



October 2006

Collaboration in Electronic Scholarly Communication: New Possibilities for Old Books

Shawn Martin

University of Pennsylvania, shjmarti@indiana.edu

Follow this and additional works at: http://repository.upenn.edu/library_papers

Recommended Citation

Martin, S. (2006). Collaboration in Electronic Scholarly Communication: New Possibilities for Old Books. Retrieved from http://repository.upenn.edu/library_papers/53

Postprint version. Published in *Journal of the Association for History and Computing*, Volume 9, Issue 2, October, 2006).
Publisher URL: <http://mcel.pacificu.edu/jahc/jahcix2/articles/martin.htm>

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/library_papers/53
For more information, please contact libraryrepository@pobox.upenn.edu.

Collaboration in Electronic Scholarly Communication: New Possibilities for Old Books

Abstract

The Text Creation Partnership (TCP) project at the University of Michigan, a project that maximizes the respective strengths of scholars, libraries, and commercial publishers, has created a new model for academic scholarly publication and collaboration in the humanities. Such new electronic resources come with the tremendous possibility of changing the study of history, yet with such collaborative endeavors come many questions about use and collaboration. These endeavors also bring up new questions about the use of such resources by historians and how scholars of history should be proactive in the creation of digital scholarship.

Comments

Postprint version. Published in *Journal of the Association for History and Computing*, Volume 9, Issue 2, October, 2006).

Publisher URL: <http://mcel.pacificu.edu/jahc/jahcix2/articles/martin.htm>

Collaboration in Electronic Scholarly Communication: New Possibilities for Old Books

Shawn Martin
Text Creation Partnership Project Librarian
University of Michigan - University Library
shawnmar@umich.edu

Index

- [.01 Abstract](#)
 - [.02 The Promise of Electronic Resources: The Text Creation Partnership](#)
 - [.03 The Text Creation Partnership: Background and History](#)
 - [.04 The Benefits of the Text Creation Partnership](#)
 - [.05 Models of Scholarly Text Creation](#)
 - [.06 Technical and Philosophical Issues](#)
 - [.07 The TCP Case Study](#)
 - [.08 Scholars' Use of Texts](#)
 - [.09 Creators of Digital Texts](#)
 - [.10 Future Critical Issues](#)
 - [.11 Electronic Preservation](#)
 - [.12 Commercial Publishers versus Scholarly Consumers](#)
 - [.13 Collaboration](#)
 - [.14 Financial Arrangements](#)
 - [.15 New Modes of Scholarly Inquiry and the Problems of Electronic Scholarly Communication](#)
 - [.16 New Models for Scholarly Communication](#)
 - [.17 Conclusions](#)
 - [.18 Notes](#)
-

.01 Abstract

The Text Creation Partnership (TCP) project at the University of Michigan, a project that maximizes the respective strengths of scholars, libraries, and commercial publishers, has created a new model for academic scholarly publication and collaboration in the humanities. Such new electronic resources come with the tremendous possibility of changing the study of history, yet with such collaborative endeavors come many

questions about use and collaboration. These endeavors also bring up new questions about the use of such resources by historians and how scholars of history should be proactive in the creation of digital scholarship.

.02 The Promise of Electronic Resources: The Text Creation Partnership

It's a little like waking up in the British Library after closing time. The rare books of the British Library, Harvard, the Folger, the Huntington, and many others are suddenly accessible in their original appearance. [\[1\]](#)

Certainly electronic resources have increased accessibility to rare materials. Text-based resources like the Text Creation Partnership (TCP) at the University of Michigan enable scholars of history and other disciplines not only to read sources in ways they have done in the past but also to interact with their sources in ways previously unimaginable. The TCP, just one example of an electronic resource, has allowed innovative scholarship, pedagogy, and scholarly communication through a unique partnership among publishers of electronic image databases, university libraries, and the scholars at universities around the country.

The TCP project, in addition to being a model of new collaborative electronic publication, also serves as a conduit for discussion of many related issues such as the value of private/public cooperation, historians' use of electronic resources, and scholarly communication in the electronic age. This paper addresses a wealth of questions about such endeavors.

- What are the advantages and disadvantages of collaboration between university libraries and commercial publishers?
- How does a university library maintain its values of public domain and free flow of information when it is working with a corporate partner who would seemingly have conflicting values?
- How does a library deal with the pressure of becoming a publisher of text and a consumer of the same text?
- What value does collaboration bring to the scholarly community?
- What potential does a resource and model like the TCP have to change scholarly communication?
- What should the TCP do to maximize the benefit of its product and model to all parties involved (scholars, libraries, publishers)?

These are but a few of the questions that the TCP has asked itself over the past five years. Now, with a certain amount of experience behind it and a database of many thousands of texts, it is possible to begin answering these questions and to propose new solutions to some of the problems facing scholars in the electronic age. It is hoped that the TCP can continue a dialog among scholars, librarians and publishers about the importance of collaboration and the need for enhanced communication among all involved parties.

.03 The Text Creation Partnership: Background and History

In 1998, ProQuest Information and Learning published the Early English Books Online (EEBO) database containing every book printed in England or in English between 1470 and 1700 totaling roughly 125,000 titles, or in essence, the entire bibliography of the Short Title Catalog I (Pollard and Redgrave) and II (Wing), the Thomason Tracts, and the English Tract Supplement. These were scanned images of their already existing microfilm collection. The publication of EEBO was naturally a major step forward in electronic availability of primary source titles. Nevertheless, at least for the University of Michigan Library, it did not present a major step forward in holdings. The University Library already held all of these titles in microfilm. Although the ability to view individual titles from one's home computer and have multiple viewers of the same book were great access advantages, they did not add greatly to capabilities of the existing collection.

