Laura Estill

@Laura\_Estill

Texas A&M University

#shakeass2016

**“EEBO-TCP: The Searchable (Print) Text and Manuscript Studies”**

Thanks, Jonathan, for bringing us together for this discussion.

Today, I’d like to talk about EEBO-TCP from the point of view of a manuscript scholar. Maybe you’re sitting in an archive with a manuscript that’s filled margin to margin. Or maybe you’re at a computer looking at an online facsimile of a verse miscellany. It might be instinctual, or it might be the amateur bibliographer in us (or, in some cases, the professional bibliographer), but one of the first responses to a miscellany is to catalogue and organize it: to attempt to reverse engineer its contents.

**SLIDE**

The face of early modern studies is changing—and one of our current shifts is a return to the archive and a re-evaluation of its contents. Digitization has changed archival research: we have increasing access to online facsimiles, for instance, at the Henslowe-Alleyn Project or the recently launched Shakespeare Documented site.[[1]](#footnote-1) There are not just more manuscript pages images online, but also more transcriptions online: such as the Early Modern Manuscripts Online project underway at the Folger.[[2]](#footnote-2)

Today, though, I want to discuss how it is not just manuscript digitization projects that change the way we undertake manuscript studies. It is also the unprecedented amount of text made digital and the EEBO-TCP that changes how we can undertake the study of historical handwritten texts. It might seem paradoxical, but this giant print digitization project facilitates our understanding of manuscript.

John Lavagnino, in his 2012 keynote at the EEBO-TCP conference, suggested three main ways scholars could use the TCP: “simple quotation finding,” “larger-scale trawl for materials” and “computational analyses.”[[3]](#footnote-3) Today, I’d like to focus on the first of Lavagnino’s categories, “simple quotation finding”—which, I would argue, is far from simple, and is the bedrock upon which much scholarship is based.

There are many problems with EEBO: terrible page images, lack of digitized text, and a price point that make it inaccessible to many scholars. EEBO-TCP solves some of those problems, but of course, introduces others. The text is searchable—but it is based off the terrible microfilm images, and the transcribers were non-native speakers instructed not to guess: the transcriptions are, in their way, incredibly accurate. They are a very accurate rendition of terrible page images.

**SLIDE**

 Here, for instance, is an example tweeted by Davenant’s Gondibot, a bot by Jacob Tootalian tweeting all of Gondibert.[[4]](#footnote-4) Of course, as Meag has shown, we need better digital texts: and not just so that we can have more accurate Twitter bots.

We need more digital editions (the *Folger Digital Anthology of Early Modern Drama*, which Meag has previewed for us, and *Digital Renaissance Editions* are two wonderful projects that come to mind).[[5]](#footnote-5) And while these are canon-expanding, they are still necessarily bound by a modern editor’s selection.

Having the most extensive, searchable corpus of early modern English Literature is important for our study of manuscripts, which in turn allows us to expand our notions of canonicity, literary importance, and textual culture.

**SLIDE**

One of the most common forms of early modern manuscript is the miscellany. As the names implies, these volumes have miscellaneous contents: commonplaces, poetry, receipts (recipes), accounts, letters, book lists, and more. A vast amount of text was copied into individual miscellanies: some of these circulated from person to person, for instance, at the universities.

Miscellanies and commonplace books tell us about the texts that early readers actually read and the ways they interacted with those texts. Readers in Shakespeare’s lifetime did not read only works collected in the *Norton Anthology of Renaissance Drama*. They were not limited to the texts we can buy in a Barnes & Noble (or even in the book room here at SAA). Early readers copied from William Peaps’ *Love in Extasie* alongside Shakespeare.[[6]](#footnote-6) Early readers were not always reading first editions—yet the EEBO-TCP corpus privileges the first printed version of a text.

Much of the work of scholarship about manuscript miscellanies is looking at the page and saying “what is this? how did it get here? when was it copied? by whom and for whom?” The “What is this?” question is both the easiest (the “simple quotation finding” that Lavagnino mentions) and the hardest.

Before digitized texts, scholars had to rely on their individual knowledge of early modern literature in order to identify unattributed selections. Now, we can turn to Google or *Literature Online*…or the EEBO-TCP. *Literature Online* is limited in two major ways: firstly, it’s paywalled (which TCP is not), and secondly, it’s, as its title suggests, limited to Literature. Limiting ourselves to the literary means applying an often anachronistic label (that is, “literature”): it also offers an incomplete (and wholly inadequate) understanding of early modern textual cultures. Google, on the other hand, can offer a good starting point, but its contents, similarly, rely on those 19th century editions often digitized by the internet archive or poorly OCR’d by Project Gutenberg. EEBO-TCP extends the searchable text from Gutenberg, LiOn, or Google, but still does not offer comprehensive searching. EEBO-TCP is a carefully selected corpus, but is far from representing all printed works in English. It is especially imperative that students and scholars recognize EEBO-TCP's (ever-expanding) limits: the size of its corpus, its metadata, and the search functionality.

**SLIDE**

With Beatrice Montedoro, I’m working on *DEx: A Database of Dramatic Extracts*, which will offer transcriptions of those parts from plays that early readers and audience members copied into their plays. Right now, when possible, we are linking to edited versions of these plays. We have been considering what we can do with the fantastic amounts of text made available by the TCP. However, our requirements for this project mean that we need, where possible, line numbers and modernized spellings for searchability—we need multiple editions of transcribed texts. We don’t want to host these plays: we want to link out to them.

**SLIDE**

And though you can access nicer interfaces than the straight-up TCP page (or the paywalled search at EEBO), for instance, this one from Wolfgang Meier from eXist,[[7]](#footnote-7) we still can’t figure out how to make this dataset meet our needs.

The EEBO-TCP, however, is invaluable to our project precisely because it allows the identification of sources that we would not otherwise be able to find. EEBO-TCP has introduced me to a wealth of plays that I would not otherwise have read, such as Sharpham’s *Cupid’s Whirligig*—plays that early readers, did, indeed, read and value—even if they are often overlooked today. I’m fortunate: drama was high on the TCP’s list of things to make searchable. We will understand manuscripts better when even more text is available: for instance, later searchable texts might help us find evidence of plays circulating before they were published or might offer evidence of earlier lost texts.

**SLIDE**

Ultimately, manuscript studies cannot be separated from texts and book history any more than manuscripts themselves can be disentangled from print sources in the early modern period. Continuing improvements made to the TCP (by, for instance, Martin Mueller and company) will only expand our understanding.

I will end with a provocation: can we imagine large-scale (perhaps crowd-sourced) improvements of the TCP? And/or how would we benefit from starting afresh with better facsimiles and transcriptions?

Thank you.

1. http://www.henslowe-alleyn.org.uk & http://shakespearedocumented.org [↑](#footnote-ref-1)
2. http://folgerpedia.folger.edu/Early\_Modern\_Manuscripts\_Online\_(EMMO) [↑](#footnote-ref-2)
3. Lavagnino slides: http://www.slideshare.net/jlavagnino/scholarship-in-the-eebotcp-age [↑](#footnote-ref-3)
4. https://twitter.com/gondibot/status/710926087539322882 [↑](#footnote-ref-4)
5. http://folgerpedia.folger.edu/A\_Digital\_Anthology\_of\_Early\_Modern\_English\_Drama & dre.uvic.ca [↑](#footnote-ref-5)
6. Such as Abraham Wright, in BL Add MS 22608 (compiled mid-seventeenth century). [↑](#footnote-ref-6)
7. http://showcases.exist-db.org/exist/apps/Showcases/index.html [↑](#footnote-ref-7)