

2016-11-03

The Right Stuff at the Right Cost for the Right Reasons

Welzenbach, Rebecca

<http://hdl.handle.net/2027.42/136646>

[Slide 1]

Good morning. As Mark noted, I am Rebecca Welzenbach, Director of Strategic Integration and partnerships at Michigan Publishing. We've heard from Stephen on some of the opportunities and benefits of leaving digitization "to the professionals," so to speak, and we've heard from Rachel about some of the challenges of tackling these projects library by library--from the standpoint of planning, funding, execution, and sustainability. Now, I'd like to tell you now about one approach first taken up by the University of Michigan Library more than 15 years ago: a public/private partnership that sought to make the most of both of these types of organizations.

The Text Creation Partnership, a project born at the University of Michigan Library, brought more than 150 contributing libraries into partnership with publishers such as ProQuest, Gale, and Readex to transcribe and markup in XML the text of many thousands of early printed works that had already been digitized as images by these commercial entities, and made available to libraries by way of subscription or license to a database. The first, and most successful, of these efforts was the the Early English Books-Text Creation Partnership--or EEBO-TCP.

The title of this panel is "The right stuff at the right cost for the right reasons," so I'll try to follow this structure to point out where this effort succeeded, and where we faced challenges.

[slide 2]

So, first: The Right Stuff.

In the case of the Early English Books Online-Text Creation Partnership, the selection of the product/collection was key. In this case, we began with a product that both parties were already deeply invested in--the longstanding Early English Books microfilm collection, just emerging in digital form as Early English Books Online. Timing and circumstances were right for the kind of work that needed to be done: the digital EEBO platform opened the door to the idea that the tool

could be much more powerful than just offering search across catalog records. This is really evident in the first reviews of EEBO, which start to appear in 2001 and 2002--they point out the lack of searchable text--taking for granted that it would exist in a digital project! but the Optical Character Recognition technology didn't yet (indeed, still doesn't) exist to capture the text accurately in a cost effective and timely way. In other words, because of the nature of the material, and the moment in which it appeared, EEBO not only inspired, but demanded this kind of partnership. Without it, there was no accurate, effective, affordable solution to produce the searchable full text, that libraries would buy into as an add-on to their existing EEBO license.

[slide 3]

And that brings us to our second: At the right cost

- What was the business model?
 - What were the costs? Mostly: employing a vendor (or, frequently, several vendors) to transcribe and mark up the texts. Employing teams of editors at Michigan and Oxford, and their manager, to proofread and edit the texts. ProQuest invested about 20% of the cost.
 - 80% of the cost came from the many contributing libraries, who paid between \$50K and \$15,000 depending on the size of the library. Most of the time these payments were spread over five years.
- What did the contributors get?
 - Contributing libraries gained immediate access to the texts--that is, the XML files generated by the EEBO-TCP team--and also became "co-owners" of them they could reuse them as though they'd been created on their own campuses, as well as accessing them right in the EEBO platform. And, they invested with the

promise that ultimately the texts would be both owned by libraries and freely available to all users.

- Michigan and Oxford, as the creators/managers/producers of the project, got pretty much unfettered access to the images; financial support that helped kickstart/anchor the work as new library partners were being sought.
- ProQuest got high quality transcribed and encoded text for their product at a fraction of what it cost to create it, and endorsed by librarians. 5 year embargo to sell the completed project to libraries that hadn't already signed on to recoup investment.
- Worth noting that in this case, economies of scale weren't really a benefit or even a relevant to working with the commercial vendor: Michigan managed all of the vendor contracting and production work. We did, of course, benefit immensely from the fact of EEBO already existing--with MARC Records, image sets, etc.

OK, so--given these costs and benefits, why? What was the point?

For the right reasons....

- Both parties agreed on:
 - Creating a text accurate enough, and with enough markup, that the text could function without the image--good for searching, but also for reading/display
- Cognitive dissonance around:
 - Library ownership/access to texts, which could go into the public domain
 - Enhance the value of the PQ product, and charge additional for the enhanced version

This partnership was suitable for this project, but we have struggled (in some ways; less so in others) to replicate the model. Within the Text Creation Partnership, we tried to replicate the

model with Gale's Eighteenth Century Collections Online and Readex's Evans Early American Imprints, and while those companies were also good to work with, it was the wrong stuff: the mutual need wasn't there. For example, unlike EEBO, ECCO already offered searchable automatically generated fully digital text. Our text was more accurate, but the difference wasn't enough to make a difference--especially since users already had a full text search option, Gale couldn't sell a new product.

Differences today? Expectation of open data, DH uses, examples of crowd-sourcing, etc.

Likewise, libraries weren't eager to invest because users didn't perceive the need. I am looking forward to hearing next from Peggy about how Reveal Digital is expanding this model of partnership.