

**Kent State University**

---

**From the Selected Works of Virginia A Dressler**

---

2009

# Implementing TEI Projects and Accompanying Metadata for Small Libraries: Rationale and Best Practices

Richard Wisneski, *Case Western Reserve University*

Virginia A Dressler, *Kent State University*



Available at: [https://works.bepress.com/virginia\\_dressler/1/](https://works.bepress.com/virginia_dressler/1/)

# **Implementing TEI Projects and Accompanying Metadata for Small Libraries: Rationale and Best Practices**

RICHARD WISNESKI

*Bibliographic and Metadata Services, Kelvin Smith Library, Case Western Reserve University,  
Cleveland, Ohio, USA*

VIRGINIA DRESSLER

*Digital Library Programs, Kelvin Smith Library, Case Western Reserve University,  
Cleveland, Ohio, USA*

*Electronic text encoding marks up documents, most often those in the humanities and social sciences, in XML to capture various metadata and represent textual features important for research and analysis. It typically follows the Text Encoding Initiative (TEI) scheme. Implementing TEI projects can be particularly challenging for small institutions with limited staff and budgets, given the detail and attention to textual nuances text encoding requires. This article argues for implementing such projects and their feasibility for small institutions, using Case Western Reserve University's text encoding project as a model. It includes the rationale for TEI versus sole reliance on page images or PDFs and for digitization and text mark-up workflow.*

**KEYWORDS** *Case Western Reserve University, digital projects, Kelvin Smith Library, metadata, Text Encoding Initiative (TEI)*

## **INTRODUCTION**

Electronic text encoding marks up documents, most often those in the humanities and social sciences, in an XML format to capture various metadata and represent textual features important for research and analysis. This mark up typically follows the Text Encoding Initiative (TEI) guidelines, maintained by the TEI consortium (TEI: [www.tei-c.org/Guidelines/index.html](http://www.tei-c.org/Guidelines/index.html); Burnard

---

Address correspondence to Richard Wisneski, Kelvin Smith Library, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106, USA. E-mail: [rlw54@case.edu](mailto:rlw54@case.edu)

& Bauman, 2008). Implementing TEI projects of any size involves several processes and levels of expertise. First, the physical texts themselves often are scanned, run through optical character recognition software, and stored following digitization best practice standards. TEI Headers, which contain administrative, descriptive, technical, and preservation metadata must be created. Unless this process is automated, text encoders need to know the MARC record format for some elements' contents. Depending on the level of encoding, text encoders may be assigned to encode a document's structural and textual nuances, which requires careful attention to detail, particularly in capturing such features as lists, tables, images, names, dates, places notes, marginalia, additions, and deletions. The size and nature of a collection can further make such work quite time consuming. It can be considerably less difficult, for instance, to encode a small collection of short, typed verse published in the early twentieth century than to encode a large collection of lengthy books, letters, handwritten manuscripts, or texts written in another script.

A number of large research institutions' digitization programs and units, such as Brown University's Women Writers Project, the Maryland Institute for Technology in the Humanities (MITH), the University of Virginia's Digital Initiatives, Indiana University's Digital Library Program (DLP), the University of Michigan's Digital Library Production Service (DLPS), and the Oxford Digital Library have instituted TEI projects and been very active in such work. They have developed workflows, created tools, and benefited from staff with a great deal of expertise to encode and present their texts in detailed and dynamic ways. Small institutions, on the other hand, are not as well represented in the TEI consortium, with few TEI projects stemming from such institutions.

This article presents the rationale for small institutions undertaking text encoding projects despite small bibliographic and digitization staffs and limited budgets. It argues for doing such work in lieu of solely relying on page images, PDF files, database structures, and outsourcing. It presents the policies and procedures for a TEI project. Specifically, this article presents workflow; TEI headers metadata; structural mark up; creation of metadata records, such as MODS, METS, and Dublin Core; and, techniques to analyze policies and procedures. Throughout, documentation for all stages is emphasized in the text-encoding process.

Catalogers and practicum students in bibliographic and metadata programs can be trained in TEI and metadata creation, in part through demonstrating crosswalks to MARC. Student employees can be trained in scanning, OCR, and making typographical corrections to texts caused by dirty (inaccurate) OCR. Digital library assistants and staff can be trained in structural-level TEI markup and phrase-level encoding.

This article also highlights the grass roots efforts to establish and promote descriptive markup to those outside a library's digitization/

bibliographic and metadata services unit. This effort includes outreach to faculty, graduate students, other small libraries, museums, and historical institutions for perpetuating electronic text encoding and establishing collaborative efforts for future projects.

## BACKGROUND

Case Western Reserve University (CWRU), a private university in northeast Ohio with just under 10,000 students, is a small private institution. Its main library, Kelvin Smith, holds approximately 1.7 million volumes and has approximately 50 full-time staff members. Despite its relatively small size and staff, the library has begun to encode a sizeable collection of first edition historical texts held in its Special Collections department on the regional history of Cleveland, Ohio, and the Western Reserve. The length of books in this collection ranges from under 50 pages to over 300 pages. The first phase of this project, encoding 150 texts, is projected to be completed between 24 to 30 months. Subsequent collections pertaining to Cleveland, Ohio, history contributed by neighboring institutions are expected to be encoded afterwards on a continual basis.

Many of these books, all published before 1923, are of interest to a number of historians and scholars, not just at Case Western Reserve University (CWRU), but to universities, historical centers, and public libraries throughout northeast Ohio and the Midwest in general. Scholars have come to the region specifically to research and study Cleveland and Western Reserve history. The history of Cleveland and the Western Reserve is rich and diverse. Cleveland was the northern terminus of the Ohio and Erie Canal, a major forwarding point for grain from the Ohio heartland. It was the fifth largest city in America in the early 20th century. It was a leading U.S. manufacturing center, second only to Detroit in the automotive industry, and a leader in steel production, shipbuilding, and other industrial sectors. Cleveland had the highest concentrations of some Eastern European heritage populations outside of their home countries, being one of America's leading destinations for immigrants during the 19th and early 20th centuries. Cleveland was, furthermore, one of the leading destinations for African Americans during the Great Migration of the 1910s. It was the founding city and headquarters of some of the country's major corporations, such as Standard Oil. The list of contributions and impacts ranges from a nationally noted tradition for reform to cultural philanthropy, which created institutions such as the Cleveland Orchestra and the Cleveland Museum of Art (Grabowski & Van Tassel, 1998).

