code{4}lib JOURNAL

Mission   Editorial Committee   Process and Structure   Code4Lib

Visualizing Fedora-managed TEI and MEI documents within Islandora

The Early Modern Songscapes (EMS) project [1] represents a development partnership between the University of Toronto Scarborough's Digital Scholarship Unit (DSU), the University of Maryland, and the University of South Carolina. Developers, librarians and faculty from both institutions have collaborated on an intermedia online platform designed to support the scholarly investigation of early modern English song. The first iteration of the platform, launched at the Early modern Songscapes Conference, held February 8-9, 2019 at the University of Toronto's Centre for Reformation and Renaissance Studies, serves Fedora-held Text Encoding Initiative (TEI) and Music Encoding Initiative (MEI) documents through a JavaScript viewer capable of being embedded within the Islandora digital asset management framework. The viewer presents versions of a song's musical notation and textual underlay followed by the entire song text.

This article reviews the status of this technology, and the process of developing an XML framework for TEI and MEI editions that would serve the requirements of all stakeholder technologies. Beyond the applicability of this technology in other digital scholarship contexts, the approach may serve others seeking methods for integrating technologies into Islandora or working across institutional development environments.

by Raffaele Viglianti, Marcus Emmanuel Barnes, Natkeeran Ledchumykanthan, Kirsta Stapelfeldt

Introduction

Early Modern Songscapes (EMS) is an interdisciplinary web project co-developed by the University of Toronto Scarborough's Digital Scholarship Unit (DSU), the University of Maryland (including the Maryland Institute for technology in the Humanities), and the University of South Carolina. EMS, "aims to develop an intermedia online platform to support scholarly investigations early modern English song," by "tracing individual lyrics and musical settings that moved through different textual and performance contexts in sixteenth- and seventeenth-century England." Ultimately, the group has goals to build a community and resource for researchers and performers interested in sixteenth and seventeenth century English song. The researchers driving the project are Katie Larson, from the UTSC English Department, Scott Trudell from the University of Maryland and Sarah Williams from the University of South Carolina. The first iteration of the project focused on Henry Lawes's 1653 Ayres and Dialogues, highlighting versions of the songs that have been published, co-located with image and sound materials relating to each individual song. [2] Ultimately, Henry Lawes's work provided a reasonable starting point for study by virtue of scope, availability, and significance of his work. However, the hope is that the corpus will grow (and the openness of the project to collaboration is reflected in the absence of a strict scope for content).

At the UTSC Digital Scholarship Unit (DSU), developers and librarians collaborate on the creation of a shared software platforms on which projects like EMS can be developed. Uniquely among the DSU's past projects, EMS incorporated off-site development of a viewer that would display TEI/MEI delivered from the library's Islandora-based infrastructure.

Project Infrastructure

Islandora is an open-source software framework that combines Fedora's repository architecture with the flexibility of Drupal and Apache Solr. Islandora developers provide Drupal modules that, when installed with their dependencies, allow users to manage, create, and access Fedora repository items (digital objects comprised of files, or datastreams) in a Drupal interface. Core indexing

functions are provided through Islandora-specific modules that provide custom integration with Apache Solr. The software is often used to support digital collections and preservation, with "utility" modules providing enrichment functions such as annotation.

As Fedora repository software stores any form of XML inside repository objects, this software has been used in several TEI projects and other applications have been built to integrate the TEI into Islandora (Stapelfeldt & Moses, 2013). The wysiwyg TEI editor "CWRC writer" [3]< has been integrated into Islandora through the The Canadian Writing Research Collaboratory [4] and projects like the "Islandora Critical Edition Solution Pack" [5] through the Editing Modernism in Canada Project [6] incorporate TEI features.

