

# Making the Most of Free, Unrestricted Texts: a first look at the Promise of the Text Creation Partnership

Rebecca Welzenbach  
Text Creation Partnership Project Outreach Librarian  
University of Michigan Library

## Abstract

The [Text Creation Partnership](#) (TCP) is a library-based initiative that partners with commercial publishers to create SGML/XML-encoded editions of early print books, specifically, those captured in databases such as ProQuest's [Early English Books Online](#), [Gale's Eighteenth Century Collections Online](#) and Readex's [Evans Early American Imprints](#).

Our work is supported by both commercial publishers and libraries, and our business model seeks to balance the interests of both. Central to our mission is the fact that, after a defined period of time during which the commercial publisher has the exclusive right to distribute the texts we produce, these files will be made freely available for anyone to use.

In April 2011, the TCP [announced](#) that, with the agreement of Gale, restrictions had been removed early from the 2,231 works from ECCO that the TCP keyed and encoded between 2005 and 2009. These are the first TCP texts to be released to the general public. We were by turns pleased and intimidated, proud and embarrassed by the widespread reaction to this announcement. This paper summarizes how the restrictions came to be lifted early, why (we think) it matters, what actions and reactions the TCP has observed in the last six months, and what new challenges--practical and philosophical--lie ahead for our work. We hope to apply what we learn from the ECCO-TCP release to the much larger EEBO-TCP release coming up in 2015. We also hope it will be useful to other projects publishing open editions or texts.

## Introduction

In April 2011, the Text Creation Partnership (TCP) announced that all restrictions regarding access to and distribution of 2,231 eighteenth century texts had been lifted, and that these files were now available to be used by anyone, for any purpose.<sup>1</sup> Previously, these files were accessible only to libraries that had paid in to support their creation. This small corpus is the first of the 40,096 files<sup>2</sup> published so far by the TCP to be made freely available, and the announcement created quite a stir.

---

<sup>1</sup>Ari Friedlander, "Text Creation Partnership Makes 18th-Century Texts Freely Available to the Public," Text Creation Partnership Blog, April 26, 2011/October 10, 2011. <http://textcreate.wordpress.com/2011/04/25/text-creation-partnership-makes-18th-century-texts-freely-available-to-the-public/>

<sup>2</sup>Includes texts published online from the ECCO-TCP, Evans-TCP, and both phases of EEBO-TCP projects as of October 19, 2011.

In the wake of this announcement, the TCP quickly learned not all users had the same understanding of the term “freely available.” We discovered, to our dismay, that we were actually quite unprepared to support the expectations of some users, while others impressed us with their innovation, industry, and ideas for working with these texts. This paper summarizes how the restrictions came to be lifted early, why (we think) it matters, what actions and reactions the TCP has observed in the last six months, and what new challenges--practical and philosophical--lie ahead for our work. We hope to apply what we learn from the ECCO-TCP release to the much larger EEBO-TCP release coming up in 2015. We also hope it will be useful to other projects publishing open editions or texts.

## Main Body

### I. What is the Text Creation Partnership?

The TCP is a library-based initiative to produce SGML/XML-encoded editions of early print books, specifically, books from the [Early English Books Online](#) (EEBO), [Eighteenth Century Collections Online](#) (ECCO), and [Evans Early American Imprints](#) databases. The TCP texts are used behind the scenes in many different projects and platforms: If you have ever searched the full text of a book in EEBO, you have used texts produced by the TCP. In addition, if you have experimented with the [University of Illinois' MONK workbench](#), searched Northwestern University's implementation of [PhiloLogic](#), perused [Witches of Early Modern England](#), explored [18thConnect](#), or inspected the [Holinshed Project](#), you have used the TCP texts. The TCP produces texts (data) that are used in many places and platforms. This project has never been especially platform or interface focused, and has never expected that the texts would be available only through a single access point--a point that will be important later.

Begun in 1999, the TCP is a collaboration between commercial publishers, academic libraries, and researchers, and aims to balance the interests of its stakeholders so that all benefit from the relationship. Publishers grant the TCP permission to make liberal use of the page images that make up products such as EEBO. The TCP engages third-party vendors to manually key in the words on these page images in order to create machine-readable text with a level of accuracy greater than 99.95%. The vendors also add basic SGML mark up using a customized schema derived from the Text Encoding Initiative (TEI) Guidelines P3.<sup>3</sup> Staff at the University of Michigan and Oxford University review the keyed and encoded text, verifying its accuracy and correcting and enhancing the markup as needed.