Librarians at Michigan felt that the true value for this collection lay not in digital facsimiles, but in the possibility of full text searching. That way, researchers and students could search individual words or concepts across titles, check citations to early authors, and copy and edit text for articles and papers. ProQuest saw the advantage of adding text to their image files, but felt that the additional cost of converting these images into full text would be millions of additional dollars and add so tremendously to the price of the product that libraries would be unable to afford the product if full text were added. Rather than taking no for an answer, the University of Michigan felt that it could get support behind the creation of full text for at least a subset of the EEBO collection. Thus the TCP project was born. Under the leadership of Mark Sandler, collection development officer at the University of Michigan Library, ProQuest agreed to partner with the University Library for the purpose of creating full text for a subset of twenty five thousand titles in the initial phase, with the understanding that the project might continue depending on the support it got. [2]

.04 The Benefits of the Text Creation Partnership

What is unique about the TCP initiative is not so much the partnership between private and public enterprises as its unique structure and new prototype for cooperation among university libraries, the academic community, and commercial publishing. The TCP is not a traditional grant-funded project but a partner-funded initiative that seeks monetary contributions from academic libraries to fund the creation of full text. Additionally, the full text that the TCP creates is not just another product, but a benefit to the academic community. All texts created by the TCP also enter the public domain. TCP partners are paying for texts which they own and will have the ability to distribute beyond their own campus communities rather than just having licensed ownership of the file for their own local and restricted use, as they would for any other commercial product. This model has been largely successful with, as of the time of this writing, over 12,000 texts available and 500 texts added every month. In fact, it has been successful enough to be extended to

two other similar commercial databases - Evans Early American Imprints (Evans), a collection of every work printed in Britain's American colonies and later the United States between 1639 and 1800 (based on the Evans Bibliography) available from Readex, Inc., and Eighteenth Century Collections Online (ECCO), a database containing over 150,000 titles printed in Britain between 1700 and 1800 available from Thomson-Gale. It is hoped that this model can be extended even further to other similar collections.

.05 Models of Scholarly Text Creation

Text creation itself, however, can be done in a variety of ways. There are basically two currently existing models of scholarly text creation. First, there is the commercial model in which texts are made available by commercial publishers. This type of database usually contains large amounts of material. The texts are restricted to a small number of users (the immediate campus community), and in terms of textual markup, the commercial model usually contains light encoding, thus making it of general use to novice researchers or undergraduates, but of less use to scholars. On the other hand, academic models of electronic text production contain small numbers of titles that are usually focused on a more particular discipline or area of research. They are free and open not only to those outside of the immediate campus community but also to the general public. These kinds of projects also have heavy amounts of textual markup that are particularly useful to specialists but of less use to those outside of that specialty.

Obviously there are many other variations on these models including projects administered by university presses, scholarly societies, and digital libraries. But, what one can immediately see is that there is a kind of model clashing here. Ideally for the academic community, it would be good to have databases containing large numbers of titles with rich markup that would be free to the world. The reality is universities do not have the money to create databases the size of commercial ones. Few grants could possibly cover the costs needed for scanning, interface creation, and text creation. Also, commercial publishers do not have the expertise or will to create projects useful to particular scholarly communities. These databases are expensive to produce. It is not cost-effective to create a database that only a small number of people will buy. The challenge, therefore, is how the academic community compromises to create a database that best fits the needs of all parties involved, given limited amounts of resources. The TCP has sought a middle ground on many of these issues in creating its model. How does the TCP do this? In general, the TCP tries to find a middle ground between all of these approaches

Technical and Philosophical Issues

With regard to the technical creation of files, the TCP creates the file using SGML/XML markup under the Text Encoding Initiative (TEI) guidelines. The TCP likes to call its approach "TEI lite with additions." What this means is that the TCP creates a file containing all of the basic information with paragraphs, titles, chapter headings and the like tagged. Additionally, the TCP will mark up some elements such as colophons,

quotations, and type face changes which will be of help to scholars using this material. The TCP does not, however, create richly tagged editions of the scholarly variety. The TCP does not make editorial decisions. It is hoped that scholars may, at a later point, take ownership of these files and create enhanced editions using TCP text as a base. Additionally, these files are created to 99.995 percent accuracy for the transcribed text. The original page scan from ProQuest is typed simultaneously by two different people, then resolved by a third person, and reviewed by experts at the University of Michigan. In essence, the TCP makes sure that the text is as accurate as it can possibly be.

The TCP's funding is one of the more intriguing aspects of the model. Rather than grant funding which often is not sustainable over thousands of texts like this, the TCP has opted to let academic institutions, normally libraries but also departments and grant funds as well, contribute funds to the project which are then matched by the commercial publishers. That money is used to fund text production, and the more institutions that join, the more text can be created, thus making each text less expensive overall. So far, at the time of this writing, over 12,000 texts have been created, and there is enough money to produce over 18,000 by 2006. The TCP is well on its way to creating 40,000 texts by 2009 and is already one of the largest text corpora in existence and the largest electronic text corpus loaded at the University of Michigan. [3]

The TCP Case Study

TCP is an excellent case study that can explore larger questions about electronic resources and scholarly communication. First, how are electronic resources used by scholars and the humanities and social sciences, and what potential do they have to change current practices? Second, what potential values and pitfalls are there to the profession in creating these new models and technologies? By answering these questions, one can begin to see the issues involved and the solutions that the historical and humanistic community will need to address as digital technology progresses.