Kelvin Smith Library's institutional digital repository, Digital Case, which utilizes the Fedora digital asset management architecture, currently holds PDF versions of the texts along with more than 40 other digital collections. Over the past two years, the collection, Books on Cleveland, Ohio and the Western

Reserve Digital Text Collection, has received more than 11,000 hits, ranking it tenth among the digitized collections.<sup>1</sup> However, this collection in its current state has several shortcomings. Currently, the collection has no interface of its own to allow researchers to search strictly within the collection. Presently, the digital repository allows only for key-word searching in two ways: (1) within a qualified Dublin Core record, where one can search on title, author, subject terms, date, and other fields as supplied only in a Dublin Core record representing the text and (2) within an “OCRred” (Optical Character Recognition) text, which is unstable in that (a) some texts are not OCRred and (b) the OCRred texts have a significant error rate, with some words unaccounted for and others containing “dirty” OCR. Oftentimes, unusual fonts, charts, and diagrams will greatly diminish the accuracy of the OCR software. The texts themselves are dependent on an Adobe® Reader platform to search within the text. Further, some PDF page images are of poor quality, with inconsistent page sizes and readability.

Availability of texts in the collection is further limited. Early into the project, student workers were assigned the task to research the availability of texts in the collection in other formats and versions. Students recorded the number of copies available in OhioLINK, a consortium of 88 Ohio college and university libraries and the State Library of Ohio; World-CAT; Google Books; the Internet Archive (<http://www.archive.org>); Ohio Digital Resource Commons (DRC) (<http://info.drc.ohiolink.edu>); the Cleveland Memory Project (<http://www.clevelandmemory.org/>); the Ohio Memory Project (<http://cdm15005.contentdm.oclc.org>); the Cleveland Digital Library (<http://web.ulib.csuohio.edu/SpecColl/cdl>); and Ohio’s Heritage Northeast (<http://www.ohiosheritageneast.org>). No text-encoded versions of books are available through the Cleveland Memory Project, the Ohio’s Heritage Northeast digital collection (both of which largely contain audio and visual material), the Ohio Memory Project, and OhioLINK’s Digital Resource Commons. Only 30 titles are available through Google Books, where search capabilities are largely confined to keyword searching (see sample, Appendix A).

Given these shortcomings, and given the interest in the subject matter, librarians at Kelvin Smith Library, including those in information technology, digital library programs, and bibliographic and metadata services, explored various options for preserving and presenting this collection to the public. They consulted experts in text encoding, contacted other institutions regarding their digitized texts presentations, conducted an extensive literature review, and interviewed historians and scholars in the collection’s subject. The final recommendation was to build an interface devoted solely to the collection, and carry out TEI encoding on the texts at P5, level 3, of the TEI Consortium’s Best Practices.

One could argue that a project of this size and scope would be more suitable for reliance strictly on PDFs or page images. We found during our

review that the PDFs were sometimes of poor quality and that many had not been run through the OCR software, which limits the searching to the basic catalog fields such as author, title, and date. In scanning some texts in house, no consistent effort was made to record administrative metadata to know when the scanning occurred; inconsistent technical metadata was created to record the resolution, format, and platform for individual page images; and no descriptive metadata was recorded to provide detail on the pages scanned and the file names.

Efforts to retrace and review such images from the texts were difficult for these reasons, but a necessary process to accomplish at least in part during the first stages of the project. Furthermore, the file storage of individual digital page images was itself less than adequate. Most of the TIFFs were stored on nonarchival-quality CDs, many of which have deteriorated in a matter of a few years. Other titles were irrevocably lost due to poor documentation and a lack of any backup procedure. Some scans were too poor in quality to retain, reflecting occasional lapses in following a set of best practices for scanning and storage. Part of this project now includes a detailed workflow to record scanning, metadata, and markup, thus, preserving the material and details on digital creation and output and preventing further loss of data by adhering to set practices and guidelines that are now in place.

The only true consistency found within previously digitized titles was found primarily in materials that were outsourced for scanning. Through the order forms, librarians could at least ascertain the date digitized and the quality of the original scans (such as the dpi and the bit depth). These companies provided archival-quality CDs and offered backup copies upon request years after the projects were completed. A simple MS Excel file had been used as the primary record of the whole digital books project (Cleveland History titles as well as others), however, was found to be incomplete and inconsistent. A thorough review of hundreds of CDs containing the titles uncovered many of the mistakes in the main Excel file, and revealed the deteriorated digital files and additional titles for which there had been previously no account. The review of CDs included verifying the quality of the digital files, which displayed a lack of standards in the previous efforts in the range of digital capture (dpi, bit depth) and also in the file naming.<sup>2</sup> After the painstaking review process was completed, we were able to create a simple file-sharing program to begin work translating these titles into the text-encoding process. In some cases, this meant that we had to re-scan certain titles, while other titles were ready for further image processing required for a book viewer, used to display TEI and page images.

It was further determined that text encoding was an important part of the process in the long-term preservation, enhanced access, and dissemination efforts, rather than strictly redoing scans, improving upon file storage, and building a database. Databases, as pointed out by Julia Flanders, director for Brown University's Women Writers, emphasize structural repeatability

and predictability, and texts in our collection have meaningful variable structures. XML encoding, by contrast, represents information predictable only in complex ways. Furthermore, XML databases allow for fast searching on nuanced data.<sup>3</sup>

TEI mark up was deemed appropriate given the nature of the texts themselves and a consensus on what to encode.<sup>4</sup> All texts were published between the mid-19th and early-20th centuries, typed, and structured in such a way as to be conducive for structural text encoding. TEI in itself was deemed to be best for preservation and retrieval, an improvement over PDFs and TIFFs, which can be prone to irrevocable loss and deterioration. The mark up is software- and hardware-independent, enhances the ability to describe textual features, and is well-supported in the TEI Consortium.<sup>5</sup> The mark up would provide a way for these texts to become “mobile” and illuminated through advanced search features.<sup>6</sup>

Lastly, Kelvin Smith Library decided to purchase a “Book Viewer” for its institutional repository. This viewer is capable of displaying page images alongside the encoded text. It also simulates a book view to mimic the physical book, which can be found at <http://hdl.handle.net/2186/ksl:bookviewer>.<sup>7</sup>

## WORKFLOW

Once the rationale for doing the project was concluded, the next issue was to establish a workflow, including policies and procedures for digitization, file naming, and TEI markup. A project manager was assigned, whose task was to oversee the project, put together a project team, contribute to text mark up, and establish policies, procedures, guidelines, and training. Given previous attempts at digitization, the project manager asked the Digital Library Programs librarian to review prior digitization work and revise the scanning and digitization workflow. A set of guidelines for the digitization of materials was created as a result of the previous inconsistent efforts. This set of internal guidelines gave benchmark recommendations for scanning text and multimedia objects. Texts are presently scanned at 24- or 8-bit,<sup>8</sup> 300 dpi and dark archived in Kelvin Smith Library’s digital repository, *Digital Case*, depending on original content; color illustrations or book covers, for example, are scanned at 24-bit. Practices follow the Library of Congress National Digital Library Program (NDLP) and National Archives and Research Administration (NARA) Technical Guidelines. Filenaming procedures were adopted from Northwestern Digital Library (<http://staffweb.library.northwestern.edu/dl/documents/DLC-AFF-finalreport-july03.pdf>) and the University of Maryland’s guidelines (<http://www.education.umd.edu/ETS/web/webNamingConventionRP.html>).