Our commercial partners provide significant financial and promotional support, as the work of the TCP has the potential to add great value to their product. The rest of the TCP's work is funded by

---

<sup>3</sup>Contracts with vendors, as well as the desire to maintain consistency across the corpus, have prevented us from migrating to TEI P5 while production is still ongoing. However, as we shall see, migration to P5 is just one possibility on the horizon for these texts as they are released.

academic libraries that join the project as partners. These institutions pay a one-time fee between \$12,500 and \$50,000 based on the size of the institution,<sup>4</sup> gaining immediate access to all existing texts, as well as the promise of periodic updates as more texts are produced and released. Partner libraries co-own the files produced by the TCP, and have the right to load them and treat them as if created locally. Partners agree not to grant non-partner institutions access to the texts until the end of a defined “exclusivity period,” during which each commercial publisher has the exclusive right to distribute the texts. Central to the mission of the TCP, after the end of the exclusivity period (typically five years after the end of production), all restrictions are lifted, and the texts are made freely available for anyone to use. Because TCP is still actively creating new texts, we have not yet reached the official end of the exclusivity period for any of our texts. The biggest existing body of texts, the 25,353 EEBO-TCP texts that we call “Phase I” will have their restrictions lifted January 1, 2015.

## II. ECCO-TCP

Initially, the goal of the TCP was quite narrowly defined: to key and encode 25,000 texts from EEBO, which today includes digital scans of more than 125,000 books listed in Pollard & Redgrave's Short-Title Catalogue (1475-1640), Wing's Short-Title Catalogue (1641-1700), the revised editions of each of these, the Thomason Tracts (1640-1661) collection and the Early English Books Tract Supplement.<sup>5</sup> Indeed, the histories of EEBO and TCP are closely intertwined: EEBO was launched in 1998, and EEBO-TCP in 1999. Around 2005, the TCP sought to expand beyond EEBO and create searchable text archives for other time periods: namely, Gale's Eighteenth Century Collections online and Readex's Evans Early American Imprints. Both of these initiatives were modeled on EEBO-TCP, but with around 40<sup>6</sup> and 90<sup>7</sup> partners respectively, compared to more than 160 in EEBO-TCP Phase I, neither project was as successful in gaining support from libraries. Due to lack of uptake, active production on ECCO-TCP ceased in 2009. As a result, by mid-2010, the TCP had produced 2,231 encoded texts that were, so to speak, all dressed up with nowhere to go. These digital editions had been painstakingly created, but represented only slightly more than 1% of all the volumes in ECCO,<sup>8</sup> and could only be used by a comparatively small group of people.

---

<sup>4</sup>“EEBO-TCP Partnership.” Text Creation Partnership, October 10, 2011. <http://www.lib.umich.edu/tcp/eebo/partner.html>

<sup>5</sup>“About EEBO.” Early English Books Online, October 10, 2011. <http://eebo.chadwyck.com/marketing/about.htm>

<sup>6</sup>“ECCO-TCP Participating Institutions.” Text Creation Partnership, October 10, 2011. <http://www.lib.umich.edu/tcp/ecco/status.html>

<sup>7</sup>“Evans-TCP Participating Institutions.” Text Creation Partnership, October 10, 2011. <http://www.lib.umich.edu/tcp/evans/status.html>

<sup>8</sup>According to Gale Digital Collections, as of October 2011, ECCO contains 182,898 titles and 205,639 volumes. “Gale Catalog-Eighteenth Century Collections Online.” Gale Cengage Learning, October 31, 2011. <http://www.gale.cengage.com/servlet/ItemDetailServlet?region=9&imprint=000&titleCode=GALEN7&type=4&id=190607>

This frustrating limbo prompted us to re-examine the TCP's longstanding business model, in particular, the exclusivity period described above. In the best case, this period of restricted use and distribution serves the needs of all those who have invested in the project:

- supporting libraries get the privilege of advance access to the texts
- supporting corporations get a window to make some money from their significant investment in the TCP
- the long-term interests of all libraries and scholars are protected by a very specific limited term of restricted use, after which restrictions are lifted.

