Renaissance and Reformation
Renaissance et Réforme

Fumerton, Patricia, project dir. English Broadside Ballad Archive

Daniel Russo-Batterham

Volume 47, numéro 1, hiver 2024

URI : https://id.erudit.org/iderudit/1111831ar
DOI : https://doi.org/10.33137/rr.v47i1.43455

Aller au sommaire du numéro

Éditeur(s)

Iter Press

ISSN

0034-429X (imprimé)
2293-7374 (numérique)

Découvrir la revue

Citer ce compte rendu

Russo-Batterham, D. (2024). Compte rendu de [Fumerton, Patricia, project dir. English Broadside Ballad Archive]. *Renaissance and Reformation / Renaissance et Réforme*, *47*(1), 204–211. https://doi.org/10.33137/rr.v47i1.43455

Fumerton, Patricia, project dir.
English Broadside Ballad Archive.
Santa Barbara: University of California, Santa Barbara, 2003. Accessed 30 July 2023.
ebba.english.ucsb.edu.

Introduction

It is a rare privilege to review a web archive that has had so sustained an influence on digital scholarship as the English Broadside Ballad Archive (EBBA). Established by Patricia Fumerton in 2003 at the Early Modern Center at the University of California, Santa Barbara, EBBA provides lessons in how we might address issues of sustainability in the digital humanities as we look to preserve and disseminate digital corpora long enough to develop communities of practice around them. Broadside ballads, the central focus of the collection, were inexpensive songs printed on one side of a single sheet of paper, a format that flourished in England through the seventeenth century. These sheets typically did not include a printed melody but instead would include the names of one or more popular tunes to which they could be sung. At the time of writing, the EBBA contains "9,873 early modern broadside ballads free to the public as text, art, and music," a feat made possible by a series of grants from the National Endowment for the Humanities, as well as from the University of California and several other sources ("Funding"). Crucially, this corpus is made available under a Creative Commons Attribution-NonCommercial 4.0 International License.

Why, then, two decades after its establishment, should we revisit this archive now? Not only has the project continued to evolve over the past years, but so has the digital and data landscape in which it sits, and irrevocably so. Data science has emerged as a major force in the way we live, work, and conduct research, and computational musicology continues to develop in its wake. In this brief review, I thus focus on the texts, images, and music of ballads housed in the EBBA, as well as their digital encodings and associated metadata, rather than the layout of the EBBA website, the user-friendliness of its search features, and so on. While the latter elements are no doubt important—doing these things well can ultimately determine the success of a digital project—they have

received reasonable attention already.[15] Moreover, the raw cultural materials of the EBBA can support research in such a broad range of contexts that it makes sense to consider them somewhat in separation from the platform that holds them, as we explore ways of leveraging the material beyond manual interactions with a web browser. The scale, semantic richness, and diversity of this data means that there is still much to be discovered about broadside ballads as new tools and techniques are applied to the corpus. The EBBA also invites us to think about how we can effectively leverage multimodal datasets. The international significance of EBBA is beyond dispute, but more could be done to open up its "text, art, and music" to ensure it is used to its full potential in the years to come.

The Text Encoding Initiative

Ballads in the EBBA are accompanied by a text transcription, a simple HTML representation of the poetic text. Underlying each such page is an XML file that conforms to the Text Encoding Initiative (TEI), where "TEI tags identify the poetic and bibliographic elements of ballads explicitly as stanzas, lines, refrains, publisher, author, publication date, and so on" ("TEI-XML"). The TEI encodings provided are generated by a program (rather than hand coded) and stick to a simple but consistent formula.

The main part of the printed ballad—the stanzas of text to which tunes were sung—is encoded with a combination of columns ( ), line groups (<lg>), and line elements (<l>), with highlighted elements (<hi>) scattered through lines to mark an emphasis on words that are generally rendered through bold or italics in the HTML presentation (see Fig. 1). Each line is numbered by its overall position in the text rather than relative to the stanza in which it sits. While there is minimal encoding of prosodic structures such as rhyme, or mark-up for people, places, and so on that appear in the text, these TEI files are a valuable resource both for presentation of texts online and computational analyses.