The level of textual mark up was the next concern. Currently, the TEI consortium encodes at P5, which has several differences and advantages over P4, including

- all pointing and linking mechanisms now use the same W3C-defined mechanism; there is no longer any distinction between internal and external pointing elements
- improvements to <sic> and <corr>
- improvements to handling language and character sets
- more rigorous validation
- variant readings of different text sources can be tagged, for example:  
`<l>So hath myn <app> <lem wit="#msB #msC">herte</lem> <rdg wit="#msA">hert</rdg></app>`
- <graphic> tag can contain a URL attribute

The TEI Consortium's Best Practices Group has been establishing encoding practices within P5, covering five levels, with level 1 being the most basic (fully automated text creation) and level 5, the most advanced (requiring expert content knowledge and analysis). Briefly put, the purposes of P5, levels 1 and 2, are largely to create electronic texts that allow for key word searching and linking to page images; the text is largely subordinate to the page images. Texts encoding in P5, levels 4 and 5, involve structural mark up and attention to a text's specific textual nuances; texts are meant to stand independent of page images and allow for sophisticated search and retrieval.<sup>9</sup>

Level 3, "Simple Analysis," was deemed the most appropriate for this project. We wanted the encoding to be page-image independent, so that scholars could download and add to the textual mark up as they wished. We wanted to create a search interface that could search on TEI headers and a text's logical structure. Following the rationale for Level 3, according to *TEI in Libraries: Guidelines for Best Practice*, we knew our collection's users wanted key word searching as well as searching on the text's structure. Yet, we didn't want to invest in Level 4, "Basic Content Analysis" encoding, given the additional time and expertise required. To mark up texts in our collection, with attention to all references to dates, places, and names, would have been an overwhelming and very time-consuming task, for example, that we determined would not be efficient, especially given the length and depth of many texts in the collection. On the other hand, encoding in Level 3 encodes hierarchy and topography. Specifically, as stated in the *Guidelines for Best Practice*, Level 3 is recommended for projects that meet the following criteria:

- The material is of interest to a large community of users who wish to read texts that allow key word searching.
- Some sophistication of display, delivery, and searching based on structure of the text is desired.
- Each text will be checked to ensure that encoding decisions have been made appropriately.
- The users of the texts may have limited storage or display capabilities.



- The creator of the texts has limited or no ability to provide content expertise to analyze, tag, or review texts.
- Extensibility is desired, that is, one desires to keep open the option for a higher level of encoding to be added at a later date ([http://wiki.tei-c.org/index.php/TEI\\_in\\_Libraries:\\_Guidelines\\_for\\_Best\\_Practices#LEVEL\\_3:\\_Simple\\_Analysis](http://wiki.tei-c.org/index.php/TEI_in_Libraries:_Guidelines_for_Best_Practices#LEVEL_3:_Simple_Analysis)).

To make the workflow as manageable as possible, the project manager, with input from the library's assistant director, the head of Digital Library Programs, and the Digital Library Programs Librarian, decided to divide the work into several stages (see Appendix B). Student workers run the original TIFFs through OCR software and save the output as a text file, then convert them to TEI P5 using Open Office (an open-source software product) and a P5 XML open-source extension.<sup>10</sup> Student workers also clean up the files, including fixing typographical errors due to dirty OCR, deleting garbage characters (most often from images and illustrations), removing hyphens from words, and marking areas where images and extraneous textual nuances occur.<sup>11</sup> To expedite this work, macros were created for students to easily insert page break tags and tags for images. Hence:

```
<pb facs=".jp2" n="" />
<figure place="">
  <graphic url="" />
  <figDesc> </figDesc>
</figure>12
```

Students, and all involved in the project, use oXygen software, a relatively low-cost XML editor that is largely regarded as the de facto XML editor industry standard.

Kelvin Smith Library cataloging staff simultaneously create TEI headers for documents.<sup>13</sup> Once students complete clean-up work and catalogers complete the headers, both are sent to the project manager, who does quality control, which includes validating XML documents and checking the XML file against the digitized texts. The project manager then combines the TEI headers with their respective edited XML file, and sends the files to text encoders, who do structural mark up at P5, Level 3, according to TEI Consortium Best Practices standards. Particularly, encoders examine the texts first, create an outline for structural mark up, and then encode the texts. All graphics, tables, lists, indices, and the table of contents are encoded as well.

## TEI HEADERS

TEI headers do not necessarily replicate a MARC record. Indeed, TEI headers serve many functions, including recording administrative, structural, and

descriptive metadata. Nonetheless, professional catalogers are enlisted for creating this part of TEI documents given their prior experience in creating MARC records for the texts and given their knowledge of metadata. The project manager created a template for the catalogers to use and provided references the TEI Consortium BPG's work on TEI headers (Appendix C).<sup>14</sup> Catalogers are able to create headers for all texts in a relatively short amount of time, in addition to performing their other work obligations (see Appendix D for a sample TEI header).

## STRUCTURAL MARK UP

The project manager and two other staff members execute the structural text mark up. To make this work as efficient and manageable as possible, the project manager created a worksheet for encoders to fill out first (Appendix E). The idea here was akin to what writers are often taught when writing a thesis: that is, before writing the paper itself, it is often helpful and necessary to spend time outlining and taking notes first. In text encoding, encoders spend time going through texts and marking up their structural components. They take note of all instances where such structural features as tables, images, notes, and lists occur and then mark the pages on which they occur.

This preliminary work, arguably, saves considerable time on the actual text mark up. After practice, some preliminary work, such as marking up all occurrences of tables and lists, can be bypassed, but the preliminary structural creation remains so as to see the larger picture inherent within each text. In the early stages of our project, this worksheet provides a good way to visualize the elements particular to each text and ensure that all the elements are completed in the mark up.

## IMPLEMENTATION OF MODS, METS, AND DUBLIN CORE

After texts are fully encoded, they are sent to the project manager for another quality control check. The manager then has the professional catalogers create Metadata Object Description Schema (MODS), Metadata Encoding and Transmission Standard metadata (METS), and Dublin Core records for each text, making use of MarcEdit, a free Windows-based MARC editing tool. These metadata records contain administrative, structural, and descriptive elements. Simple, unqualified Dublin Core records are the public display in the digital repository. MODS metadata are incorporated into a METS record and serve as a schema for bibliographic content, partly following MARC structures. The METS records were considered particularly important, in part

as a means to have an archival record of all images and documents related to each text and of the structural relation for all parts. METS records addressed the problem with previous work on the project, namely, the lack of a record to trace the location, format, and creation of page images, metadata, and PDFs.

