**Origin stories and other bibliographical tales**

**[Opening slide with NEH/Folger logos]** The Folger’s Digital Anthology of Early Modern English Drama explores the complex nature of digital play texts through metadata presentation, documentary editing, and pedagogical workshops. **[slide of splash page]** The aim of the Digital Anthology is to create an online hub for discovering the dramatic publications of early modern playwrights other than Shakespeare. Users will be able to read first editions of the plays online and download them at multiple curation levels to use in research and as the basis for further editorial work. The project scope is limited to plays written for the professional stage between 1576 and 1642, which were printed before 1660 and for which there is an extant early edition. Only 403 dramatic works fit this definition, which excludes a variety of masques, tilts, entertainments, and university plays performed by amateurs, as well as professional drama that never made it into print, or for which a printed edition no longer survives. Although our corpus varies widely from others you’ll hear about in this panel, we’re ultimately based on the same EEBO-TCP materials and we are invested in documenting the creation and potential interoperability of these datasets.

 The Digital Anthology team is working with the Roy Rosenzweig Center for History and New Media at George Mason to create a search and browsing portal **[BB Browsing slide]** through which users can explore, for example, an author’s works, a company’s repertoire, or a publisher’s output. **[WBW slide]** An individual play’s page is broken into sections that describe the play and its earliest printed witness. The information in the black header contains edition-independent information, such as a standardized title, date of first performance and printing, author, and genre. Although each play is represented by one edition for now, we have planned for the possibility of adding subsequent editions in the future.

In the edition-specific portion, our metadata is broken down into categories detailing the play’s publication, performance, bibliographical references, and a contextual history of the electronic development of our file. The publication section focuses on information listed on the title page, such as the date and any theaters or companies in a standardized format, as well as searchable original-spelling transcriptions of the full title, performance statement, and imprint. The physical description contains lists of contextual materials and a browsable format classification. **[Slide of Quartos]** Spoiler: 315 out of the 403 are quartos. **[Back to play slide]** We will be crosslinking works printed in collections so that each play points to the others, acknowledging that readers—as opposed to theatrical audiences—would have encountered Middleton’s *Women Beware Women* alongside his *More Dissemblers Besides Women*. **[slide showing same play in DEEP]** Our bibliography section will eventually link each of our play pages to the Database of Early English Plays to facilitate easy access to DEEP’s data, with its greater documentation of print-history, such as the entries in the Stationers’ Register and its wider corpus, including variants, issues, and subsequent editions.

**[return to play page]** Although our corpus selection is limited to plays performed London’s professional theaters, the performance section of our play page is bare-bones, including only the date and company associated with the first performance, if known. Future phases of the project may add links to documentation of early performances, stage histories, or even video of modern interpretations. (We’d love suggestions).

The user will be able to download the text as transcribed by the Early English Books Online Text creation partnership or as emended and curated by Martin Mueller’s Shakespeare His Contemporaries project, which you will be hearing more about later in this panel. Building on Muller’s work, the Digital Anthology project will also produce 40 documentary editions over the next year and a half, with our own additions and emendations to both the text and TEI-XML encoding.

The play pages represent the textual history of these files through a breadcrumb trail, **[Click circle]** which provides links to the holding repository of the physical playbook and to each digitally-available transcription and encoding. In crafting this documentation we’re attempting to make the digital transmission and remediation process both human-readable and easily accessible, and to encourage users to access the greater bibliographic detail often contained in each library’s online catalogue. Linking to the catalogues emphasizes the easily elided point that EEBO and its derivatives are reproductions of a unique copy of a printed book, allowing us to point users to where they can double check our work themselves. Importantly, it also allowed us to compare repository and microfilm records to the available transcripts, which flagged instances where the works originally imaged by UMI were incomplete, where the reproduction and digitization of the images resulted in conflated copies, or where the transmission process was otherwise problematic.

The EEBO-TCP transcriptions which are the basis for our documentary editions are the result of a series of remediations that deserve critical attention. Many users are blissfully unaware of just how many steps separate the original printed book from the digital objects now available to them for reading and analysis. Significantly, each stage of the remediation process allowed for a new set of assumptions to change the presentation of the work and its description. Our notes field attempts to highlight disagreements in bibliographical sources and discrepancies between files, as well as clarify some of the changes that occurred during the photographic stages of remediation (both physical and digital) that may complicate interpretations of these works.

As a significant contributor to UMI’s imaging project, the Folger has a full set of the microfilms that are the basis for EEBO’s image sets, which have provided essential data for documenting the transmission of these texts from physical playbook to encoded transcription. **[“Tracing a Book” slide w/Love’s Sacrifice montage]** While photographing a book (or, in some cases, a photostat of the original book), the technicians at University Microfilm International would also image a small card with general repository information and (when we’re lucky) a shelfmark. This information was later translated into UMI’s Guides to the microfilm, and either from there or the microfilm itself to EEBO. If the card photographed with the item was not legible, however, no holding information is recorded in later guides. We tracked down Ford’s *Love’s Sacrifice* based on its bookplate, which was imaged in the microfilm but not reproduced electronically.

 In other instances, problems crept in when the holdings information was digitized for EEBO. **[Gamester image]** James Shirley’s *The Gamester* lacks a “Copy from” field, but is both clearly labeled in the first image with a Folger Shakespeare Library ruler and in the UMI reel guide as a Folger-held item. Librarians at the University of Cambridge, University of Illinois Urbana Champaign, Williams College, and University of Texas at Austin all helped confirm data, several with images of title pages or newly created online catalogue entries, while Durham University and Trinity College Dublin staff helped by double checking that items listed in EEBO were **not**, in fact, in their collections.

This sort of metadata raised questions about *The* *Country Captain* when EEBO’s “note” field listed both British Library and the Huntington as the source of photographs. **[Country Captain record]** The record for *The Country Captain* claims “Pages 25–51 have print show-through and are tightly bound with loss of text in filmed copy. Pages 22–51 of Pt. 1 Photographed from British Library copy and inserted at end. Reproduction of original in Huntington Library.” **[EEBO TCP for Country Captain]** A user of the TCP transcription will encounter a note which identifies a gap at the front of 21+ pages, and, further down, notes for 24 pages in the middle of the transcription which were skipped for being “duplicates.” What actually happened: in the microfilm, the Huntington copy is wholly imaged and complete; however, because the copy is tightly bound, there is loss of text in the gutter. UMI decided to supplement the Huntington images with additional images from the British Museum, in the terms of the guide, “splicing” them in at the end. In the digitization process, the images from the British Museum copy were inserted in place of the preliminaries and pages 1-28 of the Huntington copy. So the text, as presented by the TCP, represents a mixed and incomplete transcription of two disparate copies. Importing the metadata description (albeit updated to read “British Library”) without unpicking the works’ transmission history would continue the game of telephone which describes the British Library contribution as appearing “at the end.” In our record, **[slide of DA play page]** we note the origin of what was transcribed and what is missing and point users toward the microfilm for a complete image set; eventually we hope to retranscribe and clarify this incomplete play. In selecting our priority list, we’ll fully admit to avoiding incomplete plays; this sadly eliminated some heavy hitters like Ben Jonson’s *The Alchemist.*

These mixed-copy transcriptions are not likely to cause problems for many of the big-data projects using the public domain EEBO-TCP phase one texts as linguistic corpora. For a project like ours, however, which attempts to create a corpus of reliable documentary editions with transparent editorial and encoding policies, the bibliographical spadework is essential for grounding each text in its material history. In June, the Folger will hold the first of two workshops to see how researchers and educators might continue the transmission, revision, and reinterpretation of these texts in the undergraduate classroom.