The Early Modern Songscapes (EMS) project made substantial use of existing utility modules in Islandora for storage and display. The Islandora XML Solution pack, which makes possible the storage of XML files, enabled the project to store TEI and MEI assets produced by project researchers. [7] The project also used a module called Islandora Context [8] which itself extends the functionality of the Drupal Context [9] module. The context module is very useful in this instance as it permits for the execution of code given particular user or environmental conditions. In the case of EMS, the "Context" is when an XML object is viewed in the repository, and the executed code is a JavaScript-based "Edition Viewer," authored by Raffaele Viglianti at the Maryland Institute for Technology in the Humanities at the University of Maryland. Digital Scholarship Unit developer Natkeeran Ledchumykanthan the provided an additional EMS integration module (DSU, 2019) [10]. The EMS Integration module includes a context for use with the Islandora Context module, providing a way for the Edition Viewer Javascript bundle to be loaded only when viewing XML objects.

XML Data Structures & Academic Requirements of the project

The Text Encoding Initiative (TEI) [11] is an extensible, XML-based markup language for the representation of texts. The Music Encoding Initiative (MEI) [12] mirrors this work to create markup language for music, but does not have a formal relationship to TEI. These two XML formats both apply descriptive markup principles to text and musical text encoding, which privilege semantic descriptions (i.e. what some text is) over procedural ones (i.e. how some text looks). Both TEI and MEI were required for this project, as the Lawes's book that formed the initial corpus for the EMS project is at the nexus of literature and music and pays careful attention to both the literary and musical aspects of the songs. In the digital editions produced by the project, the TEI is used for verse and the MEI format for notated music. This approach renders both editorial intervention and annotation machine readable. Making this information explicit allows for the preservation of editorial scholarship in the Fedora repository as well as the development of an interactive publication.

Combining TEI and MEI

Despite the lack of formal relationship between the two initiatives, schemas in TEI and MEI can be combined, and both formats provide elements that make the connection meaningful; namely the <notatedMusic> element in TEI and a number of elements in MEI that contain textual data, such as on-staff directives <dir>, poem divisions <verse>, or the more generic <anchoredText>. Combining these two formats is becoming a fairly established practice, see for example the Thesaurus Musicarum Latinarum [13] (Menegozzi 2011: 841-842), but is not without challenges as existing tools and workflows are usually focused on either TEI or MEI. The hierarchical nature of these formats also requires one of the two to contain the other or, in other words, take a primary position. In EMS, the data model priorities TEI, in part because TEI has a well established metadata header in which bibliographical information is stored and this was determined to be desirable. Moreover, the single entry point provided by the TEI element <notatedMusic> elements makes it possible to embed separate MEI documents within the TEI using XInclude, [14] a W3C recommendation that simplifies the inclusion of XML from multiple sources. See figure 1 for a schematic representation of the encoding data model for a song, with TEI including MEI data. The song shown is "When on the Altar of my Hand." Facsimile from Early English Books Online (first published here).

Figure 1. A schematic representation of the encoding data model for the song *When on the Altar of my Hand*, a Facsimile from Early English Books Online, with TEI/MEI data.

The lyrics underlaid under the music notation are encoded both in the MEI as syllables under notes and again as a TEI-encoded verse. This duplication is intentional: subsequent verses, when present, are added after the music notation and follow editorial patterns of early modern printing practice, which is best encoded with TEI. More information is available about the editorial decisions made by the team. [15]

Re-encoding the first verse (typically laid under the notation) allows project editors to represent their interpretation of how it may appear if separately formatted. Because Early Modern songs are often studied as poetry (i.e. removed, for better or worse, from their musical context), this repetition of the first verse may become useful if the text were to be presented without the music.

Collating Versions of Text & Music

A number of songs in the collection have multiple versions coming from different sources, such as manuscripts or other contemporary printed editions, which can be with or without music notation. For these cases, the EMS scholars produced a variorum edition that is a presentation of the text that showcases differences across the sources without privileging one over the

other. Both TEI and MEI are well equipped formats for modeling textual variance, but both assume that one text will be the main reading text (often called the 'base text') and only variant text will be encoded from other sources. [16] To overcome this apparent limitation, the project presents separate combined TEI and MEI documents that represent a collation – a document that lists where the differences between the sources of one song are to be located. In practice, this document encodes a set of pointers grouped by variant; each pointer refers to one or more element in the source XML documents by simple ID references.

```
1   <app>
2     <rdgGrp>
3       <rdg wit="#L638_1">
4         <ptr target="TEI-L638_1.xml#v4"/>
5       </rdg>
6       <rdg wit="#C709">
7         <ptr target="TEI-C709.xml#v4"/>
8       </rdg>
9     </rdgGrp>
10    <rdg wit="#BL_53723">
11      <ptr target="TEI-BL_53723.xml#v4"/>
12    </rdg>
13  </app>
```