In the case of ECCO-TCP, all three of these aspects were failing:

- the TCP was not seeking new library partners, so the incentive of gaining early access made no difference
- Because of particulars of Gale's technology, Gale had not incorporated the TCP texts into the ECCO interface and was not actively selling access to the texts.<sup>9</sup>
- Review of the agreement between Gale and the University of Michigan revealed that the terms of the exclusivity period were somewhat vaguely defined--in particular, there was no specific procedure for how to proceed if the project ended early.

Early in 2011, the TCP approached Gale with a proposal that all restrictions on the 2,231 ECCO-TCP files be lifted. Gale agreed, granting us permission to make these texts freely available, and to invite others to do the same.

### III. Reactions: What did people think?

On April 25, 2011, the TCP announced to a handful of outlets that restrictions had been lifted on the texts: our own blog<sup>10</sup>, the TCP's Twitterstream<sup>11</sup>, and to listservs such as SHARP,<sup>12</sup> LES-L, and TEI-L<sup>13</sup>, to a flurry of reactions. Within a month of the release, we received 12 direct requests for the files, and became aware of a handful of others working with the texts on their

---

<sup>9</sup>Gale's technological infrastructure includes marking plain text generated by optical character recognition (OCR) with coordinates to link each word to its proper location on the corresponding page image. Gale was not able to apply these coordinates to the encoded ECCO-TCP texts, so these texts were not folded back into the ECCO product as the EEBO-TCP texts were incorporated into EEBO.

<sup>10</sup>Friedlander, Text Creation Partnership Makes 18th-Century Texts Freely Available, October 10, 2011.

<sup>11</sup>@TCPStream, "Okay, here's the big news: TCP and Gale lift proprietary restrictions on 2,231 ECCO-TCP texts effective immediately." Twitter.com, April 26, 2011/October 10, 2011. <http://twitter.com/#/TCPstream/status/62537897181585408>

<sup>12</sup>Ari Friedlander, "Text Creation Partnership Makes 18th Century Texts Freely Available to the Public." Email, April 25, 2011. <https://listserv.indiana.edu/cgi-bin/wa-iub.exe?A2=ind1104&L=SHARP-L&P=R4757&D=0&H=0&O=T&T=0>

<sup>13</sup>Rebecca Welzenbach, "Text Creation Partnership Makes 18th Century Texts Freely Available to the Public." Email, April 25, 2011. "<http://listserv.brown.edu/archives/cgi-bin/wa?A2=ind1104&L=TEI-L&T=0&F=&S=&P=14587>

own. (Unfortunately, we have not tracked the number of downloads since the files have been made available online on our site or elsewhere, so we do not have a good sense of the number of users beyond those who get in touch with us). Responses ranged from delighted to frustrated and, perhaps most telling of all, confused. It is important to note that the evidence of reactions below is purely anecdotal. The TCP has not administered a survey or otherwise formally gathered responses, nor do the responses we have gathered constitute a statistically significant sample of all users. Instead, these provide a snapshot of reactions from users whose interest in the project was strong enough that they sought us out to seek help, clarification, or share their ideas and suggestions.

### *Delight*

Power users, such as digital humanities scholars, long-time friends and users of the TCP, and experts in XML and the Text Encoding Initiative (TEI) Guidelines offered enthusiastic responses thanking us for making these texts available, asking about how to get a copy of the corpus, and expressing ideas about what to do with them. For most of these knowledgeable creators and manipulators of electronic resources, their interest was less in 18th-century history or literature than it was in having a large set of consistently encoded texts to work with.

A perfect example of this user is Sebastian Rahtz at Oxford University Computing Services who, within a week of the release, had transformed all of the ECCO-TCP texts to TEI P5 and EPub, and made these available for download from Oxford.

### *Frustration*

In contrast, some scholars and librarians were frustrated because their expectations and assumptions about the meaning of “freely available” and “unrestricted access” differed from what the TCP was prepared to provide. This group consisted of those who are primarily consumers of electronic texts. That is, scholars of the 18th century who rely on the corpus for their research, and librarians who think of ECCO-TCP as a resource to which they subscribe, rather than a data set whose creation they have helped to fund. For these users, who want to quickly get at the content of the texts, not work with the data itself, XML files are not useful unless they are indexed by a search engine and presented in a web interface.

An example of a response along these lines was posted as a [comment to the Text Creation Partnership blog](#) shortly after the release was announced, by a librarian writing on behalf of a faculty member: “please send info as it becomes available on website upon which public can access the ECCO TCP freely available texts.”