Scholars' Use of Texts

How are electronic resources used by scholars and what potential do they have to change current practices? Librarians have been studying these issues for many years. [4] In fact, some have even diagnosed and identified a significant disparity in the resources that libraries and archives keep as opposed to those that scholars, historians in particular, need to use. As the amount of information available to libraries has increased exponentially (the exact kind of information that historians use for social analyses) and the capacity of libraries both monetarily and spatially to store that information has increased only partially (or in some cases not at all), libraries, as a result, have had to be more selective in the kind of information they store. This, according to some, has led to a divide between repositories of information - libraries and archives - and the users of that information. [5] This certainly holds true for paper resources and to an extent electronic ones. Nevertheless, the problem of electronic resources would seem to be an overabundance of them, or at least a potential overabundance. [6] This requires historians to select among

many different kinds of resources with differing degrees of success. Recent studies indicate that historians want depth and comprehensiveness from electronic databases. They would rather turn up too many references to potentially unhelpful material rather than turn up small numbers of references to relevant materials with the chance that potentially useful material might slip through the cracks. On the other hand, historians generally mistrust electronic resources that they often see as too narrow, of more use to undergraduates, and inherently unreliable as compared to an original source. In all, although historians might use electronic resources more frequently now than in the past, they still only use them to supplement "traditional" print or manuscript material held in libraries and archives. [7]

Looking at the TCP's use between 2002 and 2005 reflects many of these general findings. New discoveries seem to be made every day with the resources particularly in EEBO and EEBO-TCP. With the ability to search across the words contained within works (rather than just the catalog records), scholars can do previously impossible research. By searching a word or name in the EEBO and EEBO-TCP databases, they can find hundreds or even thousands of references within seconds. The same search would have taken a lifetime to find in libraries and archives scattered around the world. So, these resources not only open up access to previously untapped works, they also enhance access to these early books and allow for research that would previously have been impossible.

People have related many examples of such searching. Tom Izbicki from Johns Hopkins University wrote about his experience in finding references to Nicholas of Cusa, a medieval Catholic theologian, in a variety of titles including Protestant tracts. Jennifer Danby, a graduate student at the City University of New York had similar success finding references to an actor she was working on, Michael Mohun, using EEBO-TCP. She searched his name and found it in cast lists, medical texts, treatises on vocal training, and legal works. Others have reported similar success searching on topics like Aristophanes and Erasmus. This writer had luck doing his own research on Jeremy Collier, a seventeenth century anti-theatrical pamphleteer. His name shows up in works of French poetry, something one would never expect.

Another particularly useful avenue of research using the TCP text has been linguistics. Now a large corpus of English words spanning the fifteenth through the eighteenth century is available, and scholars at Northwestern and the University of Pennsylvania are searching frequency and proximity of words like "God" and "Man." Perhaps the best statement about the potential uses of new scholarship comes from an undergraduate who won a prize for his essay in the EEBO in Undergraduate Studies essay contest in 2003. Brian Platzer in his essay *Colonial Environmentalism: Harriot and Raleigh's Manipulation of the New World Landscape* said, "EEBO provided me with a bank of searchable texts that had probably never been studied in conjunction with Raleigh's Discovery. I researched to the point when criticism stopped, and using EEBO, I was able to go a step farther by searching and then incorporating texts contemporary to these writers. This paper demonstrates the power of EEBO keyword search on analysis of early English texts." [8] In a way, access to electronic resources like EEBO democratizes

access to primary information. Undergraduates, especially those at small institutions, would never have dreamed of doing primary research like Brian Platzer's. Large research universities have always had access to primary resources in their own special collections. Now, even the smallest institution has access to the collections of the British Library, Folger, Huntington, Newberry, Harvard, and many others. One great example of this is Washington College, an undergraduate institution of around 1,200 undergraduates in Eastern Maryland. Heidi Atwood, an undergraduate from there won the Undergraduate Essay Contest in 2005. Resources like EEBO allow study that prompted Richard Bailey from the University of Michigan to say "My undergraduates are doing research at a doctoral level."

Nonetheless, one of the TCP's paramount concerns has been that despite its high usage by these few enthusiastic scholars, overall usage has been relatively low over the past few years. Taking the University of Michigan as an example, and assuming our peer institutions perform more or less the same, for the entire year of 2004, there were around 5,000 uses from the University of Michigan for EEBO. For EEBO-TCP there were just over 1,000 uses in the same year. This may sound large. Yet, comparing those statistics to those of the appropriate renaissance and early modern studies journals contained in JSTOR there were nearly 30,000 uses from the University of Michigan alone in 2004 (out of approximately 1.2 million uses overall). Admittedly some of this gap can be accounted for because those journals contain articles relating to disciplines other than early modern English studies. Nonetheless, when a library has to justify purchase and maintenance of a large resource like EEBO, a classic measure used by administration is usage. One can make the argument that resources like EEBO are essential to early modern scholars in a way that JSTOR might not be to other disciplines. Such an argument becomes increasingly difficult as budgets become tighter.

Two conclusions are possible from these data. Either scholars do not see them as useful or they do not know about them. In order to determine the answer to that question, the TCP with the help of the School of Information at the University of Michigan studied user behavior with the TCP projects and held four task forces over the past years to determine how people are using the databases. [9] From these meetings, it was determined that scholars generally found these resources incredibly useful for themselves and their graduate students, but also felt that the community was unaware of the resources available to them. Primarily, and perhaps more importantly, faculty were unaware of the advantages the TCP offered them.