Listing 1. A TEI apparatus entry with pointers to three source documents

The example above shows a TEI apparatus entry with pointers to three source documents identified by a witness code (e.g. #L638_1). Each pointer resolves to text that the editors consider being a variant of each other. Note that #L638_1 and #C709 are grouped by the <rdgGrp> element, which indicates that they "agree" or carry the same text (they resolve to the word "deserted"); #BL_53723, on the other hand, carries a different readings of the text (it resolves to the text "forsaken").

With this approach, the editors are able to encode, and preserve in Fedora, each source separately to the degree of detail that we deem appropriate without worrying about tessellating multiple sources in one place. The applicability of this approach goes beyond Early Modern song and has been previously applied to other texts by Viglianti (2016) and is currently being improved and documented thanks to EMS and the ongoing Frankenstein Variorum edition (Beshero-Bondar and Viglianti 2018). EMS however, is the first project to extend this model to music encoding. [17]
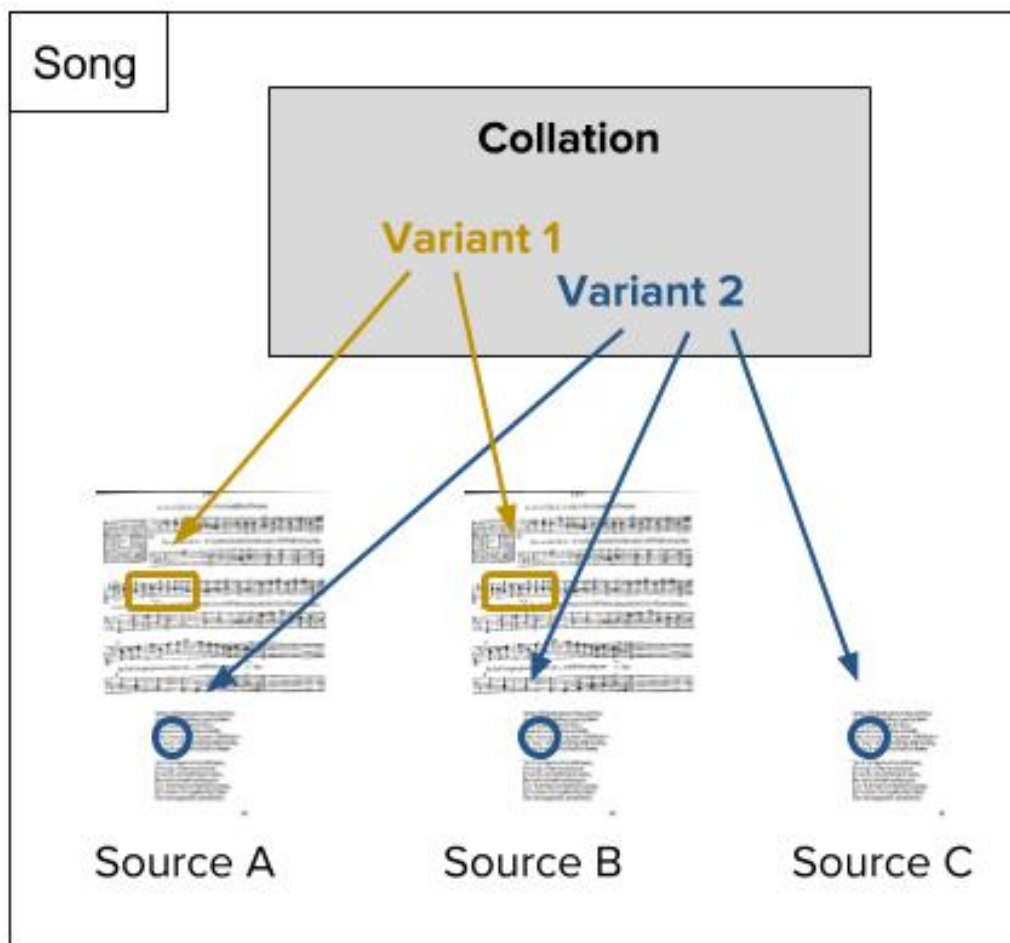
Figure 2. Diagram demonstrating how the collation document points to variants.