Professional catalogers and other participants in the project were trained in METS and MODS through in-house workshops over the course of two days. Currently, they are creating METS and MODS records for those texts numbering fewer than 100 pages. The IT department is also working on a means to partly automate METS record creation for longer texts.

## ANALYZING POLICIES AND PROCEDURES

Although a great amount of planning and preparation went into the entire project prior to its implementation, the project manager and his team knew the importance of building a review mechanism into the workflow to address any issues or problems that could occur. Virtually no game plan is flawless, and such review is necessary to address moments when difficulties arise or when things do not go according to plan.<sup>15</sup> Initially, student workers were asked to insert tags for tables and lists using macros created in MS Word. At first, this approach seemed simple enough, but students found this work time consuming, as it required careful attention to the source texts and working with formatting that, with some texts, was a bit difficult. Catalogers had to re-revise the TEI headers they created to address some revisions the TEI Consortium BPG made to them. Text encoders were also asked to mark references to dates and names within texts. Some texts, however, have multiple names and dates, making such work too time consuming to pursue. And, periodic meetings with IT personnel were held to discuss how best to make use of the TEI markup and metadata records. A software package, called "TEI Book Viewer," was purchased to display TEI alongside page images and to allow for downloading TEI and PDF versions. At present, the technology staff at the main library is investigating enhancements to this viewer. Specifically, there are other options for institutions to display encoded texts and page images rather than such viewers. Many institutions, for example, make use of XSLT for display and retrieval, such as the University of Miami Ohio's "Poetess Archive" (<http://unixgen.muohio.edu/~poetess/>).

Bimonthly meetings are now in place with project team personnel, historians, and the library assistant director to periodically review progress and procedures. It is anticipated that further tweaks to the workflow and to policies will occur; the key ingredient is active communication with all interested parties.

## CONCLUSION

As stated at the outset, Kelvin Smith Library is relatively small, with a limited budget, medium-level scanning equipment, and a relatively small staff. Nonetheless, the library undertook this project and continues to make steady progress on it. As additional titles are incorporated into the Books on Cleveland, Ohio and the Western Reserve Digital Text Collection, these titles would be put into consideration for text encoding as well.

Consistent, regular communication during the early stages of the project was essential, and modifications in the workflow and procedures reflected the adjustments necessary for a logical production schedule and also consideration of problems as they arose (such as titles which had to be re-scanned). A workflow was devised in a way that would allow each text to be processed in stages, first, assessing the page scans; then, running the text through OCR software and performing minimal image processing; and, last, preparing the text for encoding. This method allowed library staff to divide the work into these main tasks, which has aided the staff in further advancing the project while also maintaining an online shared workflow document to record completed and in-progress work at each stage of the process.

Additional assistance was found in enlisting practicum students at a nearby library science program. Given the ever-increasing challenges that new library and information science graduates face in obtaining gainful employment at the collegiate level, practicum students have been grateful for the opportunity to work on the project, and have made significant contributions. Their assistance has been invaluable.

Kelvin Smith Library is also seeking a NEH Preservation and Access Grant to, among other things, hire a programmer, purchase better and faster scanning equipment, and hire more text encoders. Applying for this grant has led to collaboration with librarians, faculty, and students at nearby universities and public libraries. The long-range goal is to train librarians at these institutions in TEI encoding and to work with other faculty and students in adding their institutions' relevant holdings to the collection. Our belief is that developing this collaborative environment will make text-encoding work more approachable and less daunting. Moreover, it will lead to a more robust and useful end product for researchers and scholars in the subject.

Future considerations for this endeavor will be further investigation into transferring content into different applications (such as other book readers or other software capable of reading TEI markup), data migration for the purpose of database upgrade, consistent review of other TEI applications and projects, and following the rules and regulations of the TEI consortium.

With proper planning, training, consultations, and collaborative efforts, TEI work is attainable for even small libraries. The short-term and long-term benefits, in our opinion, far outweigh the initial reservations about

undertaking such work. As a result of structured text encoding, it is our belief that enhanced access of the collection will result, as well as a different kind of reading experience with the newly implemented Book Viewer. The TEI mark up will also provide a stable and long-term solution to preserving texts, given that it is XML based and extensible into other environments.

## NOTES

1. Digital Case also holds a copy of the *Encyclopedia of Cleveland History*, an online encyclopedia encoded in HTML and sponsored by the university's history department. Since 2006, the *Encyclopedia* has received 162,531 hits on Digital Case, ranking it first among the 36 digitized collections.

2. Internal guidelines were set using (in part) Library of Congress National Digital Library Program guidelines (<http://memory.loc.gov/ammem/dli2/html/lcndlp.html>) and mentioned in more detail later in the article. File-naming procedures were adopted from Northwestern Digital Library and the University of Maryland. Links to Northwestern's procedures (<http://staffweb.library.northwestern.edu/dl/documents/DLC-AFF-finalreport-july03.pdf>) and Maryland's guideline (<http://www.education.umd.edu/ETS/web/webNamingConventionRP.html>). ISO 9660 are used for CD-ROM

3. See Julia Flanders, *What is text encoding?* (2007). (<http://www.wwp.brown.edu/encoding/guide/html/encoding.html>). Article accessed 10 June 2008.

4. Both points—understanding your texts and believing in the integrity and utility of selective transcriptions—are discussed in John Lavagnino's 2006 article "Electronic Textual Editing: When Not to Use TEI."

5. For further discussion of the history of and benefits to TEI, see Cummings (2007), The Text Encoding Initiative and the Study of Literature; (Unsworth, 2002) *Electronic Textual Editing and the TEI*; Renear (2004). Text Encoding; Smith (2004), Electronic Scholarly Editing; and Bauman and Burnard (2008). TEI Guidelines P5.

6. The dynamic qualities of text and mobility are discussed in Buzzetti and McGann (2006) Critical Editing in a Digital Horizon.

7. It bears mention that there are other options for institutions to display encoded texts and page images rather than such viewers. Many institutions, for example, make use of XSLT for display and retrieval, such as the University of Miami Ohio's "Poetess Archive." Kelvin Smith Library plans to make use of such transformations later in the project's development.

8. Depending on original content, color illustrations or book covers, for example, would be scanned at 24-bit.

9. For more detailed information, consult the Consortium's Best Practices Guidelines (2009) ([http://wiki.tei-c.org/index.php/TEI\\_in\\_Libraries:\\_Guidelines\\_for\\_Best\\_Practices](http://wiki.tei-c.org/index.php/TEI_in_Libraries:_Guidelines_for_Best_Practices)). Accessed 15 March 2009).