Given this analysis of current use of electronic resources, what is their potential? Edward Ayers has pointed out the possibilities of hypertext history that he defines as a "culmination of a long held desire to present a more multidimensional history and a threat to standard practice." [10] Others have pointed out the possibilities of creating new historical landscapes and historical narratives in Geographic Information Systems (GIS). [11] New technologies open up entirely new areas for historical inquiry. [12] Interactions between text and user, social memory, and multiple editions of electronic text each with different interpretations and marked up stages become a complex network of historical

and literary analyses or what Jerome McGann refers to as "radiant textuality." [13] In all, the possibilities are numerous, and, to date, largely untapped.

Creators of Digital Texts

In the ideal world, scholars could create resources that could then be used by others for research and teaching. Indeed some projects have already begun to move in this direction like The Valley of the Shadow at the University of Virginia and History Matters at George Mason University. [14] Nonetheless, given the cost of rights to images, copyrights, imaging, and production, the likelihood of individual scholars or even groups of scholars producing these kinds of electronic resources seems small. [15] Additionally, scholars find themselves with additional roles of publisher, web master, network administrator and librarian. Rather than dividing their efforts, scholars, librarians and publishers should unite to create reliable, peer-reviewed electronic resources of sufficient depth and comprehensiveness to be of use both for scholarship and pedagogy. [16]

TCP is certainly a project that combines the efforts of publishers, scholars and librarians. Such an arrangement is ideal and draws on the respective strengths of all concerned. Publishers can use their money and their production capability to disseminate resources of sufficient depth. Librarians are able to use their expertise in archiving and standards to ensure that such resources integrate across platforms. Scholars can use their expertise in content to select appropriate titles for digitization and double check to make sure the texts are sufficiently accurate for their needs.

So far in the TCP project scholars have met in several task force sessions to discuss how best to go about aiding scholarship, and the TCP has been able to put into effect many of the task force's recommendations. [17] Additionally, ProQuest has been able to enhance the product based on suggestions from librarians and scholars. The TCP has put ideas for class syllabi and assignments on its website. [18] Eventually a website like that of EEBO-TCP could serve as a conduit for collaboration and communication among students and teachers, serving as an ideal model for ways in which collaborative projects can and should work.

Future Critical Issues

On the other hand, any project that attempts collaboration on so grand a scale in ways that are largely untraditional is bound to raise additional concerns about the role of the academic community in this changing environment. The TCP model has looked at many of these issues over the past few years. They include:

- Electronic preservation
- Tensions between commercial publishers and the scholarly community
- Collaboration
- Financial arrangements

- New modes of scholarly inquiry and the problems of electronic scholarly communication

The following sections address each of these issues in turn.

Electronic Preservation

How does a publisher deal with the issues of electronic preservation? A primary concern for libraries and scholars is the importance of finding materials essential to their research years from now. Additionally, libraries desire print-like stability of content and the academic legitimacy that stability brings. [20] From a library perspective, there are two schools of thought on how to create and preserve electronic text of primary historical materials. First, there are those who say that it is important to create richly tagged text with much editorial markup. This allows scholars to interact with the text in new ways and ask new questions of the text - seeing patterns that would previously have been impossible to discover. [21] The counter argument to that is for electronic preservation libraries' need to encode only the basic structure of the text itself (those data elements that scholars will not change as time progresses) and leave scholars to take that basic text and create new editions later. [22] This second kind of text encoding does not provide scholars with the ability to interrogate the text in the way that the former approach does. Nevertheless, it does preserve the words themselves and the basic structure. Libraries have to preserve text for a large community of multi-disciplinary scholars. They also must preserve content over a long period of years in such ways that it will endure beyond particular scholarly fashions and be accessible to not only a narrow range of specialists but also to non-specialists, undergraduates, and the general public.

Additionally, the more specialized mark-up added to a text, the more expensive it becomes to produce. Therefore, libraries can produce a canon of well encoded text, but it becomes nearly impossible to encode a large corpus of material in such a way. The TCP project has chosen to adopt the lighter mark-up model given concerns of electronic preservation, multi-disciplinary use, the broad range of covered material, and the desire to preserve materials for a long period of time. Since libraries will eventually own all texts that the TCP produces, it will become possible for scholars at a later date to take ownership of files and create new, richly tagged editions of the file.

At the same time, the University of Michigan will maintain control of the original TCP text file. The University of Michigan has taken the stand that digitization is a form of preservation and the University Library is committed to preserving the infrastructure and means of preserving this content through shifts in technology across time. In particular, the library participates in a number of initiatives such as the Making of America, the Humanities Text Initiative and the Google library that will promote preservation of a wide range of documents. [23]

Commercial Publishers versus Scholarly Consumers

How does an academic library grapple with the seemingly paradoxical cultures of a commercial publisher and a scholarly consumer of text? Now that the library has adopted this collaborative endeavor, it puts the library in a strange and unfamiliar territory of having to adhere to production timetables, to administer invoices and funds, to restrict access to resources, and to deal with customer use and complaints. All of these are familiar territory for our commercial partners, but not for the academic community. Even more important than the administrative details are the more fundamental questions: When a library begins to become more like a publisher does it start losing its own values of free flow of information? Does an academic institution have to compromise so much that it is subservient to a more powerful commercial partner. Again, the TCP has had to walk this tightrope over the last several years, and although no partnership is perfect, there are some insights into how this uncharted territory can begin to be mapped out. Rather than saying that our values are conflicting, it is perhaps best to say that they are evolving. Certainly universities cannot and should not give up on any of their values, and it is incumbent upon the academic community to convince publishers that our values and their values coincide.