### *Confusion*

Most telling of all the reactions, though, were those that were simply confused. A conversation that took place on the [Open humanities list between April 26-28, 2011](#) opened my eyes to how bewildering this announcement was to many people (it also reveals remarkable community spirit and commitment to sharing resources). The confusion we have seen breaks down into several categories, all revealing areas where the TCP was not as clear as it could or should have been in making this

announcement--sometimes because we were not yet quite sure what we meant ourselves. The most confusing aspects of this announcement seemed to be:

- Differentiating ECCO-TCP from other (still restricted) projects such as EEBO-TCP
- Differentiating TCP texts from ECCO commercial database
- Assumption that a platform/interface would be provided to search the corpus
- What are the terms under which these texts can be used?
- Where can the files be retrieved?

Now, it is simply a fact that the balance of partnerships and cooperative effort, restrictions and free access that make up the TCP, is inevitably confusing and difficult to summarize in a press release. Nevertheless, for at least half of the points above, the TCP should have prepared answers (and the resources to go with them) before announcing the release of the texts. As it happens, the flurry of questions following this announcement prompted us to publish another blog post several days after the release, explaining just what we meant by declaring the texts “freely accessible.”<sup>14</sup>

#### **IV. Actions: What did people do?**

Despite these challenges, in the six months since the restrictions were lifted, many people have managed to do interesting and innovative work with the ECCO-TCP texts, manipulating them and making them available in a number of different ways. At the TCP, we are aware of at least seven ways that these texts, or information about them, are currently being shared and published online:

##### *A. TCP's encoded text files*

The full set of ECCO-TCP files produced by the TCP, both in their original SGML format and in a “displayable” XML version, can be downloaded from the University of Michigan.<sup>15</sup> In truly raw form, these files are bundled into the original batches in which they were produced and released, ready for loading into DLXS, the publishing platform used by the Michigan and Oxford groups (see below).

##### *B. University of Michigan's DLXS interface*

Since their creation and initial release, the ECCO-TCP texts have been searchable and browsable (according to a number of criteria that take special advantage of the XML structure of the text) in a DLXS collection hosted at the University of Michigan.<sup>16</sup> This implementation of the texts is the only one that links the encoded text, page by page, to the corresponding page image in ECCO. Because

---

<sup>14</sup>Ari Friedlander, “What the Public Release of the ECCO-TCP Texts Means for You, Now and in the Future.” Text Creation Partnership Blog, April 27, 2011/October 10, 2011. <http://textcreate.wordpress.com/2011/04/27/what-the-public-release-of-ecco-tcp-texts-means-for-you-now-and-in-the-future/>.

<sup>15</sup>“ECCO-TCP Files Currently Available.” Text Creation Partnership, May 11, 2011/October 10, 2011. [http://www.lib.umich.edu/tcp/docs/texts/ecco\\_files.html](http://www.lib.umich.edu/tcp/docs/texts/ecco_files.html)

<sup>16</sup>“Eighteenth Century Collections Online Text Creation Partnership.” University of Michigan Library. October 10, 2011. <http://quod.lib.umich.edu/e/ecco/>.

the ECCO page images are available only to institutions that subscribe to ECCO, this interface is currently accessible only to original ECCO-TCP partners.

### *C. 18thConnect's plain text*

About a year before restrictions were lifted on the ECCO-TCP texts, the TCP provided a set of the texts to 18thConnect, an initiative to review and federate search across digital resources related to the 18th Century. 18thConnect is interested in doing research to improve optical character recognition (OCR) on 18th century texts, and so used the TCP texts as a control group against which to measure their success with OCR.<sup>17</sup> As part of their work with these texts, they stripped out all of the tags and generated a set of plain text files. When the restrictions were lifted, they offered to distribute these upon request.

### *D. Comprehensive Knowledge Archive Network (CKAN) entry*

John Levin requested the plain text and metadata for the files from 18thConnect and created an ECCO-TCP entry on CKAN,<sup>18</sup> “a registry of open knowledge datasets” that “makes it easy to find, share and reuse open content and data, especially in ways that are machine automatable.”<sup>19</sup>

### *E. ARTFL's searchable interface*

The Project for American and French Research on the Treasury of the French Language (ARTFL), hosted at the University of Chicago, has already had the ECCO-TCP texts indexed in its PhiloLogic search engine for some time.<sup>20</sup> Since the restrictions on the texts were lifted, ARTFL has removed the wall in front of these texts and made them available for the public to search and view.

