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
@TCPStream
tcp-info@umich.edu





« TCP Releases Over 4,000 New EEBO-TCP Texts
What the Public Release of ECCO-TCP Texts Means for You, Now and in the Future »

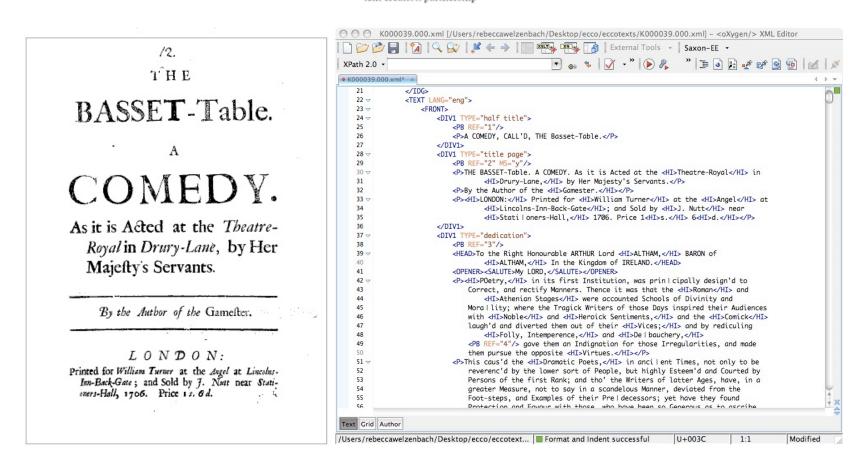
Text Creation Partnership Makes 18th-Century Texts Freely Available to the Public

By arifried

We are pleased to announce the opening to the public of 2,231 searchable keyed-text editions of books from Eighteenth Century Collections Online (ECCO). ECCO is an important research database that includes every significant English-language and foreign-language title printed in the United Kingdom during the 18th century, along with thousands of important works from the Americas. ECCO contains more than 32 million pages of text and over 205,000 individual volumes, all fully searchable. ECCO is published by Gale, part of Cengage Learning.



What is TCP?



Standardized, structured, searchable digital editions of early printed books

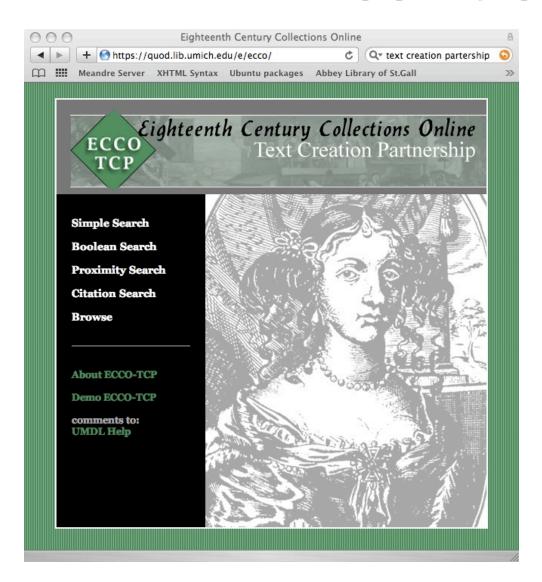


What is TCP?

- Transcribe and mark up works from EEBO, ECCO, and Evans, with permission and support from Proquest, Gale Cengage and Readex.
- Funded directly by more than 150 libraries, who co-own the files
- Eventually, all texts will be freely available for anyone to use.



ECCO-TCP



Result: 2,231 text files all dressed up and nowhere to go



All dressed up and nowhere to go

Functional exclusivity period:

- Supporting libraries get immediate access
- Supporting corporations have a window to see a return on their investment
- Clear limitations protect all libraries
 & scholars



All dressed up and nowhere to go

Functional exclusivity period:

- Supporting libraries get immediate access
- Supporting corporations have a window to see a return on their investment
- Clear limitations protect all libraries & scholars

Dysfunctional exclusivity period:

- •No new libraries signing up, production halted
- •Gale not interested in selling
- •Limits of exclusivity period unclear

When we asked, Gale Cengage said: "If you love them, set them free!" (or something like that)



And there was much rejoicing!

....sort of.



Reactions were mixed

From textperts: innovation

"You'll recall that ...I made TEI P5 versions, and ePub...In conversation with the folks at Apple, they said they'd like to get these ebooks into the Apple iBookstore, and promote them a bit."