Collaboration

Collaboration like the TCP enhances a commercial product and provides universities the capability to create resources of broader depth than would ever be possible just in grant funds or other traditional endowments. Academic institutions do not have the kind of resources either to fund or to produce a database as large as the three with which the TCP works. No university would ever be able to create a database of over 100,000 titles nor would any scholarly project get the wide publicity that the TCP enjoys with the benefit of the marketing machines commercial companies employ. Yet no publisher would have the expertise to know how scholars use these titles, nor the knowledge to enforce standards needed for integration with other library resources. Additionally, commercial publishers are less well-equipped to go to conferences, hold task forces, and solicit feedback from the academic community than are the TCP project librarians. So, both sides benefit from this kind of collaborative endeavor.

There are some compromises and drawbacks that both universities and publishers must face when moving into this new electronic publishing age. First, universities must realize that they cannot have everything they want. Publishers have rights over the content they produce and in order for libraries and scholars to get access to that content they need to work with publishers. The TCP has proved a successful model so far, but there is only a limited amount of money and the stability of the model must be tested over time. Is such a model applicable to other institutions? Michigan and Virginia already have a large digital library infrastructure behind them. Would it be possible for smaller universities to provide the same kinds of models among institutions without that kind of support behind them?

Although the TCP certainly provides long-term benefits for the academic community, there are also some short-term drawbacks. These texts do not enter the public domain for many years after their creation. So, for academic projects that are just now starting up and

would like to use texts, it is impossible for them to use TCP texts without some limitation. This is only fair to the publisher, but not particularly helpful for scholars in the short term. Also, digital libraries contain many different kinds of materials. Can the TCP model extend to other types of materials? Will the same principles be applicable to images, journal articles, and multimedia resources scholars also need to access? Finally, as technology changes and becomes less expensive, how should the academic community respond? Will the TCP model still be applicable as the gap between what publishers are able to produce and what universities are able to produce closes? We have all seen how the price of technology has changed over a period of years. Is it possible that one day universities might have the ability to produce resources at the same rate as commercial publishers? If so, how will new models need to be developed or modified?

Financial Arrangements

An additional and very substantial drawback for libraries especially is the monetary expense it takes to sustain such models. When libraries are investing thousands of dollars to produce such large projects, is it worth the effort? How can universities ensure that their investment is being maximized when it comes time to justify the budget to a provost or dean? Conversely, publishers too have expenses and need to convince their CEOs that such collaboration is profitable for the company. Some of the benefits have already been discussed - further publicity for libraries and enhanced functionality for commercial publishers. It is important to realize, however, that all parties will need to increase their efforts in addition to the money invested in the project. Libraries and publishers have to raise awareness in the scholarly community to make sure that these resources are used enough to justify the expense. Additionally, the project itself has to create opportunities for scholars to be involved in selecting texts, creating editions, and performing research. University administrations need to reward this new kind of research in tenure guidelines and promotion. Libraries need to provide tools for use of these kinds of materials in scholarship and pedagogy with syllabi and course tools. Publishers need to continually enhance their products to meet these new needs and create new products that might further their own interest and respond to the new needs the community will face. In all, it will take commitment to make these new types of collaborations and partnerships work. The rewards will be in the form of research performed by leading scholars and interdisciplinary resources for technologically-savvy faculty and students. Publishers working with universities will have new opportunities to create user-centered products while working in collaboration with rather than against their consumers.

New Modes of Scholarly Inquiry and the Problems of Electronic Scholarly Communication

Some of the most significant obstacles, however, in facilitating this kind of collaboration are the inherent weaknesses of electronic scholarly communication and inquiry. Archival research has been the mainstay for humanistic scholars for many centuries. Yet, with the advent of electronic resources, the need for going back to the original sources has become more rare. Why should I look at the original if it is available in EEBO? Conversely, and

potentially more concerning, is the belief that if a work is not available in EEBO, then it does not exist. These two concerns have led to hesitation among some in the academic community. Clearly these apprehensions are justified to an extent. For instance, if a scholar goes to an archive and sees a poorly printed octavo versus a richly printed folio with hand colored woodcuts, one can already deduce something about these books. The octavo may have been a surreptitious book meant to be hidden in a pocket; the folio may have been a book intended for public display in a library or church. Thus, electronic versions of these books scanned in black and white and devoid of all sense of scale may cause what one faculty member termed "stripping away potential layers of meaning." Also, if the younger generation of scholars turns increasingly to electronic resources and is not trained in use of other sources like print bibliographies, it could have deleterious effects. Finally, and more practically from a library standpoint, if scholars are using EEBO and not consulting books in the library, what purpose does the library of primary resources serve and will its funds be cut because of fewer uses? [24]

There are many answers to these questions. In summary, it is important to say that the role of the library itself is changing and although one can make the argument that special collections will be less consulted as digitization increases, one can also make the argument that although use of particular types of material may decline, use of others will increase. For instance, the Bodleian Library at Oxford University reports an overall decrease in paging of books in EEBO; it also reports an increase in paging of manuscript material which has not been digitized. So, although certain resources are being used less, others are being used more. [25] Therefore, one could argue that resources like EEBO allow scholars to focus their studies and use their time in archives more efficiently. Some of the study needed to restore "layers of meaning" should certainly still be performed; an electronic version cannot replace the original. In essence there are two kinds of research researchers seem to be identifying. The first is informational, that is reading a book for the contents and words contained within. The second is artifactual or analyzing a book as an artifact and deriving meaning from physical characteristics like print size and paper type. Although EEBO and EEBO-TCP can revolutionize informational research it cannot begin to replace artifactual research. The key is for scholars young and old to understand and teach how to navigate these differences.

