Incorporating Text Encoding Initiative (TEI) Projects in Technical Services: An Examination of Possibilities and Potential Pitfalls

Richard Wisneski, Head, Bibliographic and Metadata Services
Stephanie Church, Assistant Acquisitions and Metadata
Suzhen Chen, Cataloger and Metadata

Kelvin Smith Library
Case Western Reserve University
May 2011
“Stronger involvement of libraries in text development will enhance functionality of electronic texts and improve information retrieval….but some may see [this] as crossing professional boundaries”

Text encoding marks up a document in XML to capture metadata (administrative, descriptive, technical, preservation) AND represent textual features important for research.

Examples:

- **Women Writers Online** (Brown University)
- **Illinois Harvest** (University of Illinois)
- **Poetess Archive** (University of Miami at Ohio)
- **Louisa A. Doane’s Journal** (Ohio State University)
- **Jane Austen’s Fiction Manuscripts**
- **Phillis Wheatleys’ Poems on Various Subjects** (Brown University)
TEI Headers are akin to Cataloging. Encoding texts themselves is NOT cataloging, but rather performing descriptive mark-up.

Text encoding does NOT attempt to provide one unique, authoritative version of a work. It often pairs the document with interpretation (markup and metadata).

Text encoding does NOT provide one static, permanent markup for a document. While there can be alternative markup in certain instances, there can be incorrect markup.

Text encoding (TEI) is NOT meant to have an encoding recommendation for all possibilities, but rather intends to be customized and modified within TEI guidelines.

TEI Consortium Best Practices for Libraries
Text from *Uncle Remus*:
One day, arter Brer Fox bin doin' all dat...
[One day, after Brer Fox been doing all that...]

Mark-up can be 1 of 4 options:
A. `<choice>
   <sic>arter</sic>
   <corr>after</corr>
</choice>
B. `<orig>arter</orig>`
C. `<reg>after</reg>`
D. `<choice>
   <orig>arter</orig>
   <reg>after</reg>
</choice>`

• Tailor searching under specific genres (e.g. verse, drama, prose)

• Search different formats (e.g. manuscript, poetry, drama)

• Search across collections (e.g. Illinois Harvest)

• Search by historical or geographic period

• Search by title, author, date, subject headings, etc.

• Represent structural features in text, including:
  – Structure (chapters, sections, sub-sections, parts, etc.)
  – Deletions, insertions, strikethroughs, supralinear additions, etc.
  – Footnotes
  – Captions
What Is the Text Encoding Initiative (TEI)?

• **Technically:** a standards organization for humanities text encoding
• **Organizationally:** an international membership consortium
• **Socially:** a community of people and projects


Cataloging : AACR2 :: Text Encoding : TEI Guidelines
Why TEI?

• To allow researchers to have access to an electronic text that does not require special-purpose software or hardware
• To analyze information – provide a standard text-encoding scheme and metadata language which accommodates searching, retrieval, etc.
• To share information – have a standard format for data interchange in humanities research
• To create stand-alone electronic text with hierarchy identified
• To provide nuanced search capabilities that go beyond keyword or phrase-searching
• Emphasis on divisions within text, tables, lists, notes, front and back matter
THE PROJECT
Manuscripts Relating to the Early History of the Western Reserve, 1795-1869

• There are over 230 documents in the Manuscripts collection. Project is to digitize and text-encode many items in this collection.

• Those items not being text-encoded are due to practical matters, such as length, extensive damage to pages, illegibility, material type [financial logbooks])

• Manuscripts are currently not available except on-site. Microfilm copy available at Western Reserve Historical Society and Cleveland State State University

• Charles W. Whittlesey (1808-1886) responsible for the acquisition of the materials

• Collection consists of business records, government documents, land surveys and field notes, journals and diaries, correspondence, writings, statements, addresses, newspaper clippings, court proceedings, and notes
Cleveland, Nov 30, 1868 Ch Whittlesey, Esq.

Dear Sir,

In compliance with your request I proceed to state some facts that recur to me relating to the early history of Cleveland.

I came to Cleveland in June 1826. It was then claimed by the citizens of the village to contain the inhabitants. Of the clergymen residing here at that time I can recollect only the Rev. L. Bradstreet, of the Presbyterian denomination, of physicians, Dr. David Song and Dr. McIntosh, of lawyers, Leonard [Cor], Saml. Cowles, Reuben Wood, John W. Wiley, J.J. Andrews, John M., Allen, 7 and Saml. [Moriswaker], of merchants, Nathan Ferry, corner of Superior and Walton Streets, F.M. Weddell, corner of Superior and Band streets and Joel Stratton, on the South side of Superior Street. Herchel Foot had a small book store, corner of Superior Street and the Public Square. John Blair and W.M. Golding had each a forwarding and commission store on the River.

