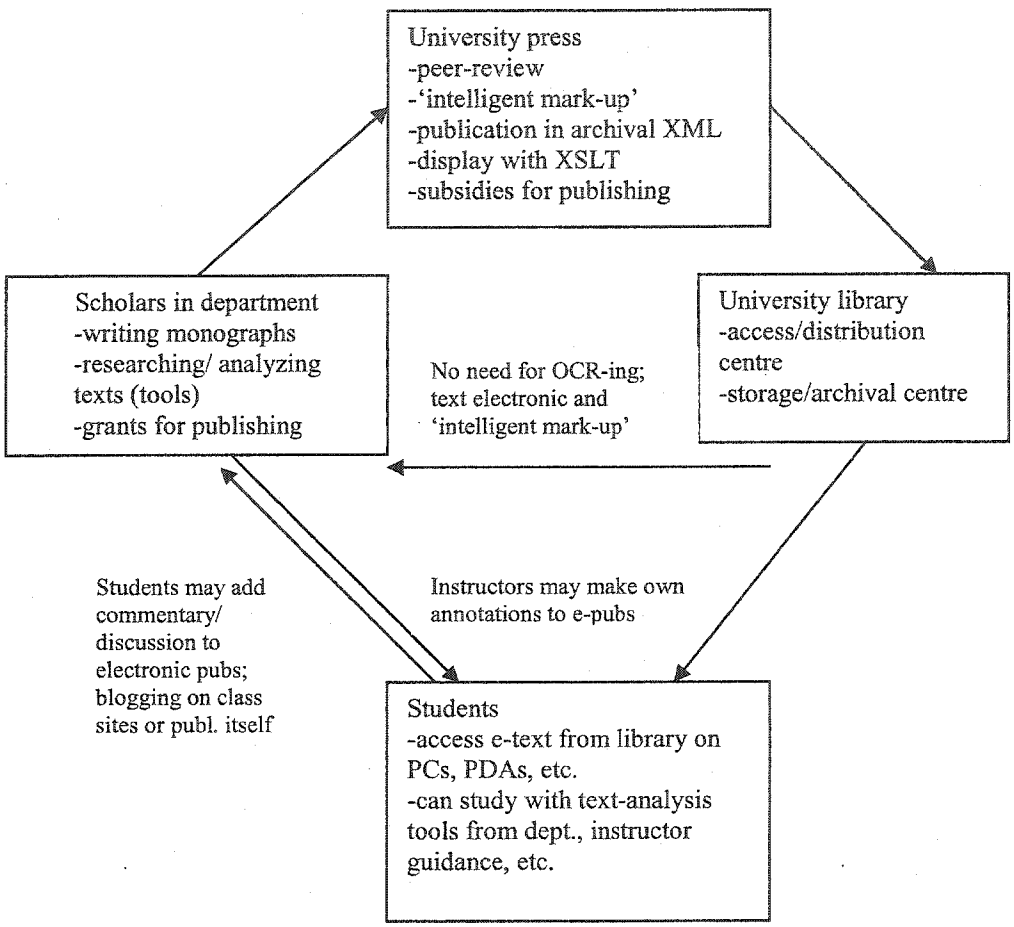
This new model of publishing in Figure 2 can be applied almost directly to the humanities and social sciences, although the speed at which electronic articles can be published is not as much of a driving force in the move to electronic publications, hence the social sciences and humanities' slight lag. But perhaps the need for speedy 'publication' of ideas has been addressed in the humanities and social sciences by Weblogs and email listservs such as the Humanist Discussion Group,[19] in which humanities scholars rapidly post and disseminate information. It is likely that electronic publications in the humanities would combine properties of these electronic information dissemination technologies that are already in use, thus changing not only the way in which we disseminate academic knowledge, but also the form in which this knowledge is organized.

Figure 3 is a model of what an open-access scholarly electronic publishing system might look like in an environment that involves the collaboration of the university press, the library, scholars in departments, and students:

[19] Available online at <http://www.princeton.edu/~mccarty/humanist/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3:



In Figure 3, publications are produced through a process similar to BioMed Central's system of electronic publication. Scholars in departments provide the processing fee to the university press when submitting their article for publication. The university press then provides peer-review for the publication, encodes it with intelligent mark-up in archival XML with XSLT[20] for publication, and distributes the XML and XSLT files to the library. The library would deposit the publication in its archive that makes publications available to students and researchers, with

[20] The use of XML and XSLT in publishing will be explained in Chapter 3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

any tools that academics may need to use with the text, such as text-analysis tools and search engines. Making the publications available on the Web through the library would also allow students to comment on the publications in the form of Weblogs associated with each article, allowing user discussion communities to develop around publications. This would also eliminate the need to scan a print text through Optical Character Recognition (OCR) scanners to produce electronic texts, thereby reducing the amount of labour spent by academics converting from print to electronic documents.

The above process would only work with the university press, rather than private presses, in part because of the grant funding that could be acquired through collaborative publishing ventures in the academic department, university press, and library, and the seamless integration that being part of the same non-profit community would provide. Presses that operate based on a profit would not benefit from the above model, and would charge the library for access to their publications even if the scholar submitting the publication provided the processing fee. The university press, on the other hand, receives subsidies for its publications in certain subject areas, and combined with print-on-demand technology,[21] would allow the university press to continue its mandate of producing publications that may focus on a small audience or specific subject-matter, without incurring financial loss through excess copies.

[21] Print-on-demand technology will be discussed in Chapter 4.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Credibility of Academic Hypertext: Evolution of Peer Review in Scholarly Hypertext Articles

> "Although there are a great many texts online, there have been few attempts to incorporate online texts into online scholarly articles. This seems a shame, for hypertext might allow us to better interrogate the object of study, to bring our evidence before the bar, as it were." (Thurston 252)

An analysis of experiments in the peer review of scholarly hypertext articles provides a useful study of the current state of peer-review in one genre of academic electronic publications. The production of hypertext works has remained largely within the experimental domain of creative fiction, and although hypertext works have been studied since the early-1990s by theorists such as George Landow and in recent works by N. Katherine Hayles, hypertext articles have not received the peer review and recognition that they deserve as valid modes of communicating research. This practice needs to be examined, as hypertext has been shown to be a powerful tool for the creation of not only fictional creative works, but also journal articles and works of criticism.

Though there have been some experiments with online hypertext editions of journal articles and hypertext editions of canonized originally-print works, the use of hypertext thus far has centred around its use as a supplement to the book medium, rather than as a potentially revolutionary tool for academic publication in its own right. By undertaking an analysis of existing academic hypertext works, from experiments with hypertext journal articles, to hypertext editions of literary works, we can determine the potential of this new tool for the study of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

literary texts, and perhaps embark on new ways of employing it effectively within the humanities. Roy Rosenzweig, guest-editor of a special issue of *American Quarterly* (AQ) in which he produced an experimental online hypertext . instalment of the journal, writes in his call for proposals to this special issue that, "although there has been much theorizing about hypertext and scholarship, there are very few concrete examples of scholars using hypertext and new media to present the results of sustained inquiry into the subjects that they study" (Krasniewicz and Blitz, "Original Call For Proposals"). This special issue of *AQ* was a unique opportunity for authors to create hypertext articles that disseminated information about their research, as other publication venues would not likely have accepted submissions in the form of an experimental hypertext article. *AQ* thus attempts to remedy this lack of an experimental test bed for academic hypertext publications, and within it we can see the first stages of evolution in the academic hypertext article. As commented by one of the authors of this issue's hypertext articles, this experiment attempts to remedy the constraint imposed by the standard accepted form of the academic article:

> Through the centuries, the definitions of science and knowledge have forced a type of academic writing that is severely constricted and narrow. Despite challenges by literary studies, history and anthropology that have introduced personal narratives, reflexivity, autobiography, raw interview data, and other experimental writing forms, the standard academic text still reigns and those who want to be legitimate members of the community

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

must produce it, despite its inherent limitations. (Krasniewicz and Blitz, "Part IV: Hyperdreams")

Hypertext thus allows academic writing to break some of the bonds of traditional academic standards, but at this point in the development of hypertext articles, it must have a 'safe' environment in which to develop, such as *AQ*'s experimental issue. Projects such as these are important to electronic publication, as they allow for experimentation in not only creative works, but in scholarly publication as well—an opportunity that the editors of this special issue recognize as rare and unique: "we soon realized that if we asked people to submit hypertext essays for review and then, as would be expected, we rejected many of them, the authors would have no other journals where they could send their work" (Rozenzweig 238). They amended this situation by requesting proposals from authors, rather than completed articles.

The articles contained within this special issue are quite different from one another, and can be observed in terms of two phases borrowed from N. Katherine Hayles in *Writing Machines*. The first phase employs hypertext as an aid or extension to the book medium, while second-phase hypertexts explore the possibilities for academic hypertext publication beyond how it can accentuate the book. Hayles discusses these two phases in the context of creative or fictional hypertext works. First-generation hypertexts, which were "largely comprised of text, making little or no use of graphics, animation, and sound," had "relatively simple navigation systems that consisted largely of clicking on links to go from one lexia to another" (Hayles 37). First-generation hypertexts functioned

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

primarily as a transition phase between the book and what would become second-generation hypertexts, as "these first-generation works were more like books than they were like second-generation electronic literature, because they operated by replacing one screen of text with another, much as a book goes from one page to another" (Hayles 37). Second-generation hypertexts, on the other hand, began to exploit the navigational potential of hypertext to create effects that differed from the print medium and did not ignore the materiality of the machine:

> Navigation was envisioned as taking place not only between lexias but between images and words, and more profoundly between the text and the computer producing it. This was something very different than moving from lexia to lexia; it was an effect print could not duplicate. (Hayles 38)

Some creative hypertext writers, such as Shelley Jackson in *Patchwork Girl*, explicitly drew connections between the machinery and the text through certain passages that asked about the relationship between the viewer's consciousness and the gaps between screens, and how the medium interacted with the body of the viewer. The transition to second-generation hypertexts in creative works, then, heralded a movement away from the effects of the book medium and to hypertext works that discovered and played with the potential of the electronic medium and the machine.

Creative genres, such as the novel, have tended to be the breeding ground for testing boundaries of a new medium and modes of expression, and fictional creative hypertext works have been the vessel for testing its inherent possibilities. It is important, at some point in the evolution of a new medium such as hypertext,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to examine whether it may be advantageous to employ techniques of this new tool in the production of scholarly publications. We do not need to produce electronic publications that mimic as closely as possible their print counterparts, such as first-generation creative hypertexts have done, although some aspects of the print medium such as linear argumentation may be appropriate to incorporate in certain electronic publications. But within academe, we do not really have any representative works of second-generation hypertexts in our publications besides what can be found in creative works or experimental issues such as *AQ*. Hypertext does have potential beyond its use for first-generation hypertext purposes such as quick linking to textual annotations and navigational points within an article, and only through further experimentation and analysis can we discover ways in which we may harvest the potential of second-generation hypertexts to include multimedia and a consciousness of the medium.

As we have seen with the evolution of creative fictional hypertexts, the first phase is usually a transitory phase that mimics its preceding phase—in this case, the book. Critical hypertext writing has followed the same evolutionary pattern, although its creative counterpart has moved much further into the second-phase. The use of hypertext in scholarly articles has remained first-generation in its employment of hypertext to complement conventional book forms such as for quick links to textual annotations, or to navigate between sections, as will be demonstrated below. Creative works have followed this transition, but academic articles have not, not only as a result of conservative scepticism by reviewers,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

editors, or publishers to a new mode of scholarly communication and research dissemination, but due to the lack of experimentation in academic publications.