This leads to a second problem facing electronic scholarship, the gap between what scholars know about electronic resources and what librarians are able to teach them. Up to now, librarians have adopted the role of educating their patrons about electronic resources in the same way they have always done with print. The University of Michigan, for instance, teaches many such classes, but the attendance rolls show which classes would seem to be more valued. For example, the most recent Enriching Scholarship series which the University Library sponsors every year and with which I am involved, offered a series of classes on issues like web searching, the recent Google initiative, and creating effective PowerPoint presentations. All of these classes had over one hundred people registered (and some even added sessions to accommodate more). My class on "Old and Rare Books Online" about EEBO, Evans, ECCO, the TCP, and some of the other related projects had five people. From conversations with my colleagues teaching similar classes, this is not unusual. Why? Scholars seem to prefer hearing from their

colleagues about new potential uses and changes to their research methodologies, not from librarians. This is understandable. Librarians are generally aware of the resources, but often unaware of some of the potentials for scholarship. Scholars are aware of the potentials, but often unaware of the resources. Thus it would seem that scholars view libraries as a place to learn technical rather than research skills

Scholars also seem to see little need for new paradigms for electronic research because of a range of causes varying from the lack of reward in tenure and promotion for electronic scholarship, the large number of databases available, the lack of time to investigate them, and ignorance of how to use their functions. Perhaps it is best summed up by another faculty member who said "I need to see something new." He meant that the type of scholarship he practiced had not changed. Admittedly, electronic resources helped him to find things more efficiently, but fundamentally the way he went about his work and the conclusions to which he came have not changed, nor did he foresee them ever changing the future. The ultimate causes of this discrepancy are neither a lack of interest among scholars nor a lack of awareness about the resources; rather, they are a lack of knowledge about how to exploit different kinds of resources and an inherent weakness in the current scholarly communication and publication chain.

The traditional model of scholarly communication/publication has always been that publishers sell material (books, journals, and databases) to libraries and then scholars use those databases for their own research. This model has worked well for print over the years but has had problems in moving to the electronic realm. [26] The same is also true, although in a different way, for databases of primary sources. Traditionally, library collection development has centered on the scholars and students with librarians acting as intermediaries between consumers of content (researchers) and purveyors of content (publishers and booksellers). As researchers inform librarians about what they need and librarians look to see what might be needed for the future, the library collects via a so-called "just in case" model of development. In other words, the library is anticipating future needs for scholars. With the advent of so many electronic resources and a perceived (and often justified) immediate demand for them, library collection development has changed to a "just in time" model, meaning the library collects in order to meet current demands, not future ones. [27] The role of the library as intermediary has not changed, but the demand both practical and economic has changed greatly. What does this have to do with the gap between use and potential use? With the "just in time" model now common among libraries, librarians often purchase databases either without considering the perceived needs of their users, or correctly perceiving them but being too overburdened with EEBO, Evans, ECCO, the TCP, and many other databases for faculties across the university. Therefore, scholars are often left to their own devices and serendipitous discoveries without realizing the full potential of what is available to them. Additionally, many of these databases are purchased based on perceived use. Librarians do not have the same expertise as scholarly specialists. So, although there may be tremendous scholarly potential of the content within the databases, that potential is not used fully by the publishers who sell to the librarians nor is it completely understood by the librarians who buy them. Hence, the potential use of these databases goes untapped.

New Models for Scholarly Communication

Will Thomas stated,

The goal for historians working in the new digital medium needs to be to make the computer technology transparent and to allow the reader to focus his or her whole attention on the 'world' that the historian has opened up for investigation, interpretation, inquiry, and analysis. Creating these worlds, developing the sequences of evidence and interpretation and balancing the demands and opportunities of the technology will take imagination and perseverance. [28]

It is hard to disagree with the above argument but it can be taken a step farther. Not only will it be incumbent upon the historian of the electronic age to create new historiographies and new modes of inquiry. It will also be incumbent upon the historian to navigate a changing landscape in which the roles of scholar, librarian, and publisher are increasingly merged. It will also be a necessity to take an active role in steering this development in ways that will eventually enhance the mission of teaching and learning. Not to do so will lead to the domination of scholarship and pedagogy in the humanities by a conglomeration of publishers that will control access to electronic scholarship and by a handful of universities that have the money to support scholarship by paying those publishers. By changing not only the way in which historians interpret their work, but also by changing the nature of the role itself, scholars now have a unique opportunity to shape the way the discipline, indeed their very role within the academy, will be defined for years to come. Therefore it is essential that librarians and scholars take the lead in creating a new scholarly communication system that both cooperates more fully among interested parties and educates everyone involved about how to use this new system.

It is hard to dispute the fact that the current scholarly communication system, well documented in past articles, is inherently flawed in the electronic age. Scholars talk to each other about their research and what they are using. Librarians talk to scholars about their research and buy databases based on their perception of scholars' immediate needs. Librarians talk to publishers about these perceived needs. Publishers create databases based on their talks with librarians that they in turn sell back to libraries, and librarians attempt to get scholars to use these newly published databases. Clearly, libraries are at the center of this information chain and need to be proactive in changing the current situation. Similarly scholars need to be more proactive in thinking about what they want and relating that to librarians and publishers. Yet, if the three involved communities (scholars, publishers, and librarians) continue to talk more to themselves than to the other two groups, the situation will remain the same. Therefore, new models are essential in navigating this new electronic world in which we live. The TCP is one model that can help to do this, and hopefully help to focus further dialog about what the needs of all three communities are for the future.