"I have been noodling around with the ~1,400 ECCO texts that have...been linguistically annotated.... With the help of a...script, I...establish[ed] a very crude error rate for each document"



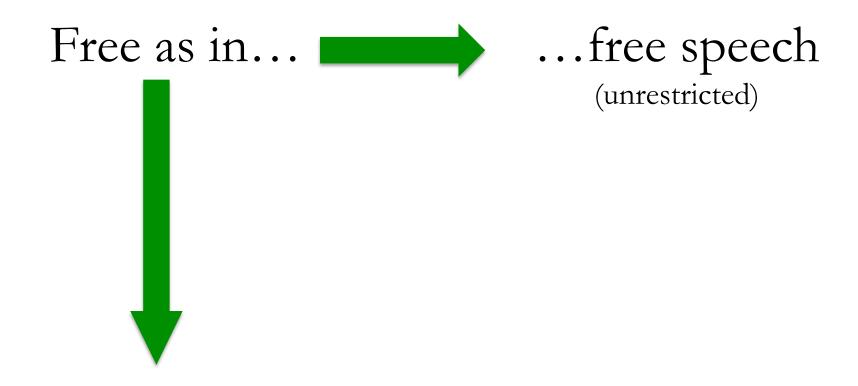
Reactions were mixed

From everyone else: frustration

"I read a press release the other day about how a certain number of texts from ECCO are now going to be available for free searching by the public in the TCP. When will they be available and what is the URL?"

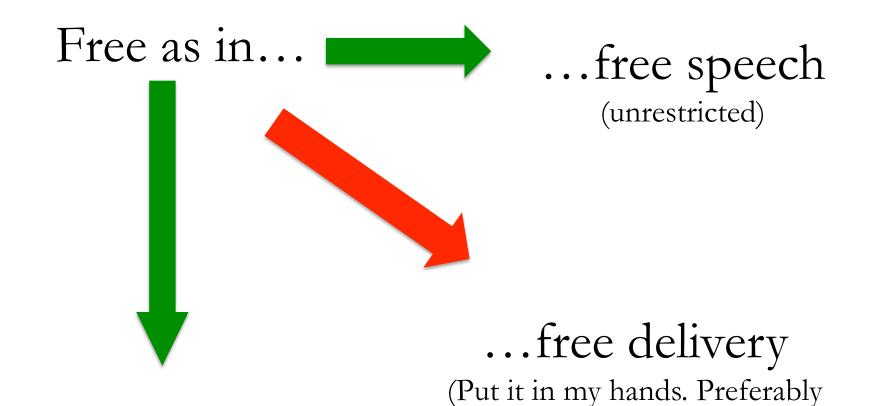
"What exactly does 'freely available to the public' mean, here and now?"





...free beer (doesn't cost any money)





assembled.)

...free beer (doesn't cost any money)



Anterotesis

Answering one question with another



← The Economics of JSTOR

Making the TCP-ECCO texts accessible

Posted on August 12, 2011 by john!

In April, the <u>Text Creation Partnership</u> released into the public domain over 2,000 eighteenth century works, in plain text. You can read more about this project and the texts on their blog:

TCP Releases Over 4,000 New EEBO-TCP Texts

What the Public Release of ECCO-TCP Texts Means for You, Now and in the Future

Unfortunately, they didn't make the texts easily accessible. To obtain them one had to apply by email to be subscribed to a Dropbox folder. There is a <u>database and search interface</u>, but it requires registration, and is unclear as to who qualifies for an account. I think that the database holds the marked up, XML texts, which have not (yet) been publicly released.

Search

Search)

Syndication

RSS Atom

Atom

> Follow @anterotesis

Archives

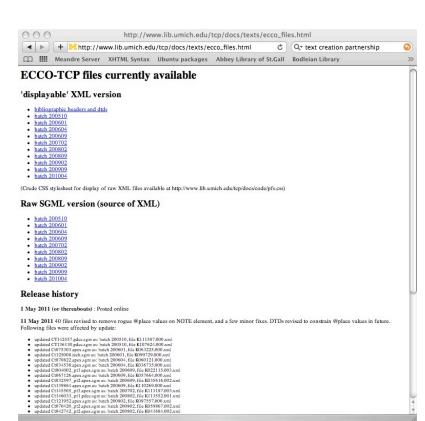
- August 2011 (1)
- July 2011 (1)
- May 2011 (2)
- April 2011 (1)
- March 2011 (2)
- February 2011 (1)
- December 2010 (2)
- November 2010 (1)
- October 2010 (1)
- August 2010 (1)
- July 2010 (5)
- June 2010 (1)