Roy Rozenzweig, guest editor of the special 1999 online issue of *AQ*[22], comments that the mixed goals of this issue were to "bring together something rather old-fashioned and established—the scholarly journal article—with something new and still emerging—the networked and digital space of the World Wide Web" (Rozenzweig 237). What makes this project especially promising to the evolution of academic hypertext publications is their shared interest in "encouraging unconventional departures in form while also retaining the conventional validation and peer review that characterizes scholarly publication" (Rozenzweig 238). The four articles contained in this special issue show the potential that is inherent in experimental projects of this nature, and they range from very minimalist uses of hypertext to articles that include elements of second-generation hypertexts.

The first article, Thomas Thurston's "Hearsay of the Sun: Photography, Identity, and the Law of Evidence in Nineteenth-Century American Courts," uses hypertext to present the actual object or 'evidence' to which the text of the article refers. As Rozenzweig comments, "it presents not just an argument about the legal status of photography in the previous century but also virtually all the evidence that supports (or perhaps even undercuts) the argument" (Rozenzweig 239). Thurston's hypertext article is structured in the way that a reader might be presented with evidence in a book—the page is divided into four frames, each of which serves to recreate the conventions of the book. At the top of the page is an

[22] Available online at <http://chnm.gmu.edu/aq/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

'index' frame containing links back to the *American Quarterly* online issue homepage, another link back to the beginning of Thurston's article, and links to the bibliography, footnotes, information about the project, resources, and a search engine. This first frame follows the conventions of an index bar at the top or side of a web page, which also represents the index of a book. A second frame encompasses the left side of the page, and is the anchoring 'text' from which images and notes in two frames on the right hand side of the page are based. Although Thurston argues that this hypertext article is quite different from its print predecessors, it presents the reader with a text that is very similar to a printed text in which the reader must flip back and forth from the footnotes to the main anchoring text. The left side of Thurston's article functions as the main text, and the other three frames serve as sections to which a reader would refer when reading a print book—an index or table of contents in the top screen, a figures section, and a notes section. The 'evidence' that Thurston claims is possible to put forward in an electronic hypertext version of this article is just as possible in a print version—all that this would involve is the reader flipping back and forth between a few pages, rather than clicking on a link with the mouse. Thurston's hypertext article therefore maintains many of the conventions of the 'book', and is a concrete example of a first-generation academic hypertext publication.

Another article in this special issue is quite indicative of a first-generation hypertext, although it also employs the electronic environment's potential to include more non-textual media. James Castonguay's "The Spanish-American War in U.S. Media Culture" uses hypertext to organize its sections into an index.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The first page consists of a title, a mail-to link to the author of the article, an introduction, and a contents list organized in a series of hypertext links to each section of the article. The article is structured in the style of a book, and it uses hypertext not only for organization but to be able to include different media types as well. For example, Castonguay is able to link to other media forms such as films in his article, in order to show as 'evidence' the object to which the text refers:

> Internet users may download and view several Spanish-American War films while they continue to read the article, either by linking to the MPEG and QuickTime movie files at the Library of Congress, or by streaming RealMedia versions I have encoded for slower modem connections. From a media studies perspective, incorporating these films into an online presentation is a clear enhancement over a static print publication. As I continue to revise and expand the article, my goal is to offer hypertext-rich analyses of every Spanish-American War film available in the National Digital Library's growing collection. In addition to this exciting opportunity to incorporate entire films, I was also able to include many more illustrations from newspapers—and frame enlargements from films—than would have been possible in a traditional print piece. (Castonguay 248)

This seems to be the extent of Castonguay's embrace of the new medium, as, in a manner similar to Thurston, he embraces hypertext's ability to bring other media forms such as the photograph and the movie clip before the reader as 'evidence'

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to replace the print medium's description of 'hearsay' textual evidence. But these hypertexts still function within the narrowness of the preceding medium of the book—they employ hypertext and extol its abilities as a way of adding to or supplementing the print article.

Castonguay does raise an interesting point about the advantages of hypertext articles in his discussion of his project, as he mentions the possibility to exploit the fluid nature of hypertext and the Internet by being able to continually revise one's work and to include a variety of media types. Hypertext allows Castonguay to bring in as evidence a variety of media types without having to describe them solely with text. This, and the ability to continually revise a text, can pose problems, though, as the limits to what one can include as evidence becomes endless, and the non-static but continually-shifting publication is never stable. This can have both good and bad effects and, as we will see in other academic hypertext works described below, it is up to the author to find the proper balance between the limitless ability to link to anything and what is practical and functional.

A third hypertext article in this special issue, "From Hogan's Alley to Coconino County: Four Narratives of the Early Comic Strip" by David Westbrook, adds to the conventions of the book employed by Thurston and Castonguay by placing hyperlinked comic-strip images throughout the article. Clicking on one of these images will open a new browser window with a larger picture of the image, and links to sections of the article that discuss this image. The way in which the hypertext and images are combined in this article is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

different from other articles in this issue, because the reader may click on a link above the image, such as 'Pioneering the tiny animal,' and the section of the image that this hyperlink is meant to accentuate will be outlined in red with a description underneath containing a link to the section of the text that discusses this part of the image. The links from images to different sections of the text adds a level of complexity that differs from the Thurston and Castonguay hypertext articles, and it can tend to disorient and lose the reader by causing the reader to jump through different sections of the text. The technique of using hypertext to draw a red outline around a section of an image is a useful technique, but, again, this article also restricts itself by employing hypertext to accentuate the properties of the book by being able to point out parts of an image.

A fourth article in the *AQ* special hypertext issue, titled "Dreaming Arnold Schwarzenegger" by Louise Krasniewicz and Michael Blitz, is noticeably different from the other articles from the very start: it includes sound and non-conventional book elements. It also self-consciously breaks from a linear narrative, by allowing the reader to begin by choosing, ironically and somewhat mockingly, from four different sections that are hyperlinks imposed on the image of a book. The 'essay' section of the article begins as such:

> You may have come to this essay first, as your choice for navigating through this project we call "Dreaming Arnold Schwarzenegger." Or you may have come to it after explorations of some of the other interconnected and unusual elements that make up this website. In either case we feel obliged to describe briefly what you will find here and why, in a very real

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> sense, you should not read this the way you probably want to read this. This essay, in four parts, describes and argues in more traditional academic terms than does the rest of this website the issues we think are relevant to an attempt to translate academic research into hypertext forms. (Krasniewicz and Blitz, "How Did You Get Here?")

Not only has *AQ* enabled Krasniewicz and Blitz to publish an experimental hypertext publication in a reputable journal, it has given them a basis on which to theorize the practical instantiation of this hypertext article—a practice that can bear much more fruitful results than theorizing without a practical experiment: "what started out as a straightforward publication of our multi-year research on Arnold Schwarzenegger as a cultural icon has turned into this theoretical and practical discussion of how and why this should be done in hypertext form" (Krasniewicz and Blitz, "Dreaming Arnold Schwarzenegger"). This also enables Krasniewicz and Blitz to draw new conlusions about the electronic medium. They use a metaphor of the dream to structure the information in the article, and to theorize about the structure of information in general in our post-modern environment:

> ...this text should be "read wrong." It should not be read as an argument that is trying to coerce or seduce you into seeing it our way. It is simply (or not) an effort to show you how to see things differently. This, of course, is the legacy of dreams. (Krasniewicz and Blitz, "Dreaming Arnold Schwarzenegger")

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

"Dreaming Arnold Schwarzenegger" allowed the authors to construct a metaphor that related the work they had done with dreams and Arnold Schwarzenegger, and to contemplate how this related to hypertext and the Internet:

> With its disparate connections, indeterminable authors, rapid changes, fluidity, emphasis on metamorphosis, non-human protagonists, and innumerable border crossings, the dream provides a better metaphor than do highways (too linear), webs (too structured), or communities (simply inevitable when humans get together). (Krasniewicz and Blitz, "Dreaming Arnold Schwarzenegger")

This shows Krasniewicz and Blitz's production as a second-generation hypertext, because it looks past the form of the book to see the medium of hypertext itself as the form of a dream: "dreams provide the basic metaphor of our present condition" (Krasniewicz and Blitz, "Dreaming Arnold Schwarzenegger"). They believe that the postmodern nature of our culture, in which we are interconnected and hypertextual, allows for one to approach almost anything from entirely different viewpoints and still come to the same conclusions. It is this that they have attempted to represent in their hypertext, and their use of hypertext thus does not draw on the form of the book, but on the nature of our society: "It is not that everything is dreamlike, but rather that the dream, in all its crazy manifestations and shifting tides, is the mode of structuring found in hypertext as well as being the illustration of our current cultural condition" (Krasniewicz and Blitz, "Dreaming Arnold Schwarzenegger").

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Each of these works experiments with and exploits the potential of hypertext in unique ways—Thurston, Castonguay, and Westbrook embrace hypertext's ability to bring other media forms before the reader as evidence, such as the photograph and the movie clip, as an alternative to 'hearsay' descriptions of second-hand accounts in the printed text that describe the photograph or film. Krasniewicz and Blitz's production shows the potential that hypertext presents for academic publications, although it may not have been produced if *AQ* had not promised it a publication spot.

There are other forms of academic hypertext works that have been developed, and these have primarily been hypertext editions of originally print works. Two examples produced through academic departments are Henry Churchyard's hypertext editions of Jane Austen's novels, and Murray McGillivray's Chaucer hypertext, *Book of the Duchess*, that both employ hypertext as a way to continually annotate a text—in other words, to more easily employ conventions of the book form. The use of hypertext as an annotation tool is discussed by Claire Lamont in "Annotating a Text: Literary Theory and Electronic Hypertext," and how one can "allow the mind to fantasize on the endless amount of annotation that could be stored in hypertext, as an example of opportunity or as a species of modern nightmare" (Lamont 55). Henry Churchyard's hypertext editions of Jane Austen's works may not be a 'modern nightmare,' but they are certainly an excellent example of how the ability to hyperlink to and from almost anything in a document can create a disruption of the linear reading process. Though other hypertext authors such as in the *AQ*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

special issue have praised hypertext's ability to disrupt the linear reading process, Churchyard's Austen hypertexts show how it must be used with balance to be effective.

Churchyard's "Pride and Prejudice Hypertext" exploits hypertext's potential as an annotation tool by providing five types of information throughout the text, which are described in his explanation of the hypertext's structure:

> This *Pride and Prejudice* e-text is fairly thoroughly hypertexted, but there are no cross references from one part of the main body of the text to another part. Instead, links go into or out of the main text, either to or from one of five indexes: The list of characters, the list of events in chronological order, the comments on random topics, the index to the motifs of "pride" and "prejudice", or the list of important places (with a map). (Churchyard)

These types of annotations assume that Churchyard's audience is quite unfamiliar with Austen and the contextual information surrounding her works, and suggest that his intended audience was junior undergraduate students who would like to receive more background information about Austen's works as they are reading her texts. Immediately, though, the reader may encounter a difficulty with the numerous hyperlinks and the lost linearity that the text creates. This is not just a result of using hypertext for annotation, though, as excessive annotation can cause the disruption of linearity in both print and electronic texts:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> Annotation obviously raises the question of linearity. There will be no linear reading of a text, in either book or electronic form, if the reader is breaking off repeatedly to read annotations. (Lamont 57)

This is precisely the consequence of Churchyard's hypertext, as he uses hypertext as a tool to annotate Austen's text. The lack of at least a semblance of linearity is evident from the moment a user first clicks on one of the links, as the user will often be brought to information that is not directly relevant to the text of the link, or to information that is embedded within a host of other unrelated information.