15. See Winkler, review of English Broadside Ballad Archive.

```
<lg>
   <l n="2" rend="left"><hi rend="italic">A</hi> S I went to Walsingham,</l>
   <l n="3" rend="indent">to the shrine with speed,</l>
   <l n="4" rend="left">Met I with a jolly Palmer,</l>
   <l n="5" rend="indent">in a Pilgrims weede.</l>
   <l n="6" rend="left">Now God you save you jolly Palmer.</l>
   <l n="7" rend="left"><hi rend="italic">Fran.</hi> Welcome Lady gay,</l>
   <l n="8" rend="indent">Oft have I sued to thee for love.</l>
   <l n="9" rend="left"><hi rend="italic">B.</hi> Oft have I said you nay.</l>
</lg>
```

Fig. 1. The first stanza of EBBA 20102 as encoded in TEI.

The TEI documents also contain rich metadata, which the creators note may exceed what is readily visible on the site: "They include data that EBBA has archived to maintain more supple search features, such as modernized ballad titles, together with an accurate history of the archive's development, such as EBBA editorial credits" ("TEI-XML"). Beyond these, the <TEIHeader> includes details about the woodcut images that accompany ballads, as well as keywords that describe the ballad texts. Parsing these TEI files programmatically provides useful insights into the EBBA corpus and its encodings. Figure 2 shows, for instance, the most commonly applied keywords in a set of more than 5,000 seventeenth-century ballads. While the EBBA uses several sets of keywords to describe the ballad texts, the most common are those developed by the Early Modern Center itself.

The header metadata also includes creation and revision dates, as well as details about the roles played by the many individual contributors, from "transcribers" and "checkers" through to bibliographers, with many people contributing in multiple capacities. Having multiple people involved with each record has, as these files recount, improved the overall quality and accuracy of the data that underpins the EBBA. While it is possible the TEI encodings are used for "text mining" analyses accompanying each ballad and the "Visualizations" section of the EBBA site (if they don't draw on an SQL database under the site), encouraging broader work with these XML documents could both lead to interesting new research and help to rectify minor data quality issues that have accreted over time, such as inconsistencies in the way contributors are named.

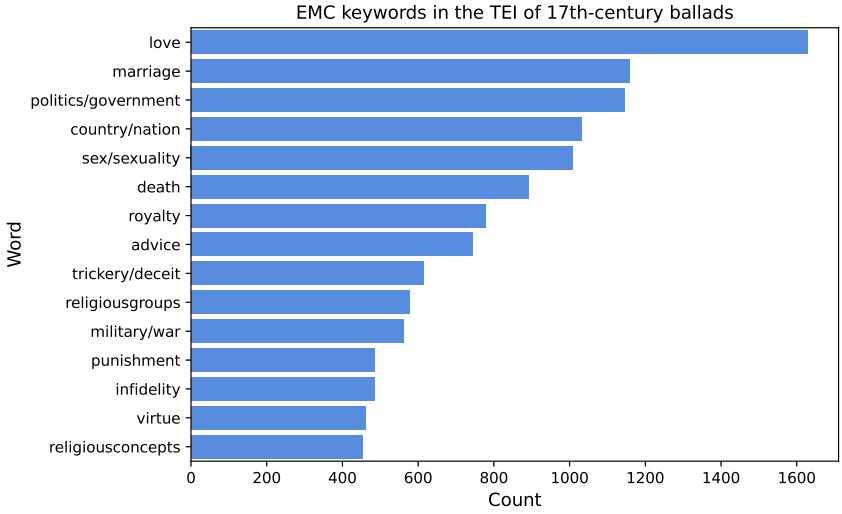

Fig. 2. The most common Early Modern Center keywords as derived from the TEI files of seventeenth-century ballads.