### *F. Oxford's Transformations*

Sebastian Rahtz of Oxford University Computing Services has transformed each of the ECCO-TCP texts to TEI P5, EPUB, HTML, and plain text. All of the files are available to download from Oxford University Computing Services.<sup>21</sup>

### *G. Linked, open data on KASABI*

Keith Alexander downloaded the metadata provided by John Levin via CKAN, and began to create an open, linked dataset from it, that will ultimately allow information about the ECCO-TCP titles to

---

<sup>17</sup>“Gale and 18thConnect Partner to Improve Access to Eighteenth Century Documents.” 18thConnect: Eighteenth-century Scholarship Online. November 4, 2010/October 30, 2011. <http://www.18thconnect.org/news/?p=19>.

<sup>18</sup>“TCP ECCO 18th century texts.” CKAN: the Data Hub, September 2, 2011/October 10, 2011. <http://thedatahub.org/dataset/tcp-ecco-18th-century-texts>.

<sup>19</sup>“About CKAN.” CKAN: the Data Hub. October 10, 2011. <http://thedatahub.org/about>.

<sup>20</sup>“ECCO-TCP.” The ARTFL Project, October 10, 2011. <http://artfl-project.uchicago.edu/content/ecco-tcp>.

<sup>21</sup>Sebastian Rahtz. “Oxford conversion of ECCO Files.” Oxford University Computing Services, October 10, 2011. <http://tei.oucs.ox.ac.uk/ecco/>.



be connected with other information, such as place and person names, dates, and more. This data is available from KASABI.<sup>22</sup>

#### *H. Discussions, research, and other uses:*

- John Levin led a session at the Open Knowledge Foundation's TextCamp in August,<sup>23</sup> where participants processed sample texts from ECCO-TCP using the IBM concordance and visualization program [Many Eyes](#).
- Martin Mueller at Northwestern University has been working on a frequency-based dictionary of the words that occur in ECCO-TCP, with the aim that it might be used to identify and correct predictable errors (or errors with predictable solutions) in the text.

### **V. Reflection**

This proliferation of uses, of sources, seems to me a classic example of an open access/open source success story: by making your work available to others, all benefit from results that go beyond what any one person or organization could achieve alone. In fact, it is this way of thinking that made the TCP possible in the first place: though each partner directly funds the conversion of only a handful of texts, all benefit from access to the full corpus. The more people become involved, the more complete the corpus will be. But are we treading in dangerous waters by scattering the texts to the wind in this way? Is there anything we can do to control how they are made available online and even if there is, should we expend resources attempting to do so?

Alastair Dunning at the JISC digitization programme wrote a blog post on just this problem last July. Citing at least four potential entry points for EEBO-TCP, Dunning wrote:

But this is part of a larger problem. If there are multiple versions of the original content, then which one is the one you use? In fact it's not only about the content. Which platform works quickest? Which gives the most 'accurate' search results? Which one provides enhanced tools for analysis? Which gives the best results for your particular area of research? Where do you send your students? Which one do you cite?

Most importantly, which one do you trust? And why?<sup>24</sup>

These are good questions, and indeed questions that a scholar searching a digital corpus ought to ask (even, or perhaps especially, if the resource exists in only one place rather than many). These issues will become apparent to ECCO-TCP users, if they have not already. We are certainly experiencing

---

<sup>22</sup>"ECCO-TCP Eighteenth Century Collections Online Texts." KASABI. October 30, 2011. <http://data.kasabi.com/dataset/ecco-tcp-eighteenth-century-collections-online-texts>

<sup>23</sup>"Text Camp 2011." EventBrite.com, October 30, 2011. <http://textcamp2011.eventbrite.com/>.