Churchyard's navigation tools, which he must have intended to be employed as an aid to using his hypertext, can be the instigators of further confusion, and his explanation of them is even more confusing:

> When you have followed a link, and the promised topic of the link doesn't seem to immediately leap into prominence, look near or at the top of the window, and then scroll back a few lines if necessary to get the immediate context of the reference. On the other hand, when there is a reference to a location near the *end* of an HTML file, some browsers (including the most frequently used graphic browsers!) will put the end of the file at the bottom of the window, with no indication of where in the window the target location is. (Complain to the software companies about these annoying browser peculiarities). (Churchyard)

Churchyard tries to amend the disruption of linearity that his hyperlinks create, but this actually produces more confusion for the user. He does use an index to divide the "Pride and Prejudice Hypertext" into sections, the further use of which

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

could have organized the entire text within an hierarchical structure in which the user could reasonably navigate—but his efforts seem to stop with his navigational tools.

Though the Austen hypertexts can be confusing and make the user feel lost, they still have some potential to provide basic contextual information to an undergraduate scholar and can also be used to search through Austen's texts. The Austen hypertexts may function largely as a supplement if you are looking for a specific location of a word in the printed text. But Churchyard himself lowers the credibility of this publication by stating, "these files make no claims to be profound scholarship or deep criticism, and necessarily sometimes reflect my own views" (Churchyard). This usefulness of Churchyard's hypertext is not enough to make it widely employed by scholars, as there are more reputable electronic editions of Austen's works available online that are also searchable. Some of these include the Electronic Text Center at the University of Virginia[23], the Alex Catalogue of Electronic Texts[24], and Carnegie-Mellon's English Server.[25] In the context of the "Pride and Prejudice Hypertext," we can learn that although annotation can be useful, and hypertext provides an easy means of producing annotation, a balanced amount of links is the key to a hypertext project's success as an intellectual resource.

---

[23] Available online at <http://etext.lib.virginia.edu/subjects/Women-Writers.html>.
[24] Available online at <http://www.infomotions.com/alex/>.
[25] Available online at <http://eserver.org/fiction/pride-and-prejudice.txt>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The hypertext edition of Chaucer's *Book of the Duchess*, by Murray McGillivray,[26] is another example of excessive linking, but it is made available for the purpose of textual analysis. Although it still uses hypertext to aid the conventions of the book medium, the hypertext *Book of the Duchess* maintains a purpose behind the amount of links employed in the text. The purpose of McGillivray's Chaucer hypertext would not be as a text for students and scholars to *read*, but as an aid to the printed text. The hypertext edition allows the user to compare different editions of Chaucer's *Book of the Duchess*, and to also compare different texts or images. For example, through the use of frames, the reader/user may compare up to six different texts and images from the Chaucer hypertext. The user may also compare the text of different editions of *Book of the Duchess* with Chaucer's French and Latin sources. In the reading edition of the actual text, though, each word is hyperlinked. Clicking on one of these terms will bring up a definition of the word in the 'glossary' frame on the right side of the page.

The structure of McGillivray's Chaucer hypertext is quite similar to some of the experimental articles of *AQ*, in its use of frames to present information sections that would normally be stored in separate sections of a book. McGillivray's Chaucer hypertext is yet another instantiation of a first-generation hypertext: it employs hypertext to augment the medium of the book, by being able to more quickly reach information about the glossary and notes. Unlike "Dreaming Arnold Schwarzenegger," it does not explore metaphors for hypertext beyond the confines of the book, and although hypertext is used to compare

[26] Available online at <http://www.ucalgary.ca/ucpress/online/pubs/duchess/Websample/mainmenu.htm>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

different editions and source texts of Chaucer's work, hypertext is not explored beyond how it can serve the already-established conventions of the book.

Interesting to note is that McGillivray's Chaucer hypertext was produced by the University of Calgary Press in a process not unlike the model proposed in Figure 3, where it was submitted electronically by the academic from a department in electronic form with the intent of being published electronically from its conception. The cost of producing, and then making available, the electronic publication was therefore minimal, and only involved the university press adding its imprint to the publication. What was left out of the publication process, though, was peer review—although publication through the University of Calgary Press would imply that the work would have gone through its standard peer-review process, it is clear from the quality of the publication that it did not undergo the same rigorous process of peer-review to which the Press would subject a for-print publication.

Hypertext has the potential to renew and rejuvenate academic publishing, but its use has been largely confined within the conventions of the book. Beyond the explorations in the change of form that hypertext presents, though, the properties of the medium itself provide very useful benefits to scholars, such as the collaborative initiative that hypertext projects require. But as Thurston points out, there are also some disadvantages to the electronic medium:

> It's somewhat of a devil's bargain, then, to commit to print observations concerning a medium as shifty as the Web. However convincingly I may argue that this is *the future of academic publishing,* that argument risks

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> being severely undermined by the not unlikely possibility that five or ten (or even two) years from now the reader of this manifesto might search for the URLs mentioned herein and find not a trace." (Thurston 251)

But although the fluidity of the Web may be a disadvantage, it is also something Thurston acknowledges that we should embrace about this new technology:

> A little more stability and restraint applied to both form and content and the medium might more closely approximate the conventions we have established for scholarly publishing. I hope such improvements will be resisted. That the Web requires continuing social relationships and can be modified indefinitely is the very thing that sets it apart from other media forms. Our task is not to make sites more like print but to explore the promise they might hold for the development of new forms of scholarly discourse. (Thurston 252)

Although there are advantages and disadvantages to this medium, we must remember that its purpose is not just to recreate that which already exists in another medium. Hypertext will be useful to us as scholars if we can be free to explore its possibilities not only within creative hypertext fiction, but within critical academic works as well. *American Quarterly* has set a good example for other academic publications about the potential of allowing hypertext works to be created within a safe environment. Through an analysis and effective peer review of these hypertext articles, we can see what has worked and how we might take hypertext into the future of academic publishing.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 3:

Discovering the Tools of the Trade: Technologies and Standards for Academic Electronic Publishing

Producing high-quality electronic academic publications relies quite strongly upon an effective peer-review system, which can be incorporated by employing the university press in the same manner as has been done with peer-reviewing print publications, but electronic publications also now require a mastery of technical resources. As Susan Hockey writes, "the production of high-quality electronic texts and resources is very labour intensive," and "it is imperative to find ways of making these resources as broadly multipurpose and reusable as possible" (Hockey 165). The effective use of the current and developing technologies to produce high-quality publications in the electronic medium thus requires two significant advancements: mastery of these tools by academics and the university press, and development of standards, best-practice guidelines, or other forms of interchange enhancement in order to facilitate as much exchange of publications as possible.

XML and Educational Repositories

An educational environment in which students could walk around with little hand-held agents and instantly retrieve information from any database connected to the Web is not too far off. In many community colleges and universities one will already see students accessing their hand-held Personal Digital Assistants (PDAs), and there have even been some educational

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

experiments in which students wandered through a wireless network set up in a forest with their PDAs in order to learn about forestry ("Pupils"). But the difference between where we are now and where we could be is that we do not yet have intelligent agents capable of retrieving information for us based on the command of a few typed or spoken keywords. We do have search engines on the Internet into which a few keywords can be typed to reveal a variety of web pages related to those keywords, but our search engines currently cannot really understand what we are trying to say to them. This does not necessarily involve computers having an 'artificial intelligence' per se, but rather 'agents' or, small personal computers, that will be able to understand what we are saying to them because the information we tell them will be semantically encoded.

Intelligent agents able to understand what we mean when we say something involves, in large part, the new vision of Tim Berners-Lee, one of the key founders of the Web. Berners-Lee envisions a Semantic Web as the next logical step in the development of the Internet. The Internet, in its current state, is largely just the architecture that could function as the backend for this second stage. The Semantic Web is another layer on top of the infrastructure we already have in place that would allow the information that is present on the web to be encoded with tags, like the tags Web browsers use to create Web pages, except that semantic tags would be called 'metadata tags'. These tags tell the machine additional information about the information object, such as a ebook, in order to give it an appropriate context.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

To show how the Semantic Web would work wonders with academic publications, let us borrow an example from Berners-Lee, James Hendler, and Ora Lassila:

> Web searches today typically turn up innumerable completely irrelevant "hits," requiring much manual filtering by the user. If you search using the keyword "cook," for example, the computer has no way of knowing whether you are looking for a chef, information about how to cook something, or simply a place, person, business or some other entity with "cook" in its name. The problem is that the word "cook" has no meaning, or semantic content, to the computer. (Berners-Lee, Hendler, Lassila)

Their vision about what the Web, or, intelligent search agents, could do with the Semantic Web in place is quite promising of a life of easy and effortless information retrieval:

> At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's *prescribed treatment* from the doctor's agent, looked up several lists of *providers*, and checked for the ones *in-plan* for Mom's insurance within a *20-mile radius* of her *home* and with a *rating* of *excellent* or *very good* on trusted rating services. It then began trying to find a match between available *appointment times* (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules. (The emphasized keywords indicate terms whose

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> semantics, or meaning, were defined for the agent through the Semantic Web). (Berners-Lee, Hendler, Lassila)

This is the important difference between what we can currently do with our handheld PDAs, and what we could do when the Semantic Web comes to life. This would not only improve searching, but allow academic publishing projects to develop archiving and delivery systems or other application development that extends the functionality of electronic resources, whether this involves cross-referencing between publications or automatic referring to related resources. The infrastructure is currently in place to make the Semantic Web come to scale through the Internet—now we need a standardization of the mark-up languages and metadata used to give context and meaning to information on the Internet.

In order for us to be able to tag on context-descriptive information to information-objects that are passed over the Internet, we need to have a standardized mark-up language. This can be accomplished through metadata and XML. Metadata is, simply, 'data about data', or, information about information. One of the remarkable features of the eXtensible Mark-up Language (XML) is that it allows us to create our own tags that are used to mark the information objects, and these tags are actually containers for information in themselves. So this means that by using XML, we can create a set of tags to describe, for example, a Web page about 'Dr. Wendy Cook' that would contain information to distinguish the name 'Cook' from preparing food and from Peter Pan's 'Captain Cook' or simply from a person.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Creating our own tags to describe a Web site is wonderful, as is filling these tags in with information such as 'general practitioner' in our 'Subject' tag or 'Dr.' in our 'Title' tag. But this alone will not make the information about Dr. Cook's Web site understandable to computers, because they will likely be unable to recognize the schema or method of tagging used by the metadata creator.

XML is vital in the development of the Semantic Web, because it allows people who are creating XML documents and metadata to use a schema (or Document Type Definition)—a set of structural definitions for the tags that are used within the XML document. Users that want to create metadata for an information object can simply use an XML schema to ensure that their metadata matches the way to create metadata as set by metadata standards and specifications.

Organizations such as the IEEE[27], the IMS[28], and offshoots such as Dublin Core[29] and CanCore[30] are working to create standards and specifications for the sets of tags that should be used to describe information objects on the Web. This crucial stage (i.e. creating a standardized schema for contextually-descriptive information about electronic publications) is where the development of a standardized set of metadata principles and practices can emerge that will allow the Semantic Web to come to scale.

The Semantic Web is particularly important to academic electronic publications and educational Web sites, especially as cut backs in educational

---

[27] Available online at <http://ltsc.ieee.org/>.
[28] Available online at <http://www.imsproject.org/>.
[29] Available online at <http://dublincore.org/>.
[30] Available online at <http://www.cancore.ca/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

funding may encourage educational institutions to share resources and to reduce spending. Linking information databases among institutions could be a benefit to all institutions involved, splitting the cost of database and server management between many institutions. But one of the more immediate ways in which institutions could benefit from sharing, and through the implementation of metadata, is in learning object repositories that would allow instructors to share course content with each other.