Education is another essential component in this new model. Since the TCP is a collaborative endeavor, faculty have the opportunity to select text, procure it for their own scholarly or educational projects, and be more involved than in most digital library

projects. It was suggested by scholars that the TCP do more outreach to the scholarly community and hopefully from those communications, the TCP could both increase use and interest in the project. This is exactly what the TCP has done. Over the past three years, the TCP has created an academic advisory group to help with selecting texts and to facilitate communication between the TCP and the academic community. Librarians from the TCP have attended several scholarly conferences. Has this effort been successful? There are two measures that can perhaps quantify the efforts the TCP has made for the academic community. The first is usage statistics - often used in library communities to measure success. For EEBO-TCP, the total usage (not just the University of Michigan) has increased from 1,700 uses in 2002, to 8,000 in 2003, to over 10,000 in 2004. Since 2002 is the first year that the TCP began many of its initiatives, the data imply that there may be a correlation between usage statistics and outreach efforts. The second, less quantifiable evidence is the TCP's use in scholarly projects. In 2002, EEBO-TCP was cooperating with two academic projects, both headquartered at Oxford University (the TCP's primary partner in editing text). As of 2006, EEBO-TCP is cooperating with projects at the University of Oxford, the University of Chicago, the University of Pennsylvania, Northwestern University, Washington University — St. Louis, the University of Western Ontario, and the University of Victoria. These projects are studying everything from standards of metadata integration to use of Spenser in early modern reading patterns. Finally, and just anecdotally, it seems that there is more activity centering on EEBO and the TCP just within the past year. At the American Library Association Annual meeting in 2005, there was a panel including both librarians and scholars sponsored by the Literatures in English Section. In September 2005 another conference on EEBO and the TCP in the UK was (de)materializing the Early Modern Text. Neither ProQuest nor the University of Michigan directly initiated any of these activities. So, these projects are beginning to take on a life of their own, independent of intervention from their creators. Clearly it would seem that the TCP is having a great impact on the academic community.

Conclusions

What can one draw from these facts? All of the communities involved have to do more to cooperate with each other. No university (much as we may hope) would ever be able to digitize thousands of early modern books. Corporate capital and efficiency are required for that task. No publisher could ever garner the expertise needed to integrate these resources in the way a library could. No publisher or librarian could ever know the scholarly potential that such resources carry without the help of faculty specialists. Therefore, it is in the best interests of all to cooperate more fully. Some of the ways the TCP has done this may serve as an outline for further cooperation. Libraries need to reach scholars more fully. One way has always been and will no doubt continue to be through classes on resources taught by librarians at their libraries. Librarians also need to go out to the scholarly community more frequently to conferences and other academic venues and raise awareness of the many issues confronting us today. Conversely, scholars need to think about the potential for such resources and, rather than creating their own scholarly projects independently, look to libraries for ways of partnering both with freely available and commercial resources. Also, scholars could look at ways of educating

themselves and their students about the potential of electronic resources (perhaps with the help of librarians), continuing to use print resources, and creating tools to facilitate learning that maximizes both. Publishers need to be open to agreements with the academic community and realize that projects like the TCP can both increase use of the resource and enhance the product by using the expertise of primary users.

These are only a few of the possibilities that a project like the TCP can offer. Although there are still many barriers to overcome, like enhancing artifactual scholarship in the digital environment and increasing usage to its maximum, by cooperating and thinking through these issues, the early modern studies community can overcome these challenges.

The TCP is by no means the only potential model for collaboration among scholars, librarians, and publishers, and in many ways this paper raises more questions than answers. Nonetheless, by investigating the successes and failures of the TCP model, perhaps it can serve as a way of thinking about these issues and what we as a community can do to change them. One often hears about the advent of electronic technology and the age in which we live as a transitional moment in history. No doubt this is true, but the challenges presented to us are in some ways no different than those present in the era that we study. Charles Blount in his *Just vindication of learning, or, An humble address to the high court of Parliament in behalf of the liberty of the press* identified many of the same problems and said "Having thus demonstrated how much the World owes to Learning and Books; let me not be altogether unmindful of Faust and Guttenburg, the promoters of both; who by their Ingenuity discovered and made known to the World that Profound Art of Printing, which hath made Learning not only Easie, but Cheap." Surely the same could also be said of our own time when electronic technology has made the reproduction of information extremely inexpensive. Yet Blount also realized that "Learning hath of late years met with an obstruction in many places, which suppresses it from flourishing or increasing, in spite of all its other helps" [29], much like many electronic resources have not met with their full potential. We, too, in the modern academic community have an opportunity to create a future for ourselves. By looking at ways that projects like the TCP have tried to cope with the inherent problems of collaboration between librarians, scholars, and publishers, and by using some of the solutions the TCP to date has employed, it may be possible to further improve on that success and create a new future for the digital library and scholarly communication system.

Notes

[1] John Schmitt, "Early English Books Online" *The Charleston Advisor* 4, no. 4, (April 2003): 5-8.

[2] Much has been written about the TCP's philosophy and model much of it by Mark Sandler, the Collection development officer at the University of Michigan library. TCP was originally Mark Sandler's idea and he has taken an active interest in it ever since. For more of his work look at: Mark Sandler, "New Uses for the World's Oldest Books: Democratizing Access to Historic Corpora," *ARL Bimonthly Report* 232 (February

2004): 4-6. Mark Sandler, "The Early English Books Online — Text Creation Partnership," *The Charleston Advisor*, 4, no. 4 (April 2003): 47-49.

[3] A good summary of the issues involved with such partnering and the needs of the many kinds of scholars involved in this project is available in: Jeffrey Garrett "The Early English Books Project meets at Northwestern" *College and Research Libraries News* 63, no. 2 (February 2002): 117-119. Hillary Nunn "Early English Books Summer Camp" *Biblio-Notes* 38 (Fall 2001): 3.

[4] Margaret F. Stieg, "The Information Needs of Historians" *College & Research Libraries* 42 (1981): 549-60.

[5] Francis X. Blouin Jr., "History and Memory: The Problem of the Archive" *Publications of the Modern Language Association* 119 (2004): 296-298.