What gave to Cleveland is first impetus in the way of prosperity and wealth was the location of the terminus of the Ohio Canal at this place.

The canal was first opened to navigation on the 4th of July 1827. At that time only 37 miles were navigable, extending from Cleveland to Lock No. 17 near Akron. Our national anniversary this year (1827) was used for a double purpose by the citizens of Ohio. Besides celebrating our Independence of Great Britain, it was used by them for celebrating the opening of navigation on the Ohio Canal. Besides the Board of Canal Commissions, the celebration of Cleveland was attended by most of the officers of state and by a large concourse of other citizens. A procession was formed which marched to the Public Square where an oration was delivered by Reuben Wood afterwards for fourteen years one of the Supreme Judges of Ohio, and subsequently twice elected governor of the State. After listening to the oration the procession returned to the Mansion House where a sumptuous dinner was served by the host, S. M. [Sugrue].

Notwithstanding the obvious benefits of the canal to Cleveland, of the canal, some of its older inhabitants, for some years often navigation in it had commenced, contended...
Goals / Wishes:

• Create separate interface for just this collection
• Include georeferencing
• Build search interface for names, dates, and places referenced
• Connect to online Encyclopedia of Cleveland History
• Include Timeline, using @when-iso within <date > elements
Our Workflow

• Manuscripts are on microfilm. Reviewed by me, noting length, complexities, potential issues and challenges regarding encoding (e.g. tables, lists, structural divisions)

• I meet with text encoder to go over text

• Text Encoders mark text following TEI P5, **Level 4**:

• I provide in-house training in XML, XSLT, TEI, accompanying metadata (MODS, METS, Dublin Core)

• I provide in-house documentation on metadata, TEI, transcribing

• I review all encoded texts before they are submitted to institutional repository
LEARNING TRANSCRIPTION, TEXT-ENCODING
Learning TEI

Learning Text Encoding as Secondary Job Responsibility

- Familiarity with XML, HTML, and metadata helpful

- Oxygen XML Editor, licensed metadata editing software, used for metadata creation

- TEI Best Practices
  http://purl.oclc.org/NET/teiinlibraries
Learning TEI

Coding a Manuscript

- Standards and Practices Documentation
- Create a Diplomatic Transcription
- Use Oxygen and Adobe Photoshop CS3 simultaneously to transcribe and view the manuscript
- Create Dublin Core, MODS, and METS records
had separated her fortunes from his, the " large property he once possessed had become " reduced, and he could not procure the necessary " bail." The result was a lodgement in this penitentiary from whence on account of his noble conduct towards the sick during the prevalence of cholera in that institution in <date when-iso="1849">1849</date> he was liberated by the president, <abbr>Gen.</abbr> <expan>[General] </expan> <persName>Taylor</persName>.<p></p> <p><persName>Ashley</persName> had died there in <date when-iso="1838">1838</date>. " After various vicissitudes they started in <date when-iso="1842">1842</date> a second " scheme of financiering, in which, if they had " succeeded, they would have rivalled the board comptrollers of the state bank. This was to encircle the world. They discounted an immense amount of bills on the U.S. Banks with which they contemplated visiting <placeName>Europe</placeName>, and even <placeName>China</placeName>, " exchanging U.S. Bank paper for the products " of those countries." [<persName>Bierce's</persName> sketches pp 45] There are incidents connected with the life and death of <unclear resp="#sap" cert="unknown" reason="illegible">[these]</unclear> polished gang of robbers that would make a striking figure in romance. If they had congregated on the banks of the Rhine instead of the Cuyahoga and instead of inhabiting taverns, groceries, and caves, had carried on their operations in ancient castles, they would furnish in
Challenges as a Transcriber

- Minimal Time Allotment
- Evolution of Institutional Practice
- Limited Opportunity for Continued Education
- Encoded Transcription as part of Whole
General Issues To / Challenges with Text Encoding Projects
The articles of association provided that
when one third of the partners in interest should
wish to have their lands set off in a tract, it
was to be done. This agreement was made
sight unseen, made but before the final draft
was seen. Proprietary representing $87,000,
valuing (not worth $84,575.85) determined to form
a new company to be called the "New Land
Company." They believed they could make
this business more economically than the company
constituted 1795, although the company...
Text Encoding Manuscripts