There are types of information objects, called learning objects, which are educational or curricular resources that can be anything from an image file to a lesson plan placed on the Web, as long as they contain the appropriate descriptive information to give them context. They are digital files (such as an image, movie, or text file) intended to be used for pedagogical purposes, which includes, either internally or via association, suggestions on the appropriate context within which to utilize the object.[31] If an instructor, for example, used an image in a lesson, it could not be considered a learning object unless it was given descriptive contextualizing information—otherwise we do not know what it is to be used for pedagogically (i.e. whether it is a religious ceremony or a celebration or a group of students reading to each other or part of a lesson about multiculturalism). Providing contextualizing information allows the image to be used as a learning object by other instructors in their courses, but this also means that those seeking to use it as a learning object must trust the contextualizing information.

---

[31] See Sosteric and Hesemeier.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Learning object repositories such as CANARIE[32] and MERLOT[33] have been providing contextualized learning objects to the e-learning community, in a repository that may be searched and accessed by individual instructors. These learning objects use metadata to describe their content and to give the learning objects semantically-understandable descriptive information that instructors can use to search for specific course-related materials, and similar repositories could be set up in academic libraries for housing electronic publications.

These learning object repositories are also a microcosmic example of what we could have with all information on the Web—images and Web pages and files that a computer can search for because they have been appended with the appropriate descriptive information. And the implications of an academic Semantic Web for scholarly electronic publishing have begun to be recognized by researchers. Collaboration with the museum as well has not been emphasized yet, but it could be invested in the same resources as libraries through digitization projects.

One of the difficulties of creating an online course is finding course materials, and learning object repositories allow this. The next step is to see how this can be implemented in other sources of search and information retrieval. If we create semantic descriptions about the information we are viewing through a computer, it changes from a random mass of information about anybody and anything, to an organized, searchable information database, somewhat like an encyclopaedia.

---

[32] Available online at <http://www.canarie.ca/>.
[33] Available online at <http://www.merlot.org/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

When Berners-Lee first demonstrated the potential of the Internet ten years ago, all he could demonstrate to his audience before this monster had begun to come to scale was its ability to link information from two storage devices across a distance.[34] The audience, and the world in general, was not much impressed until the Internet actually began to come to scale and people could see that this vision had a real existence. The Semantic Web is currently in the same state of infancy.

This is where many have criticized Berners-Lee's vision of the Semantic Web as purely a vision that could not come to fruition. The idea that computers could be 'intelligent agents' able to understand each other's information has been in the minds of people working to create robots and artificial intelligence for many years. But Berners-Lee emphasizes that the Semantic Web is *not* artificial intelligence in his paper, "What the Semantic Web can Represent":

> The concept of machine-understandable documents does not imply some magical artificial intelligence which allows machines to comprehend human mumblings. It only indicates a machine's ability to solve a well-defined problem by performing well-defined operations on existing well-defined data. Instead of asking machines to understand people's language, it involves asking people to make the extra effort. (Berners-Lee)

This extra effort involves the creation of Resource Description Framework (RDF) tags that would be included with any piece of information sent over the Web. These tags would provide a context for the information being sent and include

[34] See Dumbill.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

such information as the relationship between originally independent concepts.[35] This would allow the creation of 'intelligent agents' that are capable of correctly reading information from other agents, and without human intervention (humans would have to perform the initial work of applying the contextualizing information in the form of RDF tags, when the information material is first created, unless automated or semi-automated programs can be developed for this purpose).

The eXtensible Mark-up Language (XML) has been emerging as the new *lingua franca* of mark-up languages for the Internet and educational Web sites. The relationship between XML and Berners-Lee's new vision, the Semantic Web, has been seeping into more and more discussions involving the integration of education and technology. But as Edd Dumbill of O'Reilly Network has commented on Berners-Lee's presentation of the Semantic Web at the tenth International World Wide Web Conference: "The Semantic Web is something that a lot of people are attracted to, but find that the actual details elude them" (Dumbill). A much more general understanding of what the Semantic Web can do is important for making these 'intelligent agents' really come to life.

Metadata standardization organizations and learning object repositories will help to bring the Semantic Web into a practical existence. An image that is part of a learning object repository contains contextualizing information that helps the user find out what it is to be used for when a user searches for it within that learning object repository. But the same problems with search and retrieval as mentioned in Berners-Lee's example of the word 'cook' still applies to learning

[35] See Berners-Lee.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

object repositories, as there is still no large-scale standard for describing and retrieving these learning objects.

Learning object repositories such as MERLOT and the Campus Alberta Repository of Educational Objects (CAREO) are useful and trustworthy resources for instructors to use when searching for learning objects, but specifications such as the Canadian Core (CanCore) Learning Resource Metadata Application Profile are helping to make all information objects potential learning objects by working with the IMS to create standards for metadata mark-up and resource description framework tags. This is where the importance of XML comes in—or, a language in which the user may define their own tags and information containers in order to better describe the resource that they are creating.

The Semantic Web and Academic Publishing

An academic publishing Semantic Web has been in development recently as more academics realize the benefits of applying this technology to the academic publishing environment. Technologies are being created for this purpose, though they are still in very early stages and much research and development has yet to be invested.

Work on developing an academic publishing Semantic Web has been underway in the archiving and heritage sector. DigiCULT, a support measure for the cultural and scientific heritage sector that has established a 'technology watch' over a period of 30 months until August 2004,[36] published "Towards a Semantic Web for Heritage Resources" as one of its Thematic Issues. Though the topic of

[36] Available online at <http://www.digicult.info/pages/index.php>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

this issue addresses the question of what the Semantic Web will do for heritage institutions, it is very applicable to discussions of a Semantic Web for academic publishing. Interoperability is integral to the development of a Semantic Web, but, as with the current state of academic electronic publishing, "such interoperability is not the primary goal of heritage institutions (and intelligent software agents are not readily at hand.)" (Geser). And the goals of the library and university press are quite similar to that of the cultural and scientific heritage sector: "what the institutions are looking for are new ways of providing scholarly and non-expert users (e.g. school classes, lifelong learners) with access to their collections and related knowledge," which can be accomplished through "online collections and exhibitions that not only display objects and simple descriptions (drawn from metadata), but also allow for understanding relationships between objects (created by semantically interrelated metadata)" (Geser). The university press and the library in general face many of the same challenges as heritage institutions: the challenge is first to implement the necessary data infrastructure.[37] Unlike the conclusion of the position paper in this thematic issue, though, academic electronic publishing is not "more likely to be left behind, due in particular to the fact that for the institutions the rewards for the necessary investments are still too nebulous" (Ross). The university press and academic libraries are already behind on Semantic Web technologies, but their eventual involvement is inevitable.

According to Seamus Ross, "content as presented on the web currently is mute" (Ross). This means that "by adding descriptive information to content and

[37] See Geser.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

resources, and representing both the descriptive information and the content in well-defined, consistent, and structured ways, 'mechanised agents' could be enabled to use web information 'intelligently'" (Ross). The implications of this for academic publishing and library distribution are phenomenal. Not only can the descriptive metadata improve access to electronic publications by enhanced searching, but it can, through RDF, hold digital signatures and information about the publisher's imprint to provide both a validity and authenticity signature for the publication. And as this is automated, it can all be done without human intervention or understanding of the functioning of these technologies by the end user.

The thematic issue does highlight a very pertinent aspect of the Semantic Web that has not been as emphasized in the W3C developments: that of ontologies. In a section titled "Ontologies—the Jewels of the Semantic Web," Ross argues that "for the Semantic Web to succeed it will require not only modelling languages, such as XML, RDF, and OWL, but it will also require methodologies for extracting and defining the knowledge that is to be represented," and that "until there is such a methodology the possibilities of XML (or any other technology) as a knowledge representation language will not be achieved" (Ross). Ross uses James Hendler's definition of an ontology: "a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic" (Ross). What Ross fails to acknowledge, though, is that these ontologies can already be implemented with XML publications through the use of namespaces. Encoding

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

an XML document through an XML schema that refers to the appropriate namespace, while also employing RDF for the description of the semantic interconnections, can serve the purpose for which ontologies were created. The ontology can thereby be created through the combination of these XML technologies, and perhaps through a WYSIWYG software facility that allows the user to select an 'ontology' while automatically choosing the appropriate namespace and affixing metadata.

Standards for Academic Electronic Publishing; or, Why the TEI May Not Be Compatible with University Press Publications

The Text Encoding Initiative (TEI) is a standard for the mark-up of electronic texts that has been used for digitized republications of originally print material, but that may not apply well to original electronic publications. The TEI is useful for creating a base of tags that can be understood by different projects, but may not be beneficial for use in originally-digital publication projects. Though the TEI may actually cause problems for originally-digital publications, a standard academic electronic publishing 'namespace' would be quite beneficial along with the use of the Resource Description Framework (RDF) technology.

Founded in 1987, the Text Encoding Initiative is ideally, as defined on the TEI homepage, "an international and interdisciplinary standard that helps libraries, museums, publishers, and individual scholars represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent" ("Welcome to TEI"). Funded by substantial humanities monies from bodies such as the US

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

National Endowment for the Humanities, the European Union, the Canadian Social Science Research Council, and the Mellon Foundation, the TEI is a scholarly-based effort that began as a cooperative organization of the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing.[38] At the time when the TEI was originally conceived, it addressed a need that existed in academe due to changing proprietary software and a lack of a non-proprietary interchange format such as XML and its related technologies:

> When the Text Encoding Initiative (TEI) was originally established, scholarly projects and libraries attempting to take advantage of digital technology seemed to be faced with an overwhelming obstacle to creating sustainable and shareable archives and tools: the proliferating systems for representing textual material. These systems seemed almost always to be incompatible, often poorly designed, and multiplying at nearly the same rapid rate as the electronic text projects themselves. This situation was inhibiting the development of the full potential of computers to support humanistic inquiry by erecting barriers to access, creating new problems for preservation, making the sharing of data (and theories) difficult, and making the development of common tools impractical. ("What is TEI")

The TEI wished to create a community for communication and coordination, but this was also before the days of the Web in which common-interest group communities could form easily and 'naturally' and communicate quickly with

[38] See "What is TEI."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Web newsletters. Since the formation of the TEI, many other initiatives, such as the W3C, have developed to combat the "entrepreneurial forces which (then as now) drive information technology forward [and] would impede such integration by the proliferation of mutually incompatible technical standards" ("What is the TEI"). The results of the collaboration within the TEI's scholarly community were the TEI Guidelines that today "take the form of a substantial 1300 page reference manual, documenting and defining some 600 SGML elements which can be combined and modified in a variety of ways to create specific SGML document type definitions (DTDs) for particular purposes. This is an enormous document that can hardly be employed for practical purposes as a reference manual, and the TEI did recognize that "specific customizations of this dauntingly large document would be needed for particular user communities" ("What is TEI"). The TEI Lite is one such customization, developed "specifically to address the needs of the group forming a TEI core constituency, in electronic text centers and digital libraries" ("What is TEI"), yet TEI Lite can also be quite daunting and, if only used minimally, can even negate the purpose of employing its tagset. The TEI states of TEI Lite that:

> Very little is actually required of users, who may choose to encode as much or as little detail as they wish. The TEI encoding scheme is also highly modular. The DTD designer reviews the available "tagsets" (collections of semantically related element definitions), choosing how they are to be combined. Individual elements may be renamed, omitted, or modified, subject to some constraints, subject only to some simple

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

architectural constraints. The TEI maintains software (the "TEI Pizza Chef") which helps users construct their DTDs. ("What is TEI")

The problem with creating a standard that is both too broad and too restrictive, and then letting its users create their own tags anyway, is that it can negate the whole purpose of the standard. This is why many have suggested 'communities of practice'[39] rather than standards in order to facilitate knowledge management and sharing between institutions and organizations.[40] What seems to be in demand in the above excerpt from TEI is the XML mark-up language, and XML does address this need for both a defined tagset and flexibility on the part of the specific user community or project. But many humanities electronic text projects still only turn to their community's solution (i.e. the TEI), rather than the practical community's W3C that appropriately keeps up with the technological pace of the non-humanities world.