Encoding images: woodcuts and facsimiles

The EBBA features thousands of high-quality images that provide unique ways of reading and interacting with the underlying sources with different levels of granularity and emphasis, from the album books in which these ballads were collected ("Album Facsimiles"), through to individual woodcuts that enhance the aesthetics of each sheet. The album facsimiles—images that show not just individual ballads but also the pages of album books on which they were pasted—give insights into how collectors curated these ballads to form the large collections on which the EBBA is primarily built; the Pepys collection, for instance, held at Magdalene College, Cambridge, contains over 1,800 black-letter ballads and formed the first stage of the archive ("Holdings"). Particularly innovative are the "facsimile transcriptions," where the texts of songs have been painstakingly replaced with transcriptions in a more readable font while keeping the essential layout and character of the originals, including woodcuts. The woodcuts themselves are also stored as separate JPEG images, along with metadata that can be used to search for those with particular subjects or features.

Given the heavy investment in creating these images, any future efforts to enhance how they are viewed, discovered, and analyzed would be well placed. The EBBA platform may benefit, for instance, from integration with the International Image Interoperability Framework (IIIF), which comprises "a set of open standards for delivering high-quality, attributed digital objects online at scale" (iiif.io). Leveraging one of the excellent IIIF-compatible viewers would permit a smoother viewing experience, allowing users to examine the finer details of the facsimiles and woodcuts more easily; these viewers typically include the ability to zoom in without the need to load large, high-resolution images up front, with the IIIF image server providing tiles that fit with the current zoom level and viewport. Additionally, using IIIF standards would allow these facsimiles to be discovered more readily through data portals that bring together materials from multiple libraries and research institutions. A good example is the Biblissima project (biblissima.fr), which allows users to search IIIF "manifests" across more than 30 holding institutions, including the British Library.

Finally, exposing images through an open standard would create new possibilities for analysis of the corpora. IIIF manifests provide not only a mechanism for viewing but are themselves rich in metadata that can be interrogated programmatically. Such a change is non-trivial, however, and the existing JPEG images remain highly usable, especially when coupled with the advanced image search and similarity features provided by integration with Archv software (datalab.ucdavis.edu/archv).

Encoding music: tunes, recordings, and MEI

The EBBA includes recordings of "ballads for which extant tunes are identifiable as sung by members of the EBBA Music Team" ("Recordings"). While necessarily repetitive, these give an indication of how these melodies may have sounded, and also how they might be aligned with texts that often require "compression or elongation of words to fit the chosen melody," necessitating a "process of trial and error" ("Recording the Ballads").

Just as useful for navigating the text–music alignment in these ballads are the music transcriptions that map the first stanza to the tune in modern musical notation. These music transcriptions are currently described as "soon-to-be-launched" ("Musical Transcriptions"). There are nearly 200 (that I could find)

Verovio renderings of tunes accompanied by basic midi playback (verovio.org). While the Music Encoding Initiative (MEI) files that underpin these renderings are *technically* accessible, it seems clear that broader dissemination remains a future step, with the site outlining plans to "expose the MEI encoding of all the musical transcriptions to make evident strong and weak stresses" ("Musical Transcriptions"). Combined with the many other resources already available on EBBA, open access to MEI encodings of the tunes at any scale will help researchers to tackle a broad range of interesting questions about the relationships between these tunes and the texts with which they are associated. Beyond matters of metre and text–music alignment, the MEI files will also help us to examine more rigorously the stylistic features of tunes and answer why some tunes are more suited to broadside ballads than others. Just as EBBA currently integrates tools to cluster similar woodcut images, it would be invaluable to cluster tunes based on sequences of intervals, durations, or other features.[16]

```
<note type="Tune-Total">5</note>
<note type="Tune-1">Walsingham</note>
<note type="Tune_Simpson-1">Walsingham</note>
<note type="Tune_Modern-1">Walsingham</note>
<note type="Tune-2">Jewish dance</note>
<note type="Tune_Modern-2">Jewish Dance</note>
<note type="Tune-3">Bugle Boe</note>
<note type="Tune_Simpson-3">Bugle Bow</note>
<note type="Tune_Modern-3">Bugle Bow</note>
<note type="Tune-4">goe from my window</note>
<note type="Tune_Simpson-4">Go from my Window</note>
<note type="Tune_Modern-4">Go From My Window</note>
<note type="Tune-5">as I went to Walsingham</note>
<note type="Tune_Simpson-5">Walsingham</note>
<note type="Tune_Modern-5">As I Went to Walsingham</note>
```

Fig. 3. Tunes listed in the TEI of EBBA 20102.