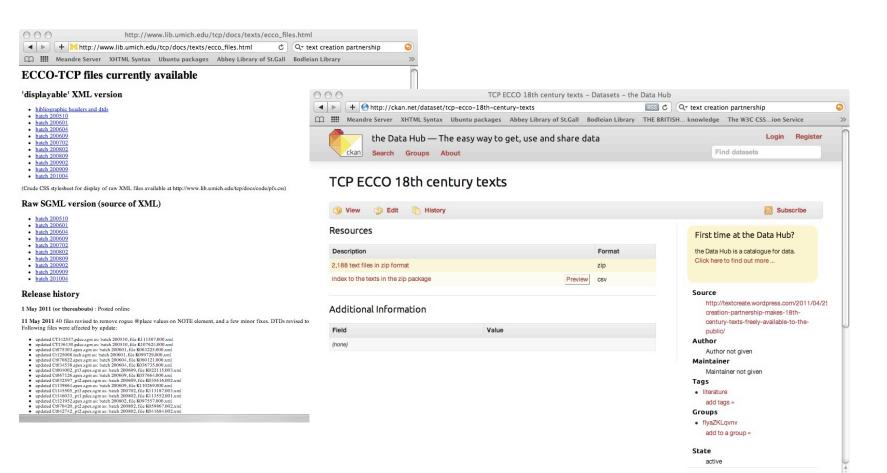
Two ways to think about representation

- Where and how are the ECCO-TCP texts represented online?
- How does the TCP represent itself, and its relationship to these texts? (And how do folks find that information?)