Perhaps an effort such as the TEI would have been augmented by cooperation with producers of original academic publications, rather than just 're-publications,' such as might be established by working with the university press. The TEI was intended to produce a standard for the encoding and interchange of electronic texts for its particular community (i.e. humanities scholars and libraries creating electronic texts), and its tagset is meant to fulfill the needs of this particular community. Yet the TEI seems to believe that the tagset is equally useful to those publishing originally electronic publications rather than

[39] "Communities of Practice are formed by groups of people who come together to learn from one another face-to-face and virtually." Definition available online at <http://www.elearningeuropa.info/doc.php?id=1483&lng=1&doclng=1>.
[40] See van Winkelen.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

digitizations of print materials: "the Guidelines are equally applicable in the creation of new resources and in the interchange of existing ones" ("Introduction"). Applying learning object metadata to the publication in XML with RDF may be more appropriate, though, and an attempt at using the TEI Guidelines in practice in an originally-electronic publication will show that this may be the case.

The deliverables of the TEI, the TEI Guidelines and the SGML-based TEI DTD (Text Encoding Initiative Document Type Definition), might be argued to accomplish the needs of the particular community to which they are aimed, but to anyone first coming to an electronic publication from the practical, tool-employing side of production, the guidelines and DTD are rather cumbersome. Though the TEI has developed an XML counterpart since its conception in SGML and is said to be developing an XML schema, many of the current technologies that we have for creating XML publications may render the TEI obsolete if it does not work more closely with its surrounding communities.

An Example of the Difficulties of Using the TEI in Practice with Electronic Publications

The Atlas of Alberta Railways, an electronic publication currently under development by the University of Alberta Press, began its electronic life through a tagset that followed the guidelines of the TEI. TEI-lite was chosen over the main TEI guidelines for its reputed practicality and more manageable size. A look at its guidelines and tagset explains why it is not very suitable for original publications that are not digitizations or electronic editions. The "starter set"

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

selection of elements that are present in TEI-lite are the "elements which almost every user should know about" ("Introduction"), but they are not a subset created for a specific user community. Perhaps if a subset were created specifically for originally-digital academic publications, with extended new tags created using XML, the TEI tagset could be more usefully accommodated for the publication of originally-digital publications.

TEI-lite was "originally conceived of as a simple demonstration of how the TEI encoding scheme might be adopted to meet 90% of the needs of 90% of the TEI user community" ("Prefatory"). Because users still found TEI-lite to be "too heavy for their needs," there was also a "Bare Bones TEI: A Very Very Small Subset of the TEI Encoding Scheme" (Sperberg-McQueen) developed by C.M. Sperberg-McQueen. This was produced in 1995, though, and has not been updated to deal with XML, unlike the main TEI Guidelines. It basically takes the information that is written in the TEI-lite guidelines, and explains their use in more simpler terms (i.e. so the tagset used still remains largely the same subset as in TEI-lite—it is just presented in a smaller, easier to understand document, with a complete example). Bad technical writing in the first version might have suggested a revision, though, rather than the creation of a new subset based on TEI Lite.

In producing the *Atlas of Railways in Alberta* (ARA), the U of A Press therefore chose to use the TEI-lite guidelines, while referring to the "Bare Bones TEI" document when a simpler explanation of the TEI-lite's tags was required. The software employed for tagging the ARA was XMLSpy, and although a DTD

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

generated by the TEI could have been used instead of a more up-to-date schema in order to define the structure of the document, the 'Pizza Chef' software used on the TEI site to generate a DTD from the TEI guidelines would only generate errors, necessitating the creation of a structuring document from scratch. The UAP therefore proceeded to create a schema document in XMLSpy for the project following the tagset proposed in TEI-Lite.

A number of problems with inappropriate tags were found in employing the TEI tagset for an originally digital publication. The main problem was the focus throughout the entire TEI Guidelines (both full and 'lite') on reproducing the printed page in electronic form, which was what the TEI was originally conceived to do in its primary application. After giving a brief overview of the history of the TEI Guidelines and purpose of TEI-Lite, there is an example on how a printed text is marked-up for electronic publication. There is an immediate assumption that the user of TEI will want to produce an electronic 'edition' of a print text:

> We begin with a short example, intended to show what happens when a passage of prose is typed into a computer by someone with little sense of the purpose of mark-up, or the potential of electronic texts. In an ideal world, such output might be generated by a very accurate optical scanner. It attempts to be faithful to the appearance of the printed text, by retaining the original line breaks, by introducing blanks to represent the layout of the original headings and page breaks, and so forth. Where characters not available on the keyboard are needed (such as the accented letter a in faàl

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> or the long dash), it attempts to mimic their appearance. ("Short Example")

The TEI asserts its use for preserving the appearance of the printed text even in this first example, and the entire Guidelines are structured this way. The mid-90s sentiment of mimicking the printed page in the electronic publication is highly emphasized in the Guidelines, yet a project wanting to create an originally-digital publication using the TEI Guidelines will not share this sentiment at all—this seems to be overlooked throughout the Guidelines. For example, the producer of an originally-digital publication will not be concerned about whether "page numbers and running titles are intermingled with the text in a way which makes it difficult for software to disentangle them" ("Short Example"); rather, the producer of the originally-digital publication would consider how the publication functions most practically in the electronic form, and would not place print layout elements in the text if this would be inappropriate in the electronic medium. Many tags designed for the purpose of mimicking the printed text are therefore inapplicable; the page divisions tag <pb> would not be used, along with many others such as <lb>, which "marks the start of a new (typographic) line in some edition or version of a text" ("Page").

The structure of the TEI text proposed in the Guidelines is also problematic for originally-electronic publications, and assumes that the encoder will be transcribing an originally-print text. The TEI Header, the metadata for the publication, such as author and title, is useful, but may not be compatible with machine-understandable metadata interchange formats that are in development.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

As TEI-Lite itself states, "The TEI header provides information analogous to that provided by the title page of a printed text" ("Structure"). Tags that deal with the electronic properties of the document may need to be included (such as its file size) that the TEI fails to even consider. The text of the TEI is also structured as a print book, with front, body, and back matter:

```
<TEI.2>
    <teiHeader> [ TEI Header information ]
</teiHeader>
    <text>
        <front> [ front matter ... ]   </front>
        <body>  [ body of text ... ]   </body>
       <back>  [ back matter ...  ]   </back>
    </text>
</TEI.2>
```

("Structure").

To follow the print layout in the electronic form is not practical, because even if the publisher would like to also produce it in print, this can be done with XSLT stylesheets rather than encoding the document's print or electronic formatting in the archival XML. It is fortunate that TEI does allow some flexibility with its , <div1>, <div2>...etc. tags, as these structural division tags may be used to define a 'Part', 'Section', 'Chapter', etc. through attributes, and these categories do allow for a structural flexibility to make divisions suitable for the electronic environment.

In short, the TEI needs to undergo reworking and updating in order to be practical in today's academic electronic publishing environment. If the TEI worked in greater contact with the W3C, it might be more applicable to current electronic publishing projects. Perhaps there is hope for such a large set of tags as the TEI, but the difficulty is that it requires such an investment of time to make

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

smaller subsets such as the 'TEI-lite' or 'Bare Bones TEI' (such as if one were to make one specifically for university press publications), when we now have technologies that can translate tags between systems. These tags can also be machine-understandable and leave the metadata input process up to the machine rather than the user.

Namespaces, XSLT, and RDF have the ability to translate between the tag sets of different projects. It is the equivalent of employing translators at international gatherings and governmental organizations, rather than making everybody first learn English or the latest 'lingua franca' (which tends to change more rapidly in the machine-language environment than in the natural language world). These and other technologies have been underway through the World Wide Web Consortium (W3C).

The W3C is a "forum for information, commerce, communication, and collective understanding" that "develops interoperable technologies (specifications, guidelines, software, and tools)" ("World"). It was created in October 1994 by Tim Berners-Lee "inventor of the Web," at the MIT Laboratory for Computer Science in collaboration with CERN, the particle physics laboratory where the Web originated. The W3C was founded to "lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its operability" ("About"). It has "earned international recognition for its contributions to the growth of the Web" and is international in scope, with over 450 member organizations ("About"). The W3C therefore involves quite a large

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

community in its development of standards and technologies for the Web, and this is one of the reasons it has been so successful and continues to grow.

A look at the W3C's discussions and technologies, perhaps alongside the recommendations of the TEI, is a smarter move for publishers in our contemporary environment. The future of academic publishing lies in putting to use in the academic environment the technologies being developed through the W3C. The W3C has published more than forty recommendations since its inception, including (X)HTML, CSS (Cascading Style Sheets), XML 1.0, XML Namespaces, XSLT (the eXtensible Stylesheet Transformation Language that works with XML), XPath and XLink for referencing between documents, PNG (Portable Network Graphics) and SVG (Scalable Vector Graphics), RDF (the Resource Description Framework Model and Syntax—the "first Recommendation on which the Semantic Web will be built," providing a "standard vocabulary and encoding techniques to attach metadata to any resource on the Web"), XML-Signature and XML Encryption ("important steps towards the Web of Trust," as they "define an XML-based framework to add digital signatures to, and to encrypt resources on the Web"), Web Accessibility Guidelines, and XML Schema that provides "functionalities above and beyond what is provided in DTDs and are essential in defining complex XML applications" ("About").

Part of the whole difficulty with the TEI is its emphasis on 'rich tagging', which brings in the issue of the individual encoder(s) deciding what should be marked up within a document. One scholar's use of a document's 'metadata' (essentially, what SGML and XML encoders using TEI are trying to invest in the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

document), and the elements that they consider important to define, will not be the same for every other scholar or project that wishes to use the same text. The TEI itself acknowledges this selectivity in encoding with TEI:

> The decision to focus on Brontë's text, rather than on the printing of it in this particular edition, is one aspect of a fundamental encoding issue: that of selectivity. An encoding makes explicit only those textual features of importance to the encoder. It is not difficult to think of ways in which the encoding of even this short passage might readily be extended. ("Short")

Any text that is tediously encoded using TEI may only be useful to the community that has encoded it—whether it is an English department, a library, or a particular scholarly society. The solution lies in putting to use some of the technologies developed through the W3C and other standards bodies such as the ISO, in combination with automating metadata technologies such as the "IMS Generator for the Masses" that "allows for the creation of IMS-conformant XML files of any and all data held by an institution, regardless of 'native' format and/or structure" (Hai). The limitations of 'hand collected metadata' have also been recognized, such as by D'Arcy Norman in the "On Search: Metadata" Weblog, in which he argues:

> If you give them too many fields (like, say, maybe IMS LOM?) they just won't do it. Or, even worse, they'll do a crappy job. Even CanCore isn't small enough to be done efficiently and effectively. Heck, even DublinCore is too big for most users to regularly enter all fields completely. ("On Search")

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

As with the TEI (and TEI-lite), metadata will not be used appropriately or at all by end-users unless there is some form of automated metadata generation in place.

Using XML Namespaces, XSLT, XML Schema, and RDF with Academic Publications

XML is the core of a family of technologies that make it possible for each community of users to "design languages that suit their particular needs and integrate them harmoniously into a general infrastructure based on XML" ("About"). This means that prescribed sets of mark-up tags (whether extremely large as in the TEI, or smaller as in TEI-lite) are no longer necessary. Each community of users, whether an academic department's digitization project of an originally-print poem, a business' personnel database, or an online ebook publisher, can encode its electronic documents with its own defined tags, and still make these understandable to the broader community.