Edition viewer

While the XML is stored in the Fedora datastore, the display and interactivity are handled entirely in the front end. A JavaScript React/Redux [18] application loads the collation file for a specific song and determines which other source files to request based on the pointers contained therein. Only one source (such as a specific manuscript or print) is shown at a time, with highlights indicating where the sources diverge. Clicking on these highlights will resolve relevant pointers to the other sources and show the different readings. The 'base' source can be switched at any time. See Figure 3 for an example of the viewer rendering the opening text and musical notation with an expanded musical variant.

Figure 3. A screenshot of the viewer showing a rendering of the opening text and musical notation of "Theseus, O Theseus, hark! but yet in vain," with an expanded musical variant in the bass line of measure 4.

TEI projects typically perform a lossy transformation into HTML for display, but in this case we are using CETEIcean (Cayless and Vigliantti 2018), a JavaScript library that registers TEI element as HTML custom elements, thus requiring minimal or no changes to the TEI source for display in the browser. For rendering MEI we use Verovio (Pugin 2016) a library that renders MEI into SVG for display. Both of these tools offer a solution for rendering these data directly in a user's browser, thus reducing the need for server-side infrastructure for TEI and MEI publications. They also provide isomorphic (that is one-to-one) renderings of the data, which allows to manipulate the rendering as if it were the actual underlying data. Given that the EMS data model strongly relies on data pointers, these isomorphic representations greatly simplify resolving the pointers: their targets will be easily found also in rendering-ready data.

Conclusions / Future directions

The Early Modern Songscapes (EMS) project demonstrates a mechanism for rendering multi-edition TEI and MEI XML data within Islandora. Future work may include overall accessibility of the project site, adding additional songbooks and related resources to the repository, as well as providing a means to annotate the audio and video objects, for example, by using the Islandora Web Annotation module (DSU, 2018).

About the authors

Dr. Raffaele Vigliantti is a Research Programmer at the Maryland Institute for Technology in the Humanities at the University of Maryland. Raffaele's work revolves around digital editions and textual scholarship. He is currently an elected member of the Text Encoding Initiative technical council and an advisor for the Music Encoding Initiative, which produces guidelines for the digital representation of music notation with a focus on scholarly requirements.

Kirsta Stapelfeldt is a librarian and the Coordinator of the Digital Scholarship Unit at the University of Toronto Scarborough Library.

Marcus Emmanuel Barnes is a software developer in the University of Toronto Scarborough Library's Digital Scholarship Unit. Marcus specializes in creating tools that enhance academic libraries for faculty, staff and students.

Natkeeran Ledchumykanthan is a software developer in the University of Toronto Scarborough Library's Digital Scholarship Unit. Nat has more than nine years of full stack development experience (LAMP, PHP, Java, Javascript, XML technologies) and is an active volunteer for community digital preservation/library projects.

References

Beshero-Bondar, Elisa E., and Raffaele Viglianti. "Stand-off Bridges in the Frankenstein Variorum Project: Interchange and Interoperability within TEI Markup Ecosystems." Presented at Balisage: The Markup Conference 2018, Washington, DC, July 31 – August 3, 2018. In Proceedings of Balisage: The Markup Conference 2018. Balisage Series on Markup Technologies, vol. 21 (2018). https://doi.org/10.4242/BalisageVol21.Beshero-Bondar01.