16. Although treating a wholly different repertory, work of this kind has been done as part of the Citations: The Renaissance Imitation Mass (CRIM) project (see crimproject.org).

For now, while MEI files remain a scarce resource, the TEI included with ballads is at times useful for identifying recordings with unique characteristics. For instance, the ballad "Frauncis new Iigge, betweene Frauncis a Gentleman, and Richard a Farmer" (whose first stanza is shown in Figure 1) moves through a handful of different tunes across its two parts, all of which are noted in the TEI (see Figure 3) but not in the citation record on the EBBA website. The delightful recording of this ballad, which features an instrumental accompaniment and multiple singers, is almost 20 minutes long.

Training in digital scholarship: past and future

The EBBA has supported an astonishing number of graduate and undergraduate researchers, many of whom acted as research assistants to the project. Over 300 are listed alongside an expansive and expert team with diverse specializations ("The Team"). The highly collaborative nature of the team is evident from the huge body of digitized materials they were collectively able to produce. It is clear from posters produced by undergraduate students that student contributions have focused on key tasks ("Undergraduate Projects"). Many describe working with Adobe Photoshop to produce the different types of facsimiles and woodcut images now available, while others focus more on the transcriptions of the texts. What rings clear across these contributions is a shared fascination with these ballads, their presentation, the surrounding imagery, and the thematic content of their texts.

EBBA's massive contribution to training in digital scholarship is as laudable as it is rare, but it also opens up questions about what kinds of research training EBBA resources will permit in future years. The platform includes a handful of interesting visualizations, from an interactive three-dimensional web tool that visualizes clusters of similar woodcut images, through to mapping of ballad seller locations, topic modelling, and quantitative representations of the prominence of different types of ballads and tunes over time. Beyond these, the "text mining" tab of each ballad entry includes some further topic modelling, frequency counts, and word clouds. These visualizations are important additions, but more could be done to present them in a way that would encourage researchers to attempt similar analyses or critically examine those provided. Although the "Geography of the London Ballad Trade" visualization does link to an "Analysis" page where there is some discussion of its current

limitations—limitations that are inherited from the underlying dataset—there is little mention of how the dataset was compiled, what manipulations were required, what tools were used, and so on.

The efforts of the English Broadside Ballad Archive team in creating rich encodings of thousands of ballads represent an extraordinary accomplishment and a decades-long commitment to excellence. Indeed, there are few digital humanities projects with such enduring success and influence. There remain, however, opportunities to disseminate these materials in ways that better reflect the technical and pedagogical landscape in which we find ourselves today. Further opening the "text, art and music" of the broadside ballad to new modes of analysis while preserving the ease of use and navigation of the current site is a worthy objective. In creating this mammoth collection of images, TEI, and MEI, the EBBA team has made a major contribution to digital scholarship. The creators of these digital encodings, with their intimate knowledge of the collection, are exceptionally well placed to provide guidance on how such encodings might be leveraged for research and digital skills training. One promising avenue to explore, but by no means the only one, would be the inclusion of computational essays in the form of Jupyter notebooks (or similar) to offer a guided journey through the archive.

DANIEL RUSSO-BATTERHAM
University of Melbourne
https://doi.org/10.33137/rr.v47i1.43455

Work Cited

Winkler, Amanda Eubanks. Review of Bodleian Library Broadside Ballads, by Mike Heaney, and English Broadside Ballad Archive, by Patricia Fumerton. *Journal of the American Musicological Society* 67, no. 3 (2014): 848–66. https://doi.org/10.1525/jams.2014.67.3.848.