XML Namespaces are a key technology behind this ability. They "provide a simple method for qualifying element and attribute names used in Extensible Markup Language documents by associating them with namespaces identified by URI references" ("Namespaces"). The motivation behind the use of namespaces is a natural outgrowth of this community that puts practicality and efficiency as high priority:

> We envision applications of Extensible Markup Language (XML) where a single XML document may contain elements and attributes (here referred

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

> to as a "markup vocabulary") that are defined for and used by multiple software modules. One motivation for this is modularity; if such a markup vocabulary exists which is well-understood and for which there is useful software available, it is better to re-use this markup rather than re-invent it. ("Namespaces")

Namespaces therefore developed to provide a practical solution to the redundancy of each particular community reinventing mark-up schemes, while also wanting to avoid a prescribed tag set. Namespaces also attempt to make the mark-up more automated and understandable between software and machines, avoiding the time-consuming process of hiring employees to tag the appropriate elements in each document being marked-up. An XML namespace is "a collection of names, identified by a URI reference, which are used in XML documents as element types and attribute names" ("Namespaces"), whereas URI references are used as a unique identifier to identify this namespace. An academic publishing project can therefore include the appropriate 'namespace prefix' and 'local part' that selects a namespace and produces an identifier that is universally unique, therefore allowing the project to define whether the tag <author> is referring to the author of a book, the editor, or the author of the book's metadata, while also understandable by software without any intervention or translation process necessary by human beings.

A technology that works along with XML Namespaces for automated interoperability of electronic publications is the eXtensible Stylesheet Language Transformation (XSLT). Combined with its companion XPath, an XML file can

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

be "transformed into any other type of XML file, for example into XHTML or SVG, for the purpose of presentation" ("About"). This means that if a document is encoded in XML, an XSL file can be created to transform this document into HTML for Web-viewing, or to another XML file that uses any other set of defined tags. For example, if the University of Kentucky Press decides to create an XML publication using a set of tags defined by one XML Namespace, and the Electronic Text Centre of the University of New Brunswick decides that it would like to change the tags that surround the data to their own set of tags, this can be done very quickly using XSLT. And if the Electronic Text Centre decides that it would like to present the data in the XML document in a way that matches the formatting and presentation style of its other publications, another XSL file can be created to present this file in their format while still preserving the original XML file as an archival document.

With the W3C's development of XML Schemas, a further progression was made in the automation of electronic publications. A schema, in general, is "a collection of classes (typically authored for a specific purpose or domain)," that are "organized in a hierarchy, and offer extensibility through subclass refinement" ("Resource"). The schema class system matches that which is used in many object-oriented programming and modelling systems.[41] The XML Schema language developed through the W3C is itself represented in XML 1.0 and uses namespaces, and "substantially reconstructs and considerably extends the capabilities found in XML 1.0 document type definitions (DTDs)" ("XML

[41] See "Resource."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Schema Part I"). The goals of the W3C XML Schema working group were to address some of the shortcomings of DTDs:

> The XML 1.0 specification defines the concepts of well-formedness and validity; it is very simple to check a document for well-formedness, while validation requires more work but allows the user to define more powerful constraints on document structure. XML validity requires that a document follow the constraints expressed in its document type definition, which provides the rough equivalent of a context-free grammar for a document type. For some uses, applications may need definitions of markup constructs more informative, or constraints on document structure tighter than, looser than, or simply different from those which can be expressed using document type definitions as defined in XML 1.0. There is also a widespread desire to allow markup constructs and constraints to be specified in an XML-based syntax, in order to allow tools for XML documents to be used on the specifications. ("XML Schema Requirements")

Structural schemas were created to go beyond DTD functionality while still creating a mechanism somewhat analogous to DTDs for constraining document structure.[42] The specific goals beyond DTD functionality were: integration with namespaces, definition of incomplete constraints on the content of an element type, integration of structural schemas with primitive data types, and to more

[42] See "XML Schema Requirements."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

easily define inheritance and 'kind-of' relationships between content models.[43] The schema language can therefore be used to "describe a class of XML documents by using these constructs to constrain and document the meaning, usage and relationships of their constituent parts: datatypes, elements and their content, attributes and their values, entities and their contents and notations," and may also "provide for the specification of implicit information such as default values" while also "documenting their own meaning, usage, and function" ("XML Schema Requirements"). Schemas are therefore more equipped to handle datatypes than DTDs:

> The XML 1.0 (Second Edition) specification defines limited facilities for applying datatypes to document content in that documents may contain or refer to DTDs that assign types to elements and attributes. However, document authors, including authors of traditional *documents* and those transporting *data* in XML, often require a higher degree of type checking to ensure robustness in document understanding and data interchange. ("XML Schema Requirements")

Schemas are also much more practical than DTDs in that "in order to create a schema slightly different from an existing one it is not necessary to 'reinvent the wheel' but one can just provide incremental modifications to the base schema" ("Resource"). The future of academic electronic publishing therefore depends upon the move away from DTDs to schemas, as schemas are able to "define, describe, and catalogue XML vocabularies for classes of XML documents" ("XML Schema Requirements"), and not just for a specific document type.

---

[43] Ibid.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The W3C provides usage scenarios representative of the "problem space to be addressed" in the development of XML Schemas that are immediately applicable to the university press ("XML Schema Requirements"). For distribution of information through publishing and syndication services, which involve "collections of XML documents with complex relations among them," structural schemas "describe the properties of headlines, news stories, thumbnail images, cross-references, etc.," while document views can be under the control of different versions of a schema ("XML Schema Requirements"). In electronic commerce and transaction processing, "libraries of schemas define business transactions within markets and between parties," and a "schema-aware processor [can be] used to validate a business document, and to provide access to its information set" ("XML Schema Requirements"). Traditional document authoring and editing can also be governed by schema constraints, as a schema definition can be used to guide an author in the development of documents.[44] The W3C gives as an example the writing of technical service manuals for a wide-body intercontinental aircraft, where an application that involves schema-constraints can be used to "ensure that the author always knows whether to enter a date or a part-number, and might even ensure that the data entered is valid" ("XML Schema Requirements"). Using XML Schemas instead of DTDs in publishing will also foster the exchange of information between projects, as "such exchange is currently hampered by the difficulty of fully describing the exchange data model in terms of XML DTDs" ("XML Schema Requirements"). According to the W3C, when "the exchange data model is represented by the more

[44] See "XML Schema Requirements."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

expressive XML Schema definitions, the task of mapping the exchange data model to and from application internal data models will be simplified" ("XML Schema Requirements"). This also ties in with the use of metadata with electronic publications, and would automate the process of transcribing and transferring metadata about a publication and between projects.

The W3C's Resource Description Framework (RDF) Model and Syntax specification builds on the other XML-family technologies by allowing metadata to be incorporated automatically into electronic publications. RDF "provides a standard vocabulary and encoding techniques to attach metadata to any resource on the Web," and is the "first Recommendation on which the Semantic Web will be built" ("About"). Researchers that address the current Web as built for human consumption will address the next-generation 'Web' as that which would surpass the currently machine-readable but not "machine-understandable" data on the WWW ("Resource"). The lack of machine-understandability of information on the Web makes it "very hard to automate anything on the Web, and because of the volume of information the Web contains, it is not possible to manage it manually" ("Resource"). This is also true of academic electronic publications—the current volume that we have cannot be classified or searched, and thus cannot be managed. If metadata were used to describe the data contained on the Web, this may make the information on the Web searchable and manageable, but the key is that this information must be *automated*, although this may not always be possible. Resource Description Framework is a "foundation for processing metadata; it provides interoperability between applications that exchange

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

machine-understandable information on the Web," and, as such, is vital to the development of an effective searching and information retrieval mechanism, as it "emphasizes facilities to enable automated processing of Web resources" ("Resource"). The W3C defines its use:

> RDF can be used in a variety of application areas; for example: in *resource discovery* to provide better search engine capabilities, in *cataloging* for describing the content and content relationships available at a particular Web site, page, or digital library, by *intelligent software agents* to facilitate knowledge sharing and exchange, in *content rating*, in describing *collections of pages* that represent a single logical "document", for describing *intellectual property rights* of Web pages, and for expressing the *privacy preferences* of a user as well as the *privacy policies* of a Web site. RDF with *digital signatures* will be key to building the "Web of Trust" for electronic commerce, collaboration, and other applications. ("Resource")

A chief advantage of RDF is that it is XML and all of the aforementioned technologies of the XML family. It also uses a schema-based system, so that "through the sharability of schemas RDF will support the reusability of metadata definitions" ("Resource"). Due to RDF's "incremental extensibility, agents processing metadata will be able to trace the origins of schemata they are unfamiliar with back to known schemata and perform meaningful actions on metadata they weren't originally designed to process" ("Resource"). RDF functions by describing the resource using three object types: resources, which are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the thing(s) being described by the RDF expression (such as a Web page); properties, which are a specific aspect, characteristic, attribute or relation used to describe a resource; and statements, which are a specific resource together with a named property plus the value of the property for that resource.[45] The RDF can actually be encoded as an XML Schema, and integrated fully with XML Schemas used for archival and presentation translation of XML documents, along with XML Namespaces for interoperability of XML publications and appropriate searching and identification.

With all of these technologies, the foundation is present to develop academic electronic publications that are interoperable with the library, the academic department, the university press, the user accessing the publication, and the world at large. For the academic that is comfortable with only Microsoft Word and basic Web-surfing for checking email (as is common), there are developments to make these WYSIWYG[46] technologies compatible with XML.[47] With technologies available for the academic or author that do not require any knowledge of mark-up or XML technologies, the information entered can be distributed through the imprint of the university press merely by applying the press' in-house stylesheet to the information (and perhaps with some slight alterations to make the publication have a unique look, not unlike the Content Management System currently being used in the University of Alberta Web

[45] See "Resource."
[46] WYSIWYG stands for "What You See Is What You Get"—it is an acronym used to describe software that shows the user the end-result rather than the mechanics behind the production of that result.
[47] Microsoft is currently working to make MS Word XML-compatible.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Project[48]). Metadata developments are also still in progress, though, and Adobe has also just released a non-proprietary Extensible Metadata Platform (XML) for embedding metadata directly files associated with applications such as Photoshop, Illustrator, and InDesign, and because it is a non-proprietary standard it can be used for other application files as well.[49] Metadata will also help students accessing the resource to change its look, perhaps to suit their own accessibility needs by providing larger text, different image colours, translating the content into another language or to output it as a voice recording, all without changing the original 'archival' XML file. The technologies for this to come to life are there—they just have to be recognized by the academic world, and put to use.

[48] Information about the U of A Web Project is available online at <http://www.uofaweb.ualberta.ca/>.
[49] Information about the Extensible Metadata Platform is available at <http://www.adobe.com/products/xmp/main.html>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Chapter 4:

Bending the Rules of the Game: Access, Archiving, and the Pragmatics of Copyright

With the establishment of credibility for academic electronic publications and the technological framework to produce these publications from university presses, there are still many considerations that could drastically alter the direction that academic electronic publishing will take. Issues of access, archiving, and copyright have been under debate for some time now, and will continue to grow with the development of new access technologies, new relationships between university presses and libraries, and with new copyright legislation for the digital world. These three intermingling areas are very important to contemporary scholarship and the university press.

Open Access

Issues of access have been at the forefront of contemporary academic publishing for some time now, and although scholarly communities have formed in the electronic environment to discuss this issue. 'Open Access' is not just 'no-cost', and the 'not-paying' that open-access might imply is not a threat to the university press—in fact, it will likely work to save the university press. Many open access initiatives and communities are banning the use of 'toll' journals—journals that charge the user to view an article online. With the availability of credible and free electronic publications, university presses may be forced to waive a toll access to electronic publications if they want readers to view them—at least by all but the most prestigious of university and private presses. It is a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

situation in which publishers will likely not be able to win out over accessibility, as content providers that want to charge a fee are now faced with a movement that seeks to provide free and open access for all, and the university press will eventually have to cooperate with this or face even more financial hardships. And some open access communities are providing exactly the solution that the university press needs.