Manuscript

Linguistic

Literary
Manuscripts

Encoding challenges

- **Revision**
  - Who, how many, how to encode
- **Analysis**
  - Analyze the text including grammar and punctuation
- **Interpretation**
  - How to interpret linguistic and literary characteristics of the text
Manuscripts

Transcription

- Describe the physical condition
- Recognize handwriting with different styles
- Encoding using contextual information
- With the aid from the background information
Some considerations

Resources
- Time
- Cost
- Staff

Evaluation
- Value
- Users
- Alternative access

Access
- Digital representation (application, software.....)
- Electronic access
- Preservation
Digital representations

- Format
- Customization
- Search and discovery tools
Answer me, year of repulses. After a certain disastrous campaigns.

Answer me, traitor! year of repulses!
How will the poets, the lecturers of ages hence, look back to you, & to me also?
What themes will they make a 0-year out of you, O year? (themes for ironical sarcastic laughter?)
What are the ballads & proofs to be finally made shown of out of you?
Are they not indeed to be shown with pride, as by poets bards descended of mine-own from me? by my children?

http://www.whitmanarchive.org/manuscripts/transcriptions/pml.00006.html
[Den Haag, 29 september 1872].

[Waarde] Theo,
Dank voor je brief, het deel mij genoegen dat je weer goed aangekomen zijt. Ik heb je de eerste dagen gemist & het was mij vreemd je niet te vinden als ik s’middags thuish kwam.

Wij hebben prettige dagen samen gehad, en tusschen de droppeltjes door a toch nog al eens gewandeld & het even en ander gezien.

Wat vreellijk weer, je zult het wel benauwd hebben op je wandelingen naar Ois[ter]wijk. Gisteren is het harddravrij geweest ter gelegenheid van de tentoonstelling, maar de illuminatie & het vuurwerk zijn uit gesteld, om het slechte weer, het is dus maar goed dat je niet gebleven zijt om die te zien. Groeten van de familie Haanebeek & Roos. Steeds

je liefh.
Vincent

5. Van Gogh boarded with Willem Marinus Roos and his wife Dina Margrieta van Aalst, who lived at Lange Beestenmarkt 32 in The Hague. The couple had no children.

http://vangoghletters.org/vg/letters/let001/letter.html
Sir George Ripley his Epistle to K. Edward unfolded.

1 This epistle as it was directed to a King who was in his generation both wise & valiant so it doth comprise the whole secret, which in it is both learnedly described & artificially veiled, yet as the Author himself testifieth that in this epistle hee doth unty the maine knot, soe I can & doe testify with him that there is nothing desirable for the true attaining of the mastery both in the Theory & Practique of which it is not in this short ep fully taught. This then I intend as a corollary or rather a key to all that I have written by

http://webapp1.dlib.indiana.edu/newton/mss/dipl/ALCH00041
S\textsuperscript{r} George Ripley his Epistle to K. Edward unfolded. 1

1 This ep. as it was directed to a K who was in his generation both wise & valiant so it doth comprise the whole secret, w\textsuperscript{ch} in it is both learnedly described & artificially veiled, yet as y\textsuperscript{c} Author himself testifieth y\textsuperscript{t} in this ep hee doth untly the maine knot, soe I can & doe testify w\textsuperscript{th} him y\textsuperscript{t} there is nothing desirable for the true attaining of the mastery both in y\textsuperscript{c} Theory & Practique of w\textsuperscript{ch} it is not in this short ep fully taught. This y\textsuperscript{n} I intend as a corollary or rather a key to all y\textsuperscript{t} I \textit{have} written by
Search and discovery tools

- Commercial products
- Locally developed tools
- Open source tool

Example: eXtensible Text Framework (XTF)
http://www.cdlib.org/services/publishing/tools/xtf/
Project: The Chymistry of Isaac Newton, Indiana University
http://webapp1.dlib.indiana.edu/newton/
Manuscripts

Maximize Value for text encoding

Projects:

• Communicate with faculty and student
• Advertise the products
• Share information with other entities
• Work collaboratively with other institutions
• Make best use of resources available
Questions:

- What is the role of tech services librarians? Should they be doing this kind of work?

- Should librarians also be involved in web development for such projects

- Feasibility of small libraries undertaking such work?

- Is it better to work with faculty directly who have research needs, or bring to light projects/collections others may be unaware of?
A Companion to Digital Literary Studies: http://www.digitalhumanities.org/companion/DLS/