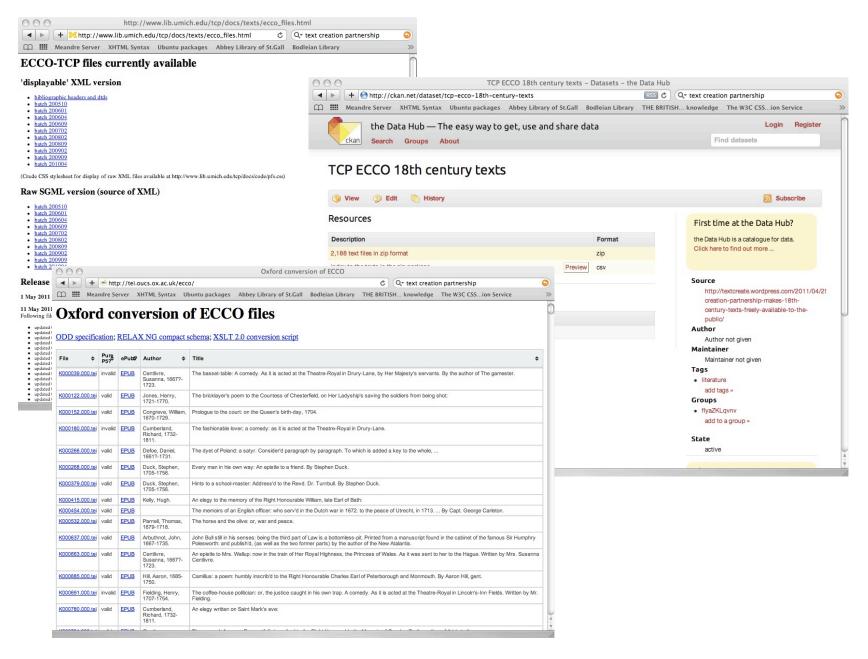




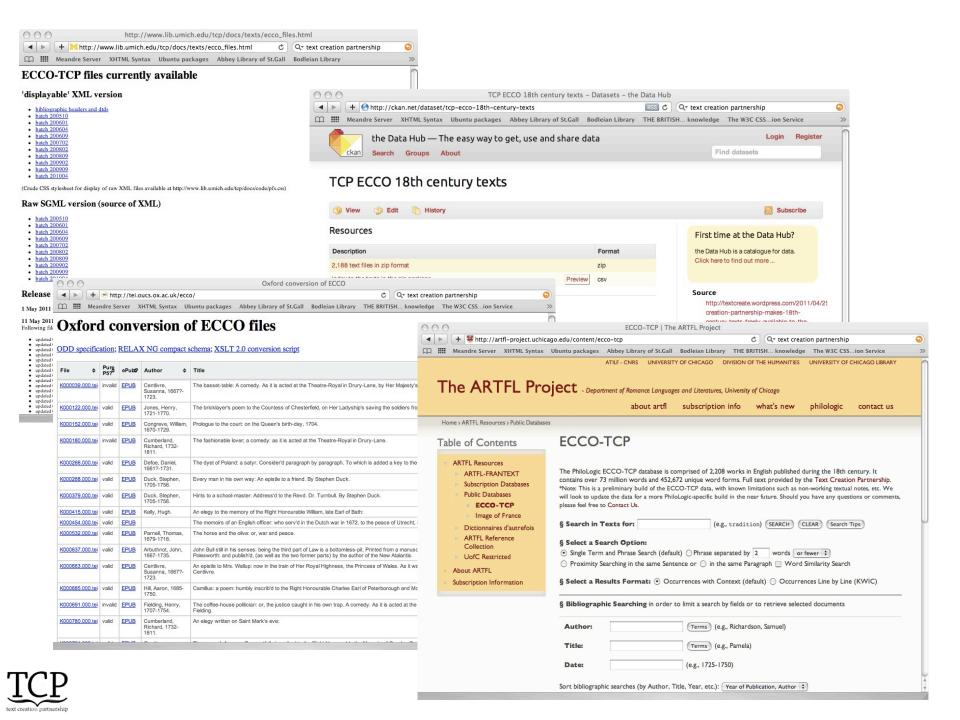


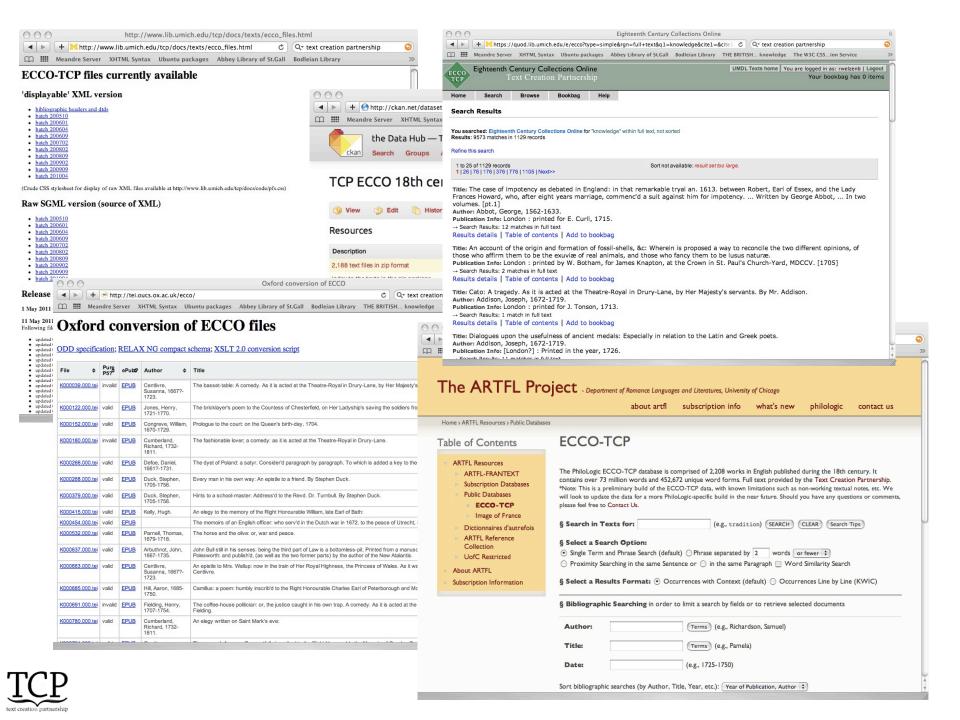


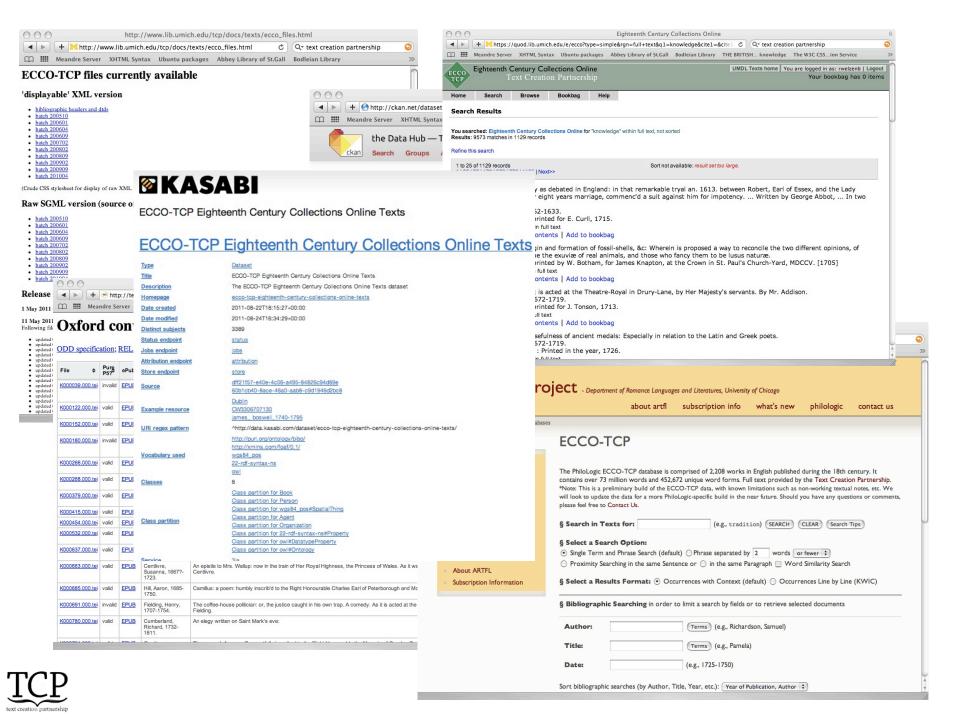


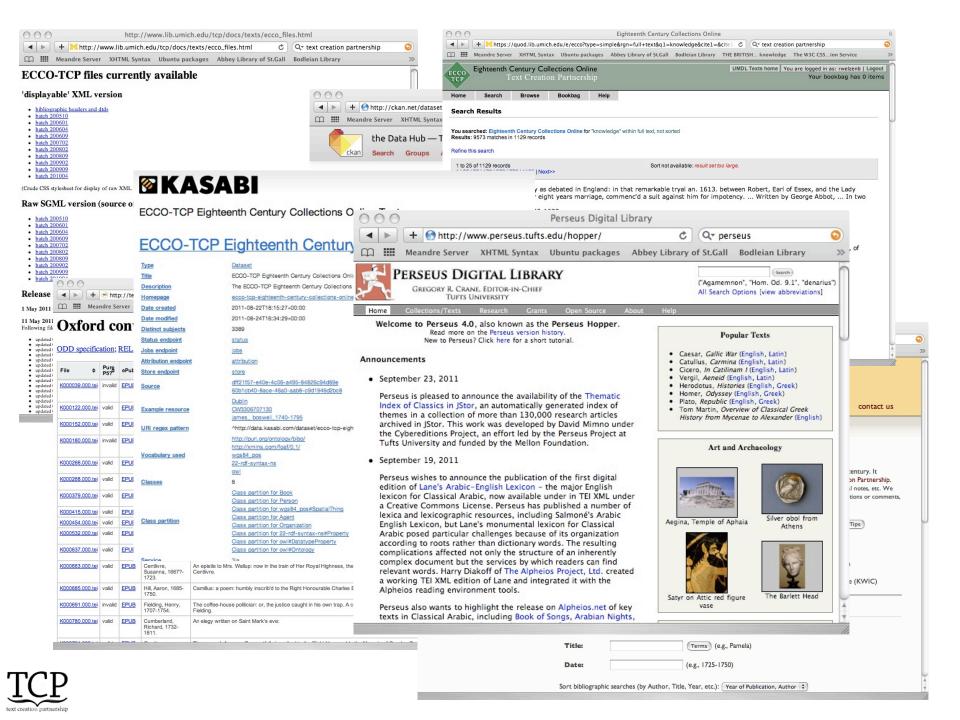












"But this is part of a larger problem. If there are multiple versions of the original content, then which one is the one you use? In fact it's not only about the content. Which platform works quickest? Which gives the most 'accurate' search results? Which one provides enhanced tools for analysis? Which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

"Crowdsourcing and Variant Digital Editions - some troubles ahead"

July 18, 2011



To roast a green Goose with green Sauce.

Roast your goose nicely; in the mean time make your sauce thus; take half a pint of the juice of sorrel, a spoonful of white wine, a little grated nutmeg, and some grated bread; boil this over a gentle sire, and sweeten it with pounded sugar to your taste; let your goose have a good froth on it before you take it up; put some good strong gravy in the dish, and the same in a boat. 'Garnish with lemon.



is your jacket. I thank you a thousand, thousand times. I shall never forget you as long as I live.

Mich Do not talk of thanks I am as