**POPHAM — Text**

<0 figures>

**Text + Creation + Partnership: Whatever Happened to the Best Laid Plans of EEBO-TCP?**

Popham, M.

When the Early English Books Online Text Creation Partnership (EEBO-TCP) was proposed by Mark Sandler of University of Michigan in 1999, the intention was to raise sufficient funds to produce accurate, marked-up, full-text transcriptions of 25,000 titles from ProQuest’s EEBO database of images. The main partners in the collaboration were the library at the University of Michigan, the Bodleian Library at the University of Oxford, and ProQuest—which, in return for making the images available to the EEBO-TCP partners, would receive a five-year exclusivity period to exploit the full-text data. Once that period elapsed, these 25,000 carefully produced texts would be released into the public domain for use by anyone in the global scholarly community and beyond. That was the dream. This paper will describe the practical realities and challenges that were faced to make this dream come true.

On 1 January 2015, the 25,000 electronic texts produced by Phase I of EEBO-TCP were released into the public domain. They represented the concerted efforts of a large number of individuals employed at both Michigan and Oxford—more than 20 editors who, between them, had contributed over 100 person-years of work towards the objectives of this $9,000,000+ endeavour. In terms of scale and scope of its ambition and its outputs, the Early English Books Online Text Creation Partnership should be recognized as a seminal project in the development of the digital humanities.

Perhaps one of the most interesting aspects of the EEBO-TCP was the mixed nature of its underlying funding model. The anticipated cost of $9M over five years was felt to be too substantial to appeal to any single foundation or funding body. Moreover, the formal collaboration with the commercial e-publisher, ProQuest, complicated matters further and, not unreasonably, raised questions about whether this could ever be a collaboration of equals.

In the United States, colleagues at Michigan enthusiastically promoted the benefits of individual institutions joining the EEBO-TCP—putting heavy emphasis on the benefits of partnership. Contribution levels were adjusted to allow institutions of different sizes to join the Text Creation Partnership on an equal footing; they could choose to contribute via a single lump sum (of $50,000 on average) or by five equal annual $10,000 payments. In return for their commitment to the work of EEBO-TCP, these institutions would gain immediate access to the textual resources as they were created—via both the commercial EEBO interface offered by ProQuest and also via a more tailored platform, built on the DLXS system developed at Michigan. They would also be contributing to the production of a corpus of essential texts, published in England between 1473 and 1700, which would subsequently be made available for the benefit of all. Despite increasingly constrained budgets, almost 150 US institutions paid to support the work of EEBO-TCP.

In the United Kingdom, we were fortunate to have the Jisc (then known as the Joint Information Systems Committee), a nationally funded service dedicated to acquiring or funding the production of content, tools, and services that would be of widespread benefit to the UK’s scholarly community. The Jisc immediately saw the merits of EEBO-TCP’s innovative approach and committed a single contribution of £1,000,000 on behalf of the UK academic community. This decision also served the Jisc, as they were involved in negotiating the relicensing of ProQuest’s EEBO database to UK universities, and the added value of full-text searching for 25,000 of these important items considerably increased its appeal to library budget holders.

The initial successes in fundraising meant that the production work of EEBO-TCP was able to begin in 2000. A keying specification was developed, and text conversion companies were invited to tender for the work. Titles were selected in light of suggestions from an international editorial board and also colleagues at the growing number of EEBO-TCP partner institutions. The digital images of the texts were sent to the chosen keying companies, and the results were subject to robust quality assurance and markup enhancement by trained teams of digital editors based at Michigan and Oxford. Texts that did not achieve the desired quality threshold were returned for rekeying, whilst those that met the standards were fed into a delivery workflow that ensured their timely appearance in ProQuest’s products, and also the TCP’s own delivery platforms. Everything was going to plan.

But even the best-laid plans need to account for unanticipated issues and obstacles, and this paper will share the lessons learned from this major international collaborative endeavour.

For example, whilst the overall production workflow worked extremely well—thanks to the careful oversight of key individuals at Michigan—there is no doubt that fundraising a work-in-progress raises some additional challenges; yet had we not adopted this approach, it is probably unlikely that we would ever have secured a significant majority of the necessary funding before beginning the work. In fact, having outputs that we could ***show*** to potential partners as the work moved along in many cases helped secure their commitment and enabled them to clearly understand what we were aiming to do. Even so, the hard work of attempting to raise funds to achieve the target of $9,000,000, whilst spending a proportion of that money each month on text production, resulted in the work taking longer than anyone had originally envisaged. We met our production target of 25,000 texts—but it took nearly four years longer than we had planned!

In the course of our work, new questions began to emerge that we had not anticipated at the outset. In 2000, nobody asked us about the employment practices and ethical standards of the keying companies selected to work on EEBO-TCP material. By 2007, some institutions that were thinking about committing to EEBO-TCP, and even end-users of the materials, wanted reassurances that the digital data had been produced in an ethically acceptable environment. Moreover, with the growing awareness of the Google Books Library programme, some users began to question the legitimacy of ProQuest’s five-year exclusivity period to the full-text data, and understandably they wanted clarity on when that five-year embargo would elapse. EEBO-TCP had agreed that ProQuest’s exclusivity period would start from the end of the year in which production was completed (originally anticipated to be 2005), but because production was necessarily extended until 2009, the texts could not be released into the public domain until 2015.

As the work of EEBO-TCP neared its end, other questions were also raised. The first and most rewarding for us as a project, was the request to carry on doing what we were doing: to produce *more* texts. Whilst this was a clear demonstration that many people valued the work of EEBO-TCP, it also raised new questions about retaining and redefining our production methods and workflows, whether we could continue with the same funding model, how to select further texts, and so forth. It was tremendously rewarding to have the Jisc commit an additional £1,000,000 to EEBO-TCP ‘Phase II’ without hesitation—but this was 2008–2009, and we have undoubtedly been directly affected by the consequences of the global recession ever since.

Perhaps some of the biggest questions about the 25,000 texts produced by EEBO-TCP (‘Phase I’, as it is now known) are around what we meant—and what we now understand by—the term ‘public domain’. Back in 1999 we blithely assumed that we would simply release this corpus of material into the intellectual wilds, and that ‘the community’ would assume responsibility for their ongoing maintenance and enhancement. Nowadays, we are constantly asked to consider the sustainability of digital resources—and to define what this might mean, who will do the work, and most importantly, how it will be resourced?

At the time of writing this abstract, the texts from EEBO-TCP have not yet been released. That will not happen until 1 January 2015. Several leading individuals and groups from around the globe have already expressed an interest in working with some or all of the corpus; for example, they have put forward ideas for how the materials can be enhanced with additional markup, or corrections crowdsourced from a community of volunteers, or their contents integrated into scholarly editions. But we do not know which, if any, of these things will happen—or if the texts will be taken up and used in wholly unexpected ways by communities with which we have yet to engage. However, by the time of the DH conference in 2015, we will be in a position to reflect on both the 15-year build-up to the release of one of the most important collections of digital texts yet to be created, and to summarize what has happened in the six months since their release. Will they have been taken up and used in new and exciting ways, been picked up by just a few people, or been resolutely ignored?

Whatever the impact of the release of the 25,000 EEBO-TCP Phase I texts on 1 January 2015, there will certainly be important lessons to be learned by the global digital humanities community.