10. See *TEI Open Office Package* ([http://wiki.tei-c.org/index.php/TEI\\_OpenOffice\\_Package](http://wiki.tei-c.org/index.php/TEI_OpenOffice_Package)). Accessed 19 January 2008.

11. Clean-up work does not involve correcting misspellings that are in the original text, only those that result from dirty OCR. Hyphens are generally removed except in instances in which they are a notable feature of the text. Removal of hyphens that separate words at the ends of lines or pages is not uncommon in TEI markup. See Burnard and Bauman (Eds.), 2008. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Oxford: TEI Consortium (p. 35, 58).

12. This way of marking images within pages of text follows P5, Level 4, conventions more so than P5, Level 3. Our rationale is to provide for the needs of print-handicapped readers. See Burnard and Bauman (2008) *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, 14.3, (pp. 443–446).

13. At present, the head of Bibliographic and Metadata Services, who heads the project, is investigating automating this procedure.

14. Also see the TEI Consortium BPG link, especially regarding headers ([http://wiki.tei-c.org/index.php/TEI\\_in\\_Libraries:\\_Guidelines\\_for\\_Best\\_Practices](http://wiki.tei-c.org/index.php/TEI_in_Libraries:_Guidelines_for_Best_Practices)). Accessed 15 July 2009.

15. See Laurie Lopatin's survey of literature on the topic (2006). Also, see Bailey-Hainer and Urban (2004).

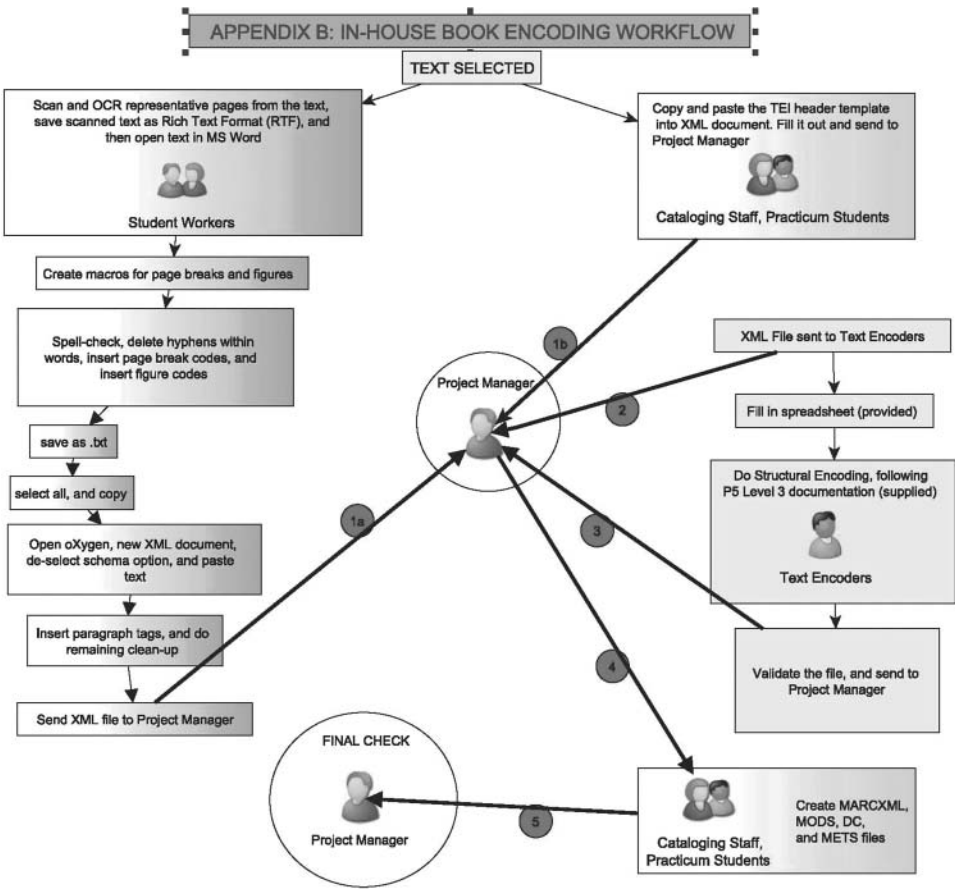
## REFERENCES

- Bailey-Hainer, B., & Urban, R. (2004). The Colorado Digitization Program: A collaboration success story. *Library Hi Tech*, 22(3), 254–262.
- Burnard, L., & Bauman, S. (2008). *TEI P5: Guidelines for electronic text encoding and interchange*. Oxford: TEI Consortium.
- Buzzetti, D., & McGann, J. (2006). Critical editing in a digital horizon. In L. Burnard, K. O. O'Keefe, & J. Unsworth (Eds.), *Electronic textual editing* (pp. 53–73). New York: MLA.
- Cummings, J. (2007). The Text Encoding Initiative and the study of literature. In R. Siemens & S. Schreibman (Eds.), *A companion to digital literary studies* (pp. 451–477). Malden, MA: Blackwell Pub.
- Flanders, J. (2007). *What is text encoding*. Retrieved June 10, 2008, from <http://www.brown.edu/encoding/guide/html/encoding.html>
- Grabowski, J., & Van Tassel, D. (Eds.) (1998). *The Encyclopedia of Cleveland History*. Retrieved from Case Western Reserve University and Western Reserve Historical Society website: <http://ech.cwru.edu/>
- Lavagnino, J. (2006). Electronic textual editing: When not to use TEI. In L. Burnard, K. O. O'Keefe, & J. Unsworth (Eds.), *Electronic textual editing* (pp. 334–339). New York: MLA.
- Lopatin, L. (2006). Library digitization projects, issues and guidelines: A survey of the literature. *Library Hi Tech*, 24(2), 273–289.
- Renear, A. H. (2004). Text encoding. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A companion to digital humanities* (pp. 218–240). Malden, MA: Blackwell Pub.
- Smith, M. N. (2004). Electronic scholarly editing. In R. S. S. Schreibman (Ed.), *Companion to digital humanities* (pp. 306–323). Malden, MA: Blackwell Pub.
- TEI Consortium (2009, March). *TEI in Libraries: Guide to best practices*. Retrieved March 15, 2009, from [http://wiki.tei-c.org/index.php/TEI\\_in\\_Libraries:\\_Guidelines\\_for\\_Best\\_Practices](http://wiki.tei-c.org/index.php/TEI_in_Libraries:_Guidelines_for_Best_Practices)
- Unsworth, J. (2002). *Electronic textual editing and the TEI*. Retrieved June 2, 2007, from Annual Convention of the Modern Language Association: <http://www3.isrl.uiuc.edu/~unsworth/mla-cse.2002.html>