The Budapest Open Access Initiative, developed by the scholarly community to "use... resources more productively to aid the transition to open access and to make open-access publishing economically self-sustaining" ("Budapest1") sums up the possibilities of open access for academic electronic publishing in its mandate:

> An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds. Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge. ("Budapest2")

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Breaking down these traditional publication access barriers is the purpose of the Open Access Initiative, and although its mandate may have sounded utopian or impossible in the pre-Internet world, publishers experimenting with different dissemination models in the electronic environment are already showing us that this is possible.

BioMed Central

BioMed Central is one of the most important examples of a successful academic publishing system that is not only credible and widely-used, but is completely free and open-access. It arose in the scientific community "based on the view that open access to research is central to rapid and efficient progress in science and that subscription-based access to research is hindering rather than helping scientific communication" ("BioMed Central: About"). BioMed Central recognizes the importance of peer-review in providing credibility for the thousands of articles housed in its journals, and emphasizes peer-review in many locations of the BioMed Central site. Because of its stringent emphasis on peer-review, BioMed Central has become an imprint in itself, now known to house credible online academic publications. And it provides these free to all, no matter what institution a user is affiliated with.

Part of BioMed Central's model is the retaining of copyright by the author, and the immediate archiving in PubMed Central, a repository of full-text, peer-reviewed articles from life science journals.[50] PubMed is the U.S. National Library of Medicine's digital archive of life sciences journal literature

[50] See "Open Access Now."

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

("PubMed") and as such is "not a publisher itself, but instead relies on the willingness of publishers to deposit copies of articles they have published" ("Open Access Now"). This aspect of BioMed Central—its cooperation with the archival body PubMed central, is integral to its success, and a working relationship not unlike this could be set up between the university press and the library. As with BioMed Central, though, it would be integral for the university press to either maintain that if the copyright remains with the author, that the author would deposit the publication in the library's archive, or that the author would allow the university press to have permission to archive the publication in such archive resources.

An alternative model to BioMed Central could be an institutional content management system (CMS) in which academics would deposit their publications online, rather than going through their own self-archiving, or having to pay $500 to publish through BioMed Central. One might argue that scientific journals article authors may have more funding from their institutions to publish research, as part of research grant funding, though this kind of money may not be as available in the humanities and social sciences. In this aspect, BioMed Central's publishing model is essentially that of a vanity press, but this is, as well, not unlike the process that already occurs at many university presses, where many publishing projects will not be taken on unless it is of a nature that the university press can acquire grant funding to publish. An alternative to subsidy publishing might be a CMS integrated with software such as EndNote. BioMed Central already offers manuscript submitters a $50 rebate off the article processing fee if

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

they use EndNote or other uniform software. A university-wide self-publishing/archiving CMS could be established for the publication of academic articles. But key to the acceptance of such a system would be incorporation with the imprint of the university press associated with the institution, which would ensure that the publications on the CMS have undergone or are in the process of being rigorously peer-reviewed. Other programs, such as "Open Journal Systems,"[51] are available that assist with every stage of the refereeing process, as well as "GNU eprints 2" for creating online archives.[52] There are also programs such as TUSTEP,[53] that function as a "professional toolbox" by providing software for processing the textual data, with a strong focus on humanities applications ("TUSTEP"). The Fedora Project,[54] an "Open Source Digital Repository Management System," is especially vital to the integration of the university press, the library, and academic departments in creating electronic publications, as it is "designed to be a foundation upon which interoperable web-based digital libraries, institutional repositories and other information management systems can be built, [and] demonstrates how distributed digital library architecture can be deployed using web-based technologies, including XML and Web services" ("Fedora").

Many of these developments, such as BioMed Central and Fedora, are being closely watched by scholarly communities devoted to developing free and open access to publicly funded information. Peter Suber, Director of the Open

---

[51] Available online at <http://www.pkp.ubc.ca/ojs/>.
[52] Available online at <http://software.eprints.org/>.
[53] Available online at <http://www.uni-tuebingen.de/zdv/zrlinfo/tustep_eng.html>.
[54] Available online at <http://www.fedora.info/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Access Project, "Free Online Scholarship Newsletter," now published as the "SPARC Open Access Newsletter,"[55] has been a key provider of information on copyright and open access issues to the academic community. SPARC is "an alliance of universities, research libraries, and organizations built as a constructive response to market dysfunctions in the scholarly communication system" that have "reduced dissemination of scholarship and crippled libraries" ("SPARC"). SPARC, as well as the Budapest Open Access Initiative, has been integral to spreading awareness of issues of free and open access, and has played an important role in the development of many open access technologies.

The Budapest Open Access Initiative specifically promotes open access on the level of the individual scholar by allowing scholars to sign on to the initiative that emphasizes a commitment to self-archiving. According to the BOAI, "For an individual, signing indicates a commitment to open access for one's own research (by self-archiving what one publishes in toll-based journals and/or by publishing in open-access-journals)" ("Sign"). For an institution, signing means "a commitment to open access by supporting institutional self-archiving and/or open-access-journals" ("Sign"). BioMed Central is itself highly involved in these initiatives and produces a newsletter on its site, *Open Access Now*, devoted to the promotion of free and open access academic publications.

Print-on-demand technologies are currently in development that might help both individual academics archiving their work and university presses. CafePress[56] has launched a 'Sell Online' technology that allows writers to publish

[55] Available online at <http://www.arl.org/sparc/home/index.asp?page=0>.
[56] Available online at <http://www.cafepress.com/cp/info/sell/books.aspx>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and sell their own books. Unlike BioMed Central or vanity presses, though, the author does not have to invest anything, not even an article processing fee. The content only needs to be formatted and then submitted through the online system, and colour-covered, black-and-white textbooks will be created "using true print-on-demand technology" ("Self Publish"). If print copies of a book are required, scholars (and university presses) can print copies as they are ordered, with "no setup fees or minimum quantities." For a $7 flat binding fee or for $0.03 per page, Perfect Bound books can be individually printed and one can choose the price that the book will retail for, and wait for the orders.[57] There are also other systems available for easily creating book publications online, such as with "Original Works"—a marketing system designed specifically for academics that will transform manuscripts, lab manuals, or course notes into textbooks.[58] Self-publishing is beneficial for academic electronic publications if the peer-review processes and imprint have been integrated into the Web environment, but for publications that are produced using reputable peer-review processes, if the authors retains copyright these articles may be self-published by the author in order to make them more freely available to others. Integral to the development of academic self-publishing and archiving is the retaining of copyright with the author, rather than 'selling' these rights to a publisher. BioMed Cental's model, along with models proposed by open access movements, all require the retaining of copyright with the author in order to function.

---

[57] See "Self Publish."

[58] Available online at <http://www.xanedu.com/originalworks/index.html>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This is a pivotal point in the future of academic publishing. We have already seen libraries become non-institutional in the sciences with BioMed Central and PubMed Central, and the W3C has developed its own 'WWW Virtual Library' that provides links to resources available on the Web in a variety of subject areas.[59] If institutional libraries, academic departments, and the university press work together to tackle the initial technological hurdles, and to cut costs on the dissemination and archiving of publications, academic institutions will be able to provide useful resources to students and scholars in the digital age.

[59] Available online at <http://vlib.org/Overview.html>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Conclusion

The field of academic publishing is not stable, and will likely remain in flux for some time. The implementation of technologies such as XSLT and RDF with metadata in academic publications will depend upon whether copyright legislations hinder a more open network of sharing publications. If advancements in open access initiatives press for a greater ability to transfer publications between users, and if these publications are more easily searchable and from trusted sources, we may see the technologies discussed in Chapter 3 come into more mainstream academic use.

Future questions to be addressed in the field of academic electronic publishing will likely involve the future of learning—how will learning change with a greater integration of online sources into teaching? If more publications become open-access to all, perhaps the same could be established for university course content? The future of learning and the future of academic publishing are tightly intermixed, and would likely follow similar pathways of development. Incorporating properties from the non-online classroom environment into the Web, such as the 'lecture', would likely fail in the electronic medium. The use of textbooks is also linked with the university press—will Web users begin to develop patterns of learning and reading that would render the traditional class textbook obsolete? If so, would university presses become involved in the peer-review of small learning modules, that can be customized by instructors for individual classes and integrated with other learning modules, rendering the textbook as we now know it impractically large and static?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In developments with peer-review of electronic publications, one might ask why scholarly institutions do not just train students on research methods and provide them with the ability to discern what are proper and credible sources. But the library has functioned as the resource where students typically go for this information, and although it is currently lagging behind, with the help of the university press it can be integrated into the Web as a useful gateway to credible academic information that a student can access at home or anywhere there is connection to the Internet.

We are already employing computers in humanities research: when a student types in keywords specific to their research, whether in a library search engine or a general Internet search engine such as Google, and all that we need to do is provide a means of having the information students are looking for seamlessly available when they type in the search. The technologies are currently available, and experimental examples have been produced. It is simply a matter of now putting these technologies to use for academic electronic publishing.

The academic world will continue to change due to technologies based on the computer, and this will require students and scholars that are flexible enough to adapt to these new developments. This can be provided by training students (as future scholars) to be adept at studying the use of computers in various academic disciplines and applying computer technology to their research. New technological developments can be integrated into the academic community through these scholars that connect the academic department, university library, and university press to developments in the technological sector and, in turn, find

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ways to benefit humanities research with these endeavours. It is these new scholars adept at both humanities research and applying computing technology that will be able to tackle these issues on academic electronic publishing and the use of computers in the humanities, in turn leading humanities programs into the digital future.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Works Cited

"A Short Example." Text Encoding Initiative. 7 Aug. 2002. 13 Aug. 2003 <http://www.tei-c.org/lite/U5-eg.html>.

"About the World Wide Web Consortium." W3C. Mar. 2000. 13 Aug. 2003 <http://www.w3.org/Consortium/>.

Baker, John F. "University Presses: 'Sleepless in St. Louis'." Publishers Weekly. 7 Jul. 2003. 27 Aug. 2003 <http://publishersweekly.reviewsnews.com/index.asp?layout=article&articleid=CA309330&display=breakingNews&publication=publishersweekly>.

Berners-Lee, Tim. "What the Semantic Web Can Represent." W3C. 17 Sep. 1998. 13 Aug. 2003 <http://www.w3.org/DesignIssues/RDFnot.html>.

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." Scientific American.com. May 2001. 13 Aug. 2003 <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.

Best, Michael, and Elizabeth Grove-White. "Copyright Issues." The Credibility of Electronic Publishing: A Report to the Humanities and Social Sciences Federation of Canada. Text Technology 11.1. (2002). 13 Aug. 2003 <http://web.mala.bc.ca/hssfc/Final/Copyright.htm>.

"BioMed Central: About Us." BioMed Central. 13 Aug. 2003 <http://www.biomedcentral.com/info/>.

"Budapest Open Access Initiative1." Budapest Open Access Initiative. 13 Aug. 2003 <http://www.soros.org/openaccess/>.

"Budapest Open Access Initiative2." Budapest Open Access Initiative. 13 Aug. 2003 <http://www.soros.org/openaccess/read.shtml>.

Burk, Alan, James Kerr, and Andy Pope. "Archiving and Text Fluidity/Version Control." The Credibility of Electronic Publishing: A Report to the Humanities and Social Sciences Federation of Canada. Text Technology 11.1. (2002). 13 Aug. 2003 <http://web.mala.bc.ca/hssfc/Final/Archiving.htm>.