Cayless, Hugh, and Raffaele Viglianti. "CETEIcean: TEI in the Browser." Presented at Balisage: The Markup Conference 2018, Washington, DC, July 31 – August 3, 2018. In Proceedings of Balisage: The Markup Conference 2018. Balisage Series on Markup Technologies, vol. 21 (2018). https://doi.org/10.4242/BalisageVol21.Cayless01.

Kepper, Johannes. "Das FreiDi-Datenmodell" in Freischütz Digital online project. (2014) https://freischuetz-digital.de/datamodel.html.

Pugin, Laurent. "Interaction with Music Encoding" in „Ei, dem alten Herrn zoll' ich Achtung gern'" Festschrift für Joachim Veit zum 60. Geburtstag. Eds. Kristina Richts and Peter Stadler. (2016): 617-631. Allitera Verlag

Mengozzi, Stefano. "Digital and Multimedia Scholarship." Journal of the American Musicological Society 67, no. 3 (2014): 839-47. doi:10.1525/jams.2014.67.3.839.

Raffaele Viglianti. "Music and Words: Reconciling libretto and score editions in the digital medium" in „Ei, dem alten Herrn zoll' ich Achtung gern'" Festschrift für Joachim Veit zum 60. Geburtstag. Eds. Kristina Richts and Peter Stadler. (2016): 728-746. Allitera Verlag

Stapelfeldt, Kirsta, and Donald Moses. 2013. "Islandora and TEI: Current and Emerging Applications/Approaches." Journal of the Text Encoding Initiative 5. http://jtei.revues.org/790. doi:10.4000/jtei.790.

DSU (The Digital Scholarship Unit at the University of Toronto Scarborough Library), March 5, 2019. digitalutsc/ems: Initial tagged release (Version v1.0.0). Zenodo. http://doi.org/10.5281/zenodo.2584156

DSU (The Digital Scholarship Unit at the University of Toronto Scarborough Library), December 2, 2018. digitalutsc/islandora_web_annotations: 7.x-1.12 (Version 7.x-1.12). Zenodo. http://doi.org/10.5281/zenodo.2483252

Notes

[1] Early Modern Songcapes; [cited 2019, Jan 28]. Available from: http://ems.digitalscholarship.utsc.utoronto.ca/

[2] Henry Lawes (1596-1662) is a 17th century composer that the project PIs deemed particularly well suited to explorations of song and text. A more fulsome argument about why Lawes's work was selected as the first contribution to this project is proffered on the project website. http://ems.digitalscholarship.utsc.utoronto.ca/content/why-henry-lawes

[3] https://cwrc-writer.cwrc.ca/

[4] https://cwrc.ca/about

[5] https://github.com/discoverygarden/islandora_critical_edition

[6] http://editingmodernism.ca/

[7] https://github.com/SFULibrary/islandora_solution_pack_xml

[8] https://github.com/SFULibrary/islandora_context

[9] https://www.drupal.org/project/context

[10] Early Modern Songscapes Viewer Integration Islandora Module https://github.com/digitalutsc/ems/releases/tag/v1.0.0

[11] https://tei-c.org/

[12] https://music-encoding.org/

[13] http://www.chmtl.indiana.edu/

[14] https://www.w3.org/TR/xinclude/

[15] http://ems.digitalscholarship.utsc.utoronto.ca/content/editorial-statement

[16] See Chapter 12 of the TEI Guidelines https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html and Chapter 10 of the MEI Guidelines https://music-encoding.org/guidelines/v4/content/critapp.html.

[17] A similar approach has been the subject of experimentation of the Freischütz Digital project, but with substantial differences. In short, the collation document (called core) carries most of the notation data, while sources import common data and provide additional detail (Kepper 2014). While this method may be more economical (reduces duplication of notation common to all sources), it makes the production of data considerably harder to achieve. We argue that our method achieves the same goal more simply.

[18] Respectively https://reactjs.org/ and https://redux.js.org/.

Subscribe to comments: For this article | For all articles