**APPENDIX A** Sample of text availability

Title	No. of Ohio Link copies	No. of World Cat copies	In Google Books	In Internet Archive	Cleveland Memory Project E-Books: <a href="http://www.clevelandmemory.org/ebooks/index.html">http:// www.clevelandmemory. org/ebooks/index.html</a>
A Chapter in the History of Cleveland	4	35	YES	YES	NO
A Comic History of Cleveland, 1796–1901	7	18	YES	YES	NO
A directory of the cities of Cleveland & Ohio, for the years 1837–38	7	12	YES	YES	NO
A few Civic Problems of greater Cleveland	2	17	NO	NO	NO
A historical report of the sixteen years work at Hiram House	2	7	NO	NO	NO
A history of Catholicity in northern Ohio and the diocese of Cleveland	1	50	YES	YES	NO
A history of Cleveland and its environs: the heart of new Connecticut, vols. 1–3	20	91	YES	YES	NO
A history of Cuyahoga County and the City of Cleveland, vols. 1–3	2	46	NO	NO	NO
Bridges of Cleveland, Cuy Valley	3	18	YES	YES	NO
Brief historical sketch of the Cuyahoga County soldiers' and sailors' monument	2	13	NO	YES	NO
Cleveland schools in the nineteenth century	14	42	YES	YES	NO
Cleveland, a prediction	3	5	NO	NO	NO
History of Cleveland Presbyterianism with directory of all the churches	1	58	NO	NO	NO
History of the charities of Cleveland, 1796–1896;	2	7	NO	NO	NO
One hundred years of Western Reserve	1	34	NO	NO	NO

APPENDIX B In-house encoding workflow





APPENDIX C TEI header

TEI HEADER	DESCRIPTION	MARC EQUIVALENT AND NOTES
<teiHeader xml:lang="en"> <fileDesc>	header following AACR2 rules contains a full bibliographic description of the TEI document	Parent element for <sourceDesc>, describing source document from which TEI doc was created
<titleStmt>	ELECTRONIC VERSION INFORMATION groups information about the electronic version of the title of a work and those responsible for its intellectual content.	1XX 1st author. PARENT ELEMENT FOR ALL CHILD ELEMENTS IN THIS SECTION
<title type="main"> or type="marc245a" or "marc245b" etc.	main title entry. Only uniform and main title should be entered. See <sourceDesc> for other forms of title for documents where a user might seek the documents under titles other than those assigned. Where a title is provided by the header creator rather than the document creator, the title should be enclosed in square brackets using standard English language conventions for editorial insertion. subheading [if applicable]	130, 240, 245, 246's. Attribute level=indicates if the text is a book, series, journal, article, or unpublished material
<title type="sub"> or "alt" etc. <author> <persName> or <orgName>	in a bibliographic reference, contains the name of the author(s), personal or corporate, of a work; the primary statement of responsibility for any bib item. Author of original source (electronic or print) should be entered into the <author> tag before the <respStmt>. Use discrete tags within <author> tag for "last name", "first name", "middle name", "date", "position title" to allow future flexibility in display, indexing, and in transferring to MARC. Whenever possible, establish or use nationally established forms of names. The name should be inverted and entered in the established form. editor of original source (electronic or print) should be entered into the <editor> tag BEFORE the <respStmt>	1XX 1st author, 534  a 1st author, 7XX 2nd and 3rd authors
<editor>		7XX all editors

(Continued on next page)

APPENDIX C TEI header (Continued)

TEI HEADER	DESCRIPTION	MARC EQUIVALENT AND NOTES
<respStmt>	supplies a statement of responsibility for the <u>electronic version</u> of a text, edition, recording, or series, where the specialized elements for authors, editors, etc., do not suffice or do not apply. The name should be inverted and entered in the established form.	
<resp>editor, etc.	The editor (also compiler, illustrator, etc.) of an <u>electronic version</u> should be entered into the appropriate tag in the <respStmt>. The name should be inverted and entered in the established form.	500, 7XX  e. Child element for <respStmt>
<name>	contains a proper noun or noun phrase for the institution responsible for the electronic version	Child element for <respStmt>
<extent>	describes the approximate size of the electronic text as stored on some carrier medium, specified in any convenient units. Best practice: use kb	
<publicationStmt>	PUBLICATION STATEMENT FOR ELECTRONIC DISTRIBUTOR groups information concerning the publication or distribution of the ELECTRONIC version of the text	PARENT ELEMENT FOR ALL CHILD ELEMENTS IN THIS SECTION
<distributor>	The party from whom copies of the file (the TEI document, not the source document) can be obtained. Often the same as <publisher>, in which case no <distributor> element should be specified.	
<publisher>	Whoever has collected the TEI document, not the source document, and has made decisions concerning it	
<pubPlace>	contains the city and state of institution responsible for the ELECTRONIC version	
<date when="20xx">	contains a date in any format, with normalized value in the value attribute, for publication of the ELECTRONIC document. For most purposes, the year date (yyyy) will be adequate. If greater detail is required, enter dates as yyyymmdd.	
<idno> <idno>	Any unique identification number for the TEI document.	099
<seriesStmt>	Whenever possible, establish the national authority file authorized form for the electronic locally created series	440
<notesStmt> <note>	Optional. Should be used for indicating questionable attributions for title, author, etc	500
<availability> <p>	supplies information about the availability of a text, for example any restrictions on its use or distribution, its copyright status, etc.	

BIBLIOGRAPHIC INFORMATION FOR TEXT ITSELF		PARENT ELEMENT FOR ALL CHILD ELEMENTS IN THIS SECTION
<sourceDesc>	supplies a bibliographic description of the copy text(s) from which an electronic text was derived or generated.	246, 700 X2  a  t, 730, 740, 534  t
<biblStruct>	contains a fully-structured bibliographic citation, in which all components of the TEI file description are present. It is suggested that child elements appear in the following order for ease of display according to ISBD: <author>, <title type="245a">, <title type="245b">, <respStmt>, <edition>, <pubPlace>, <publisher>, <date>, <extent>, <title type="series">, <note>, <idno type="isbn-13">, <idno type="isbn-10">, <idno type="OCLC-____">. Use <biblFull> for exhaustive description created by hand.	
<monogr>	to group together the elements describing the whole source document, even if the whole source document is not a "monograph", per the TEI definition of this element	
<title type="main"> or type="marc245a" or "marc245b" etc.	main title entry. Only uniform and main title should be entered. See <sourceDesc> for other forms of title for documents where a user might seek the documents under titles other than those assigned. Where a title is provided by the header creator rather than the document creator, the title should be enclosed in square brackets using standard English language conventions for editorial insertion.	130, 240, 245, 246's
<title type="sub"> <author><persName> or <orgName>	subheading [if applicable] in a bibliographic reference, contains the name of the author(s), personal or corporate, of a work; the primary statement of responsibility for any bib item. When possible, use name authority file as found in original source	1XX 1st author, 534  a 1st author, 7XX 2nd and 3rd authors
<editionStmt> <respStmt>	supplies a statement of responsibility for someone responsible for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc., do not suffice or do not apply.	534  b 245  c
<name>	contains a proper noun or noun phrase. Attributes can indicate its type, give a normalized form, or associate it with a specific individual or thing by means of a unique identifiers.	