<sup>24</sup>Alastair Dunning, "Crowdsourcing and Variant Digital Editions--Some Troubles Ahead." JISC Digitisation Programme. July 18, 2011/October 10, 2011. <http://digitisation.jiscinvolve.org/wp/2011/07/18/crowdsourcing-and-variant-digital-editions-some-troubles-ahead/>.



some anxiety that users will wind up with data altered in ways we do not know about and cannot correct. On the other hand, it has always been our intention that the TCP texts would not be an end in and of themselves, but that they would become the grist for dozens of other mills. As I wrote in a blog post responding to the JISC post, “If the texts produced by the TCP remain inert and static—if we end up with a single destination rather than a starting point for new scholarship—we will not have done our job.”<sup>25</sup>

In order to think about this problem in a productive way, it helps to separate the idea of data from the platform through which it is presented to an end user. This is not a new idea, indeed, it is central to the philosophy of XML encoding. But digital editions are so frequently published as a final product--and indeed, preliminary discussions about distributing the TCP texts have focused so much on the features such a system ought to support--that is easy to forget that the data might stand alone. I propose that for each TCP initiative, as the texts become freely available, the TCP ought to provide a reliable, flexible, robust source for the data. It need not be the only source, but it should be recognized as the source most related to the TCP so that users know they can rely upon it. On the other hand, I propose that TCP collections do *not* require an authoritative platform or interface. Indeed, many various interfaces or entry points might be desirable for various reasons, and one is not better than the others.

#### *Data source*

I want to build on an idea developed by Gabriel Bodard and Juan Garcés, that digital editions should not only be open access, but that they should be licensed much like open source software.<sup>26</sup> That is, the edition should not only be freely readable to an audience, but users should also have the opportunity to download the underlying files--the code--in order to inspect, de-bug, and modify it for themselves. Bodard and Garcés argue that:

Open Source is not so much a business model with exclusively economic implications as a strategy based on the belief that cultural advances are made by building upon the creations and publications of those who came before us. Without full access to the raw code, the documentation and the methodological statement that makes an experiment or a solution reproducible, a given publication is a dead end; it cannot be built upon.<sup>27</sup>

As they themselves note, this idea is not new to the text encoding community. Indeed, the idea that the XML underlying a digital edition ought to be visible and available to download is commonly

---

<sup>25</sup>Rebecca Welzenbach, “Our thoughts on “Crowdsourcing and Variant Digital Editions —some troubles ahead” (1/2).” TCP News & Views. July 21, 2011/October 10, 2011. <http://textcreate.wordpress.com/2011/07/21/our-thoughts-on-crowdsourcing-and-variant-digital-editions-%E2%80%94some-troubles-ahead-12/>.

<sup>26</sup>Gabriel Bodard and Juan Garcés, “Open Source Critical Editions: A Rationale,” in *Text Editing, Print and the Digital World*, ed. Marilyn Deegan and Kathryn Sutherland. (Abingdon: Ashgate Publishing Group, 2009) p 84.

<sup>27</sup>Gabriel Bodard and Juan Garcés, “Open Source Critical Editions: A Rationale,” in *Text Editing, Print and the Digital World*, ed. Marilyn Deegan and Kathryn Sutherland. (Abingdon: Ashgate Publishing Group, 2009) p 84.

shared, if less frequently carried out, and, as I hope I have demonstrated, has always been part of the TCP's mission. However, it would be even more valuable than simply offering the files to download to provide a space where modified texts could also be shared back with the same community. This year at the Text Encoding Initiative Members Meeting and Conference, Christian Wittern proposed a model for digital editions that resembles a code repository, more specifically, a distributed version control system, that would work in this way:

At the core, this is a plain text format that only contains very few traces of markup, but serves to make the textual content available to the reader. The text is published through a distributed version control system, which allows the researcher to create branches, annotate, edit or translate the text without losing the connection to the established digital edition and thus to all the other researchers, that are working on this text.<sup>28</sup>

Such a system would address a number of challenges facing the TCP:

- Most importantly, the TCP would have a stable, publicly accessible site from which users can download as much or as little of the corpus as they wish
- Enhancements to the text that are subjective--or that simply add more detail or granularity than is desired by some users, can be shared with all, without interfering with the basic digital "reading edition."
- Actual corrections (such as the completion of letters marked by the keyers as illegible) could be incorporated back into the master version.