[6] Roy Rosenzweig, "Scarcity or Abundance? Preserving the Past in a Digital Era" *American Historical Review* 108, 3 (June 2003): 735- 762.

[7] Margaret Stieg Dalton and Laurie Charnigo "Historians and Their Information Sources" *College & Research Libraries* 65 (2004): 400-420.

[8] All of these examples have been published in the newsletter of the EEBO-TCP project and are available online at http://www.lib.umich.edu/tcp/eebo/archive/archive_pub.html. Further information about the Undergraduate Essay including Brian Platzer's full essay is available at http://www.lib.umich.edu/tcp/eebo/edu/edu_win_03.html.

[9] The School of Information report was internal and not available at the TCP website, but minutes of the task force meetings, which come to much the same conclusions, are available at <http://www.lib.umich.edu/tcp>. Additionally, more information about the academic advisory group and conferences TCP has attended is also available at the TCP website.

[10] Edward Ayers "History in Hypertext," <http://www.vcdh.virginia.edu/Ayers.OAH.html>.

[11] William G. Thomas, III "Computing in the Historical Imagination" in *A Companion to the Digital Humanities* ed. Susan Schreibman, Ray Siemens, and John Unsworth, (Oxford: Blackwell Publishing, 2004): 66.

[12] Francis X. Blouin Jr., "History and Memory: The Problem of the Archive" *Publications of the Modern Language Association* 119 (2004): 296-298.

[13] Jerome McGann *Radiant Textuality: Literature after the World Wide Web*. (New York: Palgrave. 2001): 151-152.

[14] See Valley of the Shadow at <http://valley.vcdh.virginia.edu/>) and History Matters at <http://historymatters.gmu.edu/>

[15] Roy Rosenzweig and Michael O'Malley "Brave New World or Blind Alley? American History on the World Wide Web" *Journal of American History* 88, no. 2 (June 1997): 152.

[16] See Peter Roberts "Scholarly Publishing, Peer Review, and the Internet" *First Monday* 4 no. 4 (April 1999), http://firstmonday.org/issues/issue4_4/proberts; and Modern Language Association, "Guidelines for Evaluating Work with Digital Media in the Modern Languages," http://www.mla.org/guidelines_evaluation_digital

[17] Two Task Forces have been held on interface and selection for the EEBO-TCP project available on the EEBO-TCP project site at <http://www.lib.umich.edu/tcp/eebo/documents/NW.doc> and http://www.lib.umich.edu/tcp/eebo/News/Task_Force_Report.html respectively and a focus group among English faculty available at <http://www.lib.umich.edu/tcp/eebo/News/umfocusgrp.html>

[18] The EEBO-TCP website has areas for scholars to post syllabi and assignments for classes using the resource at http://www.lib.umich.edu/tcp/eebo/edu/edu_intro.html and has created the undergraduate essay contest that demonstrates the work of undergraduates using the resource in a wide variety of fields. Examples are available at http://www.lib.umich.edu/tcp/eebo/edu/edu_essay.html

[19] Margaret Hedstrom, "Research Agendas Set Course for Digital Archiving and Long Term Preservation" *RLG DigiNews* 7, no. 6 (2003), <http://www.rlg.org/legacy/preserv/diginews/diginews7-6.html#feature2>

[20] Caroline Arms, "Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress" *RLG DigiNews* 4, no. 3 (2000), <http://www.rlg.org/legacy/preserv/diginews/diginews4-3.html#feature1>

[21] Carole Mah and Julia Flanders. "Scholarly Needs, Encoding Challenges: Correction, Regularization, and Expansion." *The Brown University Women Writers Project Newsletter* 2, no. 2 (1996).

[22] Christina Powell, "SGML Creation and Delivery: The Humanities Text Initiative" *D-Lib Magazine* (July/August 1997), <http://www.dlib.org/dlib/july97/humanities/07powell.html>

[23] See Making of America at <http://www.hti.umich.edu/m/moagrp/> the Humanities Text Initiative at <http://www.hti.umich.edu/> and the Michigan Digitization Project (Google library) <http://www.lib.umich.edu/mdp/index.html>

[24] For further discussion of the concern among libraries see Peter B. Hirtle "The Impact of Digitization on Special Collections in Libraries" *Libraries and Culture* 37, no. 1 (2002): 42-52.

[25] Richard Ovendon, "Being Digital in Bodley: Early Books and Research Libraries in the Google Era" (de)materializing the early modern Text: Early English Books Online in Teaching and Research Conference Bath Spa University 8-9 September, 2005.

[26] The so-called scholarly communication crisis is well documented elsewhere and does not need repeated here, but for a good overview read Cathy Davidson "Understanding the Economic Burden of Scholarly Publishing" *Chronicle of Higher Education* 50 issue 6 (October 2003) <http://chronicle.com/free/v50/i06/06b00701.htm>

[27] For a more complete discussion of this trend and of the scholarly communication crisis in libraries, see Ruth Miller "Electronic Resources and Academic Libraries, 1980-2000: A Historical Perspective" *Library Trends* 48, no. 4 (Spring 2000).

[28] William G. Thomas, III "Computing in the Historical Imagination" in *A Companion to the Digital Humanities* eds. Susan Schreibman, Ray Siemens, and John Unsworth, (Oxford: Blackwell Publishing 2004): 66 - 67.

[29] Charles Blount, *A just vindication of learning, or, An humble address to the high court of Parliament in behalf of the liberty of the press by Philopatris*. Early English Books Online — Text Creation Partnership database, text A28439 <http://ets.umd.umich.edu/e/eebo/> consulted 9 September, 2005.