(Continued on next page)

# APPENDIX C TEI header (Continued)

TEI HEADER	DESCRIPTION	MARC EQUIVALENT AND NOTES
<extent>	describes the approximate size of the physical item. Use 2 for number of pages and dimensions	300
<publicationStmnt>	groups information concerning publisher, place of publication, and date of the text.	260
<pubPlace>	contains the name of the place where a bibliographic item was published. Do NOT repeat field; instead, use semicolons to divide multiple publication places	260, 534   c _1
<publisher>	provides the name of the organization responsible for the publication or distribution of a bibliographic item. Do NOT repeat field; instead, use semicolons to divide multiple publishers	534   c _2
<date when="xxxx">	contains a date in any format, with normalized value in the value attribute, of bibliographic item's original publication. For most purposes, the year date (yyyy) will be adequate. If greater detail is required, enter dates as yyyymmdd.	260, 534   c _3
<seriesStmnt>	establish via national authority file	440, 490
<idno> -> e.g. <idno type="LC Call number"> or type="isbn-13" or type="isbn-10"	identification number, such as LC call number	050, 090, 099
<encodingDesc> <p>	ENCODING DESCRIPTION documents the relationship between an electronic text and the source or sources from which it was derived.	PARENT ELEMENT FOR ALL CHILD ELEMENTS IN THIS SECTION
<appInfo>	records the application which edited or created the XML file	Has <application> and <label> child elements
<projectDesc>	Enter a description of the purpose for which the electronic file was encoded	
<editorialDec n="...">	Record encoding level for the content as an arabic numeral in the n attribute. As content of this element, record editorial decisions made during encoding and notes about omissions of material found in the original work	VERY IMPORTANT. Make sure you include this
<tagsDecl> <namespace name="http://www.tei- c.org/ns/1.0"> <tagUsage>	<tagUsage> must be either <tagUsage gi="div1">Numbered divs used. </tagUsage> OR <tagUsage gi="div"> Unnumbered divs used. </tagUsage>	

<classDecl>	contains one or more taxonomies defining any classificatory codes used elsewhere in the text.	Parent element for <taxonomy>, <bibl>, and <title>
<taxonomy xml:id="LCSH">	defines a typology used to classify texts either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy.	650
<bibl>	contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged	
<title>	contains title of taxonomy, e.g. Library of Congress Subject Headings	
<profileDesc>	PROFILE DESCRIPTION provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.	PARENT ELEMENT FOR ALL CHILD ELEMENTS IN THIS SECTION
<langUsage>	describes the languages, sublanguages, registers, dialects, etc., represented within a text.	
<language ident="">	Use the ISO 639-2 standard (which is the same as the MARC language codes) for the language attribute. The content of the element should be the name of the language given in ISO 639-2	Child element of <langUsage>
<textClass>	groups information which describes the nature or topic of a text in terms of a standard classification scheme, thesaurus, etc.	
<keywords scheme="LCSH">	contains a list of keywords or phrases identifying the topic or nature of a text; if the keywords come from a controlled vocabulary, it can be identified by the scheme attribute	Child element of <textClass>
<term>	contains a single-word, multi-word or symbolic designation which is regarded as a technical term.	Child element of <keywords>
<revisionDesc>	REVISION DESCRIPTION contains change, resp, and item for any revisions to encoding. Use the specific codes to note revisions rather than free text description. Include the entire date (e.g., 19991101).	OPTIONAL. USE ONLY WHEN REVISING AN OLDER, PUBLISHED TEI DOCUMENT
<change>	records each significant change to the TEI document	
<date>	the date of change in ISO 8601 form (YYYY-MM-DD)	
<respStmt> <name>	initials or other identifying information about the person or process making the change	
<item>	prose description of the change	

**APPENDIX D** Sample TEI header

```

<teiHeader xml:lang="en">
  <fileDesc>
    <titleStmt>
      <title type="main">An electronic version of Early history of the Cleveland
        Public Schools</title>
      <author> <persName key="AFREESE">Freese, Andrew</persName>
        </author>
      <respStmt>
        <name xml:id="ksl">Kelvin Smith Library, Case Western
          Reserve University</name>
        <resp>Publisher of TEI-conformant electronic version.</resp>
      </respStmt>
      <respStmt>
        <name xml:id="ID54">NAME</name> <resp>TEI encoder and Header
          creator</resp>
      </respStmt>
    </titleStmt>
    <extent>12.924 MB</extent>
    <publicationStmt>
      <distributor>collection">KSL Digital Book Collection</distributor>
      <publisher>Digital Case, Kelvin Smith Library, Case Western Reserve
        University</publisher>
      <pubPlace>Cleveland, Ohio</pubPlace>
      <date>2009</date>
      <availability status="free">
        <p>This work is in the public domain and may be freely downloaded
          for personal or academic use.</p>
      </availability>
    </publicationStmt>
    <sourceDesc>
      <biblStruct>
        <monogr>
          <author>
            <persName key="AFREESE">Freese, Andrew</persName>
          </author>
          <title type="MARC245a">Early history of the Cleveland Public
            Schools</title>
          <imprint>
            <pubPlace>Cleveland, Ohio</pubPlace>
            <publisher>Robison, Savage &Co., Book Printers</publisher>
            <date when="1876">1876</date>
          </imprint>