### *Platform*

Because the version control system imagined above does allow for many sibling branches in addition to the original source data, it should not attempt to deal with platforms to present the text. There are many reasons that the TCP should avoid imposing an authoritative platform or entry point for the texts:

- Organizations such as JISC Collections, ARTFL, and MONK may want to index these texts along with many others to which they have access in order to enable cross-collection searching--but these groups all have access to different data sets. For this reason, one authoritative searchable interface is not sufficient. We should not prevent the TCP texts from being included in any or all of these other projects.
- Some users only want to work with and publish a handful of texts. We should allow (encourage!) any number of TCP texts to be included in thematic or otherwise distinct digital

---

<sup>28</sup>Christian Wittern, "Beyond TEI: Returning the text to the reader." TEI Members Meeting 2011 Abstracts. 16 October 2011. [http://www.zde.uni-wuerzburg.de/tei\\_mm\\_2011/abstracts/abstracts\\_micropapers/#c249146](http://www.zde.uni-wuerzburg.de/tei_mm_2011/abstracts/abstracts_micropapers/#c249146).

projects, such as Witches of Early Modern England.<sup>29</sup>

- For those who just want to read, the texts should be made available and searchable on

---

<sup>29</sup>Witches in Early Modern England: A Digital Humanities Project for Unveiling Witchcraft Narratives. October 31, 2011. <http://witching.org/>.

relevant platforms, such as The Internet Archive<sup>30</sup> and wikisource<sup>31</sup>, as well in ebook format for download or purchase. The texts should be placed where people go to read; potential readers should not be required to find our website.

## Conclusions

There is considerable work ahead for the TCP to put the ECCO-TCP texts in good working order, and prepare for the 2015 release of the EEBO-TCP texts. Our action items might at least the following:

- Applying an appropriate license, such as the Creative Commons Public Domain mark,<sup>32</sup> so that those who find the files understand how they may be used
- Cleaning up the very raw files so that they stand alone as complete XML files that might reasonably be fed into any workflow
- Selecting and setting up a distributed version control system that will meet our needs
- Developing guidelines for the modification of the files, including specifications for which edits will count as corrections and be incorporated back into the master text, and which will be maintained as various optional branches

However, by focusing on developing a hub for the distribution of the encoded texts themselves, rather than on attempting to control the various platforms on which they might be made available, we will be able to do reasonable quality control without hindering innovation or imagination among scholars. In short, our approach should be to encourage users not to worry about where they discover the data, but where it comes from. Thus far, it has been extremely gratifying to see how many people are interested in working with these texts, and to learn from the expertise of our colleagues who have wasted no time in making them available in new ways. The ECCO-TCP texts will provide a useful sample with which to work as we look ahead to and prepare for the release of EEBO-TCP Phase I, coming in 2015.

## Bibliography

@TCPStream, “Okay, here's the big news: TCP and Gale lift proprietary restrictions on 2,231 ECCO-TCP texts effective immediately.” Twitter.com, April 26, 2011/October 10, 2011. <http://twitter.com/#!/TCPstream/status/62537897181585408>

“About CKAN.” CKAN: the Data Hub. October 10, 2011. <http://thedatahub.org/about>.

“About EEBO.” Early English Books Online. October 10, 2011. <http://eebo.chadwyck.com/marketing/about.htm>

---

<sup>30</sup>Internet Archive. 31 October, 2011. <http://www.archive.org/>.

<sup>31</sup>Wikisource. 31 October, 2011. [http://wikisource.org/wiki/Main\\_Page](http://wikisource.org/wiki/Main_Page)

<sup>32</sup>“Public Domain Mark 1.0,” Creative Commons, October 31, 2011. <http://creativecommons.org/publicdomain/mark/1.0/>.

Bodard, Gabriel and Juan Garcés, “Open Source Critical Editions: A Rationale,” in *Text Editing, Print and the Digital World*, edited by Marilyn Deegan and Kathryn Sutherland, 83-98. Abingdon: Ashgate Publishing Group, 2009.

Dunning, Alastair. “Crowdsourcing and Variant Digital Editions--Some Troubles Ahead.” JISC Digitisation Programme. July 18, 2011/October 10, 2011. <http://digitisation.jiscinvolve.org/wp/2011/07/18/crowdsourcing-and-variant-digital-editions-some-troubles-ahead/>.

“ECCO-TCP.” The ARTFL Project. October 10, 2011. <http://artfl-project.uchicago.edu/content/ecco-tcp>.

“ECCO TCP Eighteenth Century Collections Online Texts.” KASABI. October 30, 2011. <http://data.kasabi.com/dataset/ecco-tcp-eighteenth-century-collections-online-texts>.

“ECCO-TCP Files Currently Available.” Text Creation Partnership, May 11, 2011/October 10, 2011. [http://www.lib.umich.edu/tcp/docs/texts/ecco\\_files.html](http://www.lib.umich.edu/tcp/docs/texts/ecco_files.html)

“ECCO-TCP Participating Institutions.” Text Creation Partnership. October 10, 2011. <http://www.lib.umich.edu/tcp/ecco/status.html>.