Carlson, Scott. "Web-Loving Students Can Be Prodded to Cite Peer-Reviewed Works in Term Papers, Study Suggests." The Chronicle of Higher Education. 6 Feb. 2003. 13 Aug. 2003 <http://chronicle.com/free/2003/02/2003020601t.htm>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Castonguay, James. "The Spanish-American War in U.S. Media Culture." American Quarterly: Hypertext Scholarship in American Studies. 51.2 (1999). 1 Dec. 2002 <http://chnm.gmu.edu/aq/>.

Churchyard, Henry. "Pride and Prejudice Hypertext." The Republic of Pemberley. 1996. 1 Dec. 2002 <http://www.pemberley.com/janeinfo/pridprej.html>.

"Digital Scholarship." University of Virginia Press Electronic Imprint. 20 Mar. 2003. 13 Aug. 2003 <http://www.ei.virginia.edu/digitalscholarship.html>.

Dumbill, Edd. "Berners-Lee and the Semantic Web Vision". O'Reilly XML.com. 6 Dec. 2000. 31 Aug. 2002 <http://www.xml.com/pub/a/2000/12/xml2000/timbl.html>.

"Electronic Imprint at the University of Virginia Press." University of Virginia Press Electronic Imprint. 28 Mar. 2003. 13 Aug. 2003 <http://www.ei.virginia.edu/>.

"Electronic Text Center." Electronic Text Center. University of Virginia Library. 13 Aug. 2003 <http://etext.lib.virginia.edu/>.

"ETC Research and Development." Electronic Text Centre. University of New Brunswick Libraries. 13 Aug. 2003 <http://www.lib.unb.ca/Texts/projects.html>.

"Fedora Home Page." The Fedora Project. 13 Aug. 2003 <http://www.fedora.info/>.

Gants, David L. "Studies in Bibliography Online." Humanist Discussion Group Vol. 9, No. 725. 14 Apr. 1996. 13 Aug. 2003 <http://lists.village.virginia.edu/lists_archive/Humanist/v09/0695.html>

Geser, Guntram. "Introduction and Overview." DigiCULT Thematic Issue 3: Towards a Semantic Web for Heritage Resources. May 2003. 13 Aug. 2003 <http://www.digicult.info/pages/publications.php>.

Guédon, Jean-Claude. "Peer Review and Imprint." The Credibility of Electronic Publishing: A Report to the Humanities and Social Sciences Federation of Canada. Text Technology 11.1 (2002). 13 Aug. 2003 <http://web.mala.bc.ca/hssfc/Final/PeerReview.htm>.

Hai, Kelvin, David Palmer and Chris Dunlop. "An IMS Generator for the Masses." ARIADNE. 5 Aug. 2003. 13 Aug. 2003 <http://www.ariadne.ac.uk/issue36/martini/>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hayles, N. Katherine. Writing Machines. Cambridge, Mass.: MIT Press, 2002.

Hockey, Susan. Electronic Texts in the Humanities. New York: OUP, 2000.

"Introduction." Text Encoding Initiative. 7 Aug. 2002. 13 Aug. 2003 <http://www.tei-c.org/lite/U5-Intro.html>.

Kasdorf, William E. "Introduction." The Columbia Guide to Digital Publishing. New York: Columbia UP, 2003. 13 August 2003 <http://www.digitalpublishingguide.com/c3/guide.pl>.

Krasniewicz, Louise and Michael Blitz. "Dreaming Arnold Schwarzenegger." American Quarterly: Hypertext Scholarship in American Studies. 51.2 (1999). 1 Dec. 2002 <http://chnm.gmu.edu/aq/>.

Krasniewicz, Louise and Michael Blitz. "The Original Call For Proposals: Hypertext Scholarship in American Studies." American Quarterly: Hypertext Scholarship in American Studies. 51.2 (1999). 1 Dec. 2002 <http://www.sscnet.ucla.edu/ioa/arnold/arnoldwebpages/Anchor-THE-10944>.

Krasniewicz "How Did You Get Here?" American Quarterly: Hypertext Scholarship in American Studies. 51.2 (1999). 1 Dec. 2002 <http://www.sscnet.ucla.edu/ioa/arnold/arnoldwebpages/dhaessay.htm>.

Krasniewicz "Part IV: Hyperdreams." American Quarterly: Hypertext Scholarship in American Studies. 51.2 (1999). 1 Dec. 2002 <http://www.sscnet.ucla.edu/ioa/arnold/arnoldwebpages/dhaessay.htm>.

Lamont, Claire. "Annotating a Text: Literary Theory and Electronic Hypertext." Electronic Text: Investigations in Method and Theory. Ed. Kathryn Sutherland. Oxford: Clarendon P, 1997. 47-66.

Lombardi, Joseph. "Academic Libraries in a Digital Age." D-Lib Magazine. 6.10 (October 2000). 13 Aug. 2003 <http://www.dlib.org/dlib/october00/lombardi/10lombardi.html>

McGann, Jerome. "Literary Scholarship in the Digital Future." The Chronicle of Higher Education 13 (December 2002): B7-B9.

McGann, Jerome. "Radiant Textuality." Institute for Advanced Technology in the Humanities. 10 Feb. 1996. 13 Aug. 2003 <http://www.iath.virginia.edu/public/jjm2f/radiant.html>.

McGillivray, Murray. Hypertext Book of the Duchess. University of Calgary Press. 1997. 1 Dec. 2002

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

<http://www.ucalgary.ca/ucpress/online/pubs/duchess/Websample/mainmenu.htm>.

MLA Ad Hoc Committee on the Future of Scholarly Publishing, "The Future of Scholarly Publishing," Profession 2002 (New York: Modern Languages Association, 2002), pp. 172–86.

"Namespaces in XML." W3C. 14 Jan. 1999. 13 Aug. 2003 <http://www.w3.org/TR/REC-xml-names/>.

"On Search: Metadata." Ongoing. 30 Jul. 2003. 13 Aug. 2003 <http://www.tbray.org/ongoing/When/200x/2003/07/29/SearchMeta>.

"Open Access Now: Home Page." BioMed Central. 28 Jul. 2003. 13 Aug. 2003 <http://www.biomedcentral.com/openaccess/archive/?page=home&issue=2>.

"Page and Line Numbers." Text Encoding Initiative. 7 Aug. 2002. 13 Aug. 2003 <http://www.tei-c.org/lite/U5-pln.html>.

"Prefatory Note." Text Encoding Initiative. 7 Aug. 2002. 13 Aug. 2003 <http://www.tei-c.org/lite/U5-pref.html>.

"PubMed Central Home." PubMed Central. 21 Aug. 2003. 27 Aug. 2003 <http://www.pubmedcentral.nih.gov/>.

"Pupils Learn in Wi-fi Wood." BBC News. 2 Jul. 2003. 13 Aug. 2003 <http://news.bbc.co.uk/2/hi/technology/3035372.stm>.

"Resource Description Framework (RDF) Model and Syntax Specification." W3C. 22 Feb. 1999. 13 Aug. 2003 <http://www.w3.org/TR/REC-rdf-syntax/>.

Rosenzweig, Roy. "Crashing the System? Hypertext and Scholarship on American Culture." American Quarterly 51.2 (1999): 237-46. 1 Dec. 2002 <http://muse.jhu.edu/journals/american_quarterly/toc/aq51.2.html>.

Ross, Seamus. "Position Paper: Towards a Semantic Web for Heritage Resources." DigiCULT Thematic Issue 3: Towards a Semantic Web for Heritage Resources. May 2003. 13 Aug. 2003 <http://www.digicult.info/pages/publications.php>.

Rushinek, Avi, and Sara Rushinek. "What Makes a Publisher Important." Ubiquity 4.12 (2003). 13 Aug. 2003 <http://www.acm.org/ubiquity/views/a_rushinek_2.html>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

"Self Publish Books." CafePress.com. 13 Aug. 2003 <http://www.cafepress.com/cp/info/sell/books.aspx>.

Siemens, Ray. "Introduction and Overview." The Credibility of Electronic Publishing: A Report to the Humanities and Social Sciences Federation of Canada. Text Technology 11.1 (2002). 13 Aug. 2003 <http://web.mala.bc.ca/hssfc/Final/Credibility.htm>.

"Sign On." Budapest Open Access Initiative. 13 Aug. 2003 <http://www.soros.org/openaccess/sign.shtml>.

Sosteric, Mike, and Susan Hesemeier. "When a Learning Object is Not an Object: A First Step Towards a Theory of Learning Objects." International Review of Research in Open and Distance Learning (October 2002). 13 Aug. 2003 <http://www.irrodl.org/content/v3.2/soc-hes.html>

"SPARC: About SPARC." The Scholarly Publishing and Academic Resources Coalition. 20 Mar. 2003. 13 Aug. 2003 <http://www.arl.org/sparc/core/index.asp?page=a0>.

Sperberg-McQueen, C.M. "Bare Bones TEI." Text Encoding Initiative. Jun. 1995. 13 Aug. 2003 <http://www.tei-c.org/Vault/Bare/>.

"Structure of a TEI Text." Text Encoding Initiative. 7 Aug. 2002. 13 Aug. 2003 <http://www.tei-c.org/lite/U5-struc.html>.

Thurston, Thomas. "Hearsay of the Sun: Photography, Identity, and the Law of Evidence in Nineteenth-Century American Courts." American Quarterly: Hypertext Scholarship in American Studies. 51.2 (1999). 1 Dec. 2002 <http://chnm.gmu.edu/aq/>.

Thurston, Thomas. "New Questions for New Media: Scholarly Writing and Online Publishing." American Quarterly 51.2 (1999): 250-53. 1 Dec. 2002 <http://muse.jhu.edu/journals/american_quarterly/toc/aq51.2.html>.

Tipler, Frank J. "Refereed Journals: Do They Insure Quality or Enforce Orthodoxy?" International Society for Complexity, Information, and Design. 30 Jun. 2003. 13 Aug. 2003 <http://www.iscid.org/papers/Tipler_PeerReview_070103.pdf>.

"TUSTEP." TUebingen System of TExt Programs. 25 Nov. 2002. 28 Aug. 2003 <http://www.uni-tuebingen.de/zdv/zrlinfo/tustep_eng.html>.

van Winkelen, C. "Inter-Organizational Communities of Practice." elearningeuropa.info. 5 May 2003. 13 Aug. 2003 <http://www.elearningeuropa.info/doc.php?id=1483&lng=1&doclng=1>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

"Welcome to the TEI Website." Text Encoding Initiative. 6 Jan. 2003. 13 Aug. 2003 <http://www.tei-c.org/>.

Westbrook, David. "From Hogan's Alley to Coconino County: Three Narratives of the Early Comic Strip." American Quarterly: Hypertext Scholarship in American Studies. 51.2 (1999). 1 Dec. 2002 <http://chnm.gmu.edu/aq/>.

"What is the TEI Consortium?" Text Encoding Initiative. 1 Feb. 2003. 13 Aug. 2003 <http://www.tei-c.org/Consortium/index.html>.

Willett, Perry. "WWWP Guidelines." Victorian Women Writers Project. 10 Dec. 2002. 13 August 2003 <http://www.indiana.edu/~letrs/vwwp/vwwp-about.html>.

"World Wide Web Consortium." W3C. 13 Aug. 2003 <http://www.w3.org/>.

"XML Schema Part 1: Structures." W3C. 2 May 2001. 13 Aug. 2003 <http://www.w3.org/TR/xmlschema-1/>.

"XML Schema Requirements." W3C. 15 Feb. 1999. 13 Aug. 2003 <http://www.w3.org/TR/NOTE-xml-schema-req>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.