```

```

<extent>128 p. : ill.</extent>
<extent>23 cm.</extent>
</monogr>
<note>by <persName key="AFREESE">Andrew Freese
  </persName>; Published by order of the Board of Education.</note>
<idno type="OCLC">755713</idno>
<idno type="LC_call_number">LA348.C6 F8 1876</idno>
</biblStruct>
</sourceDesc>
</fileDesc>
<encodingDesc>
<p>Data in the sourceDesc of the header comes from a pre-AACR2 record.
  Other data follows AACR2 when applicable.</p>
<projectDesc>
<p>Text collected as part of Cleveland History Project. Encoded in P5 level
  3, beginning Winter 2009. RTF version of text transferred to Open Office,
  then saved as P5. All typographical and encoding errors fixed first, then
  structural encoding.</p>
<p>TEI header follows AACR2</p>
</projectDesc>
<editorialDecl n="3">
<hyphenation eol="none">
<p>Hyphenated words that appear at the end of a line have
  been removed</p>
</hyphenation>
<quotation>
<p>Quotation marks have been retained</p>
</quotation>
</editorialDecl>
<tagsDecl>
<namespace name="http://www.tei-c.org/ns/1.0">
<tagUsage gi="div">Only un-numbered divisions are used.</tagUsage>
</namespace>
</tagsDecl>
<classDecl>
<taxonomy xml:id="LCC">
<bibl>Library of Congress Classification</bibl>
</taxonomy>
<taxonomy xml:id="LCSH">
<bibl>Library of Congress Subject Headings</bibl>
</taxonomy>
</classDecl>
</encodingDesc>
<profileDesc>

```

```
<langUsage>
<language ident="en">English</language>
</langUsage>
<textClass>
<classCode scheme="#LCC"></classCode>
<keywords scheme="#LCSH">
<term>Public schools–Ohio–Cleveland</term>
</profileDesc>
</teiHeader>
```

**APPENDIX E** In-house encoding procedures

<b>DATE</b>	
<b>YOUR NAME</b>	
<b>TARGET COMPLETION DATE</b>	
<b>FILE NAME</b>	
<b>FULL CITATION, INCLUDING EDITION, VOLUME, AND SERIES (If Applicable)</b>	
<b>List text characteristics, with page numbers (n.b. These are SOME Characteristics; Refer to P5 Instructions for more)</b>	
<b>TEXT CHARACTERISTIC</b>	<b>PAGE(S)</b>
Table of Contents	
Dedication	
Books/Parts (major sections within text)	
Chapters	
Sections within Chapters	
Epigraph	
Frontispiece	
Acknowledgements	
Block quotations	
Italics	
Boldface	
Underline	
Tables	
Strikes	
Superscript	
Subscript	
Quotes within Quotes	
Lists	
Footnotes	
Endnotes	
Foreign words or phrases	
Poetic lines	
Poetic stanzas	
References within text	
Illustrations	
Maps	
Charts	
Photographs	
Fold-outs	
Marginalia	
Index	
Unfamiliar characters, abbreviations and macros	
Other	



**APPENDIX F** General recommendations**STRUCTURE**

1. Electronic text at all levels of encoding should begin with the transcription of the first word on the first leaf of the original work. It may be impractical or undesirable to transcribe and encode certain features of the text, such as publisher's advertisements or indexes, but if at all possible, they should be included as links to page images. Any omissions of material found in the original work should be noted in the `<editorialDecl>` in the TEI header.
2. Use unnumbered divisions `<div>`, unless your text has obvious divisions, such as chapters with no complex subdivisions, in which case begin with `<div1>`.
  - a. Whether you use numbered or unnumbered division, include the following attributes for all `divs`: `type=`, `n=`, and `xml:id=`. Remember that `n=` can contain a repeatable value, but `xml:id` cannot. Also, `xml:id` must begin with a character, not a number or symbol
  - b. Regardless of whether you use a numbered or unnumbered division, include `<front>`, `<body>`, and `<back>` in the structure:
  - c. `<front>` includes title pages, frontispieces, table of contents, introductions, and prefaces
  - d. `<body>` includes the primary text
  - e. `<back>` includes endnotes, publishers' advertisements, bibliographies, indices, colophons, and anything else following the book's body
3. Use `<milestone>` to mark a boundary point separating any kind of section of a text, as indicated by changes in a standard reference system, where the section is not represented by a structural element. Include attributes `unit=` and `xml:id=`
4. Page breaks should be encoded using the `<pb>` element, which should demark the top of a page (i.e. the text of page seven should immediately follow `<pb n="7"/>`), and should always be contained within a division (`<div>`). E.g., a page break that occurs between chapters 2 and 3 should be encoded near the top of the `<div>` that holds chapter 3 (rather than near the bottom of the `<div>` that holds chapter 2)
  - a. Include `n=` and `xml:id=` attributes within all `<pb />` to serve as reference identifiers to segments of the text: `n=` can be non-unique; `xml:id` must begin with a letter, and must be unique.

**COMMON ELEMENTS**

1. Use `<foreign>` for words or phrases in a language other than English. Include `lang=` attribute. E.g. `<foreign lang="fre" rend="italic">au naturale</foreign>`
2. Set the Title page within `<docTitle>`. Include `<titlePart>` within `<docTitle>`, with `type="main"` attributes within `<titlePart>`
3. Use `<byline>` for primary statement of responsibility, with `<docAuthor>` nested within `<byline>`
4. Use `<docImprint>` to encode the name of the publisher, place of publication, and date. Date uses `<date value= />` within `<docImprint>`
5. Quotation
  - a. Use `<q>` for direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages that are mentioned but not used. When in doubt, use `<q>`
  - b. Use `<quote>` for phrases or passages attributed by the narrator or author to some agency external to the text. Passages cited from other works
6. Use `<p>` for all paragraphs. If a paragraph runs from the bottom of one page to the top of the next, insert `<pb n=rend=xml:id= />` within the paragraph—do NOT break up the paragraph by inserting a `</p>` when the paragraph has not in fact ended
7. Delete all hyphens that occur within words, especially for words between pages. Make note of this in the header `<editorialDecl>`
8. Check all quotation marks. Change to `<q>` (most common) or `<quote>`

*(Continued on next page)*

**APPENDIX F** General recommendations (*Continued*)

- 
9. Fix `<emph>` to `<hi>` and either include `rend=` within the tag or in header's `<tagDecl>`
  10. Fix all footnotes and endnotes. Use `<note type=place=n=anchored=></note>`
  11. Fix all tables, lists, and images. Have links to images and complex tables
    - a. tables: `<table> <row> <cell>`
    - b. lists: `<list type=><item>`
    - c. images: `<figure> <image url=/> <figDesc> </figure>`

## OXYGEN TIPS

1. Use `<f4>` for spell check
  2. Use Edit → Insert from character map for symbols, etc.
  3. Use FIND—REPLACE for global changes
  4. Use Document → Source → to uppercase to change words to all uppercase
  5. Use `<ctrl> + <shift> + <y>` so that lines do a hard-return
  6. Highlight a word or phrase or section. Right-click, and scroll to “refactoring.” Use to surround what you highlighted with a particular tag
  7. To repeat the previous function, you can `<ctrl> + </>`
  8. Use `<ctrl> + <+>` to increase the font-size on the screen
  9. In Oxygen, periodically check for well-formedness (green checkbox) and validity (red checkbox)
-

Copyright of Journal of Library Metadata is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.