“EEBO-TCP Partnership.” Text Creation Partnership. October 10, 2011. <http://www.lib.umich.edu/tcp/eebo/partner.html>.

“Eighteenth Century Collections Online Text Creation Partnership.” University of Michigan Library. October 10, 2011. <http://quod.lib.umich.edu/e/ecco/>.

“Evans-TCP Participating Institutions.” Text Creation Partnership. October 10, 2011. <http://www.lib.umich.edu/tcp/evans/status.html>.

“The Free Software Definition.” GNU Operating System. October 31, 2011. <http://www.gnu.org/philosophy/free-sw.html>.

Friedlander, Ari. “Text Creation Partnership Makes 18th Century Texts Freely Available to the Public.” Email. April 25, 2011. <https://listserv.indiana.edu/cgi-bin/wa-iub.exe?A2=ind1104&L=SHARP-L&P=R4757&D=0&H=0&O=T&T=0>.

Friedlander, Ari. “Text Creation Partnership Makes 18th-Century Texts Freely Available to the Public,” Text Creation Partnership Blog. April 26, 2011/October 10, 2011. <http://textcreate.wordpress.com/2011/04/25/text-creation-partnership-makes-18th-century-texts-freely-available-to-the-public/>.

Friedlander, Ari. “What the Public Release of the ECCO-TCP Texts Means for You, Now and in the Future.” Text Creation Partnership Blog. April 27, 2011/October 10, 2011. <http://textcreate.wordpress.com/2011/04/27/what-the-public-release-of-ecco-tcp-texts-means-for-you-now-and-in-the-future/>.

“Gale and 18thConnect Partner to Improve Access to Eighteenth Century Documents.”

18thConnect: Eighteenth-century Scholarship Online. November 4, 2010/October 30, 2011. <http://www.18thconnect.org/news/?p=19>.

“Gale Catalog-Eighteenth Century Collections Online.” Gale Cengage Learning, October 31, 2011. <http://www.gale.cengage.com/servlet/ItemDetailServlet?region=9&imprint=000&titleCode=GALEN7&type=4&id=190607>.

Internet Archive. 31 October, 2011. <http://www.archive.org/>.

“Public Domain Mark 1.0,” Creative Commons, October 31, 2011. <http://creativecommons.org/publicdomain/mark/1.0/>.

Rahtz, Sebastian. Sebastian Rahtz. “Oxford conversion of ECCO Files.” Oxford University Computing Services, October 10, 2011. <http://tei.oucs.ox.ac.uk/ecco/>.

“TCP ECCO 18th century texts.” CKAN: the Data Hub. September 2, 2011/October 10, 2011. <http://thedatahub.org/dataset/tcp-ecco-18th-century-texts>.

“Text Camp 2011.” EventBrite.com, October 30, 2011. <http://textcamp2011.eventbrite.com/>.

Welzenbach, Rebecca. “Our thoughts on “Crowdsourcing and Variant Digital Editions — some troubles ahead” (1/2).” TCP News & Views. July 21, 2011/October 10, 2011. <http://textcreate.wordpress.com/2011/07/21/our-thoughts-on-crowdsourcing-and-variant-digital-editions-%E2%80%94some-troubles-ahead-12/>.

Welzenbach, Rebecca. “Text Creation Partnership Makes 18th Century Texts Freely Available to the Public.” Email. April 25, 2011. <http://listserv.brown.edu/archives/cgi-bin/wa?A2=ind1104&L=TEI-L&T=0&F=&S=&P=14587>.

Wikisource. 31 October, 2011. [http://wikisource.org/wiki/Main\\_Page](http://wikisource.org/wiki/Main_Page).

Witches in Early Modern England: A Digital Humanities Project for Unveiling Witchcraft Narratives. October 31, 2011. <http://witching.org/>.

Wittern, Christian. “Beyond TEI: Returning the text to the reader.” TEI Members Meeting 2011 Abstracts. 16 October 2011. [http://www.zde.uni-wuerzburg.de/tei\\_mm\\_2011/abstracts/abstracts\\_micropapers/#c249146](http://www.zde.uni-wuerzburg.de/tei_mm_2011/abstracts/abstracts_micropapers/#c249146).