

# 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

# Introduction

**“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”**

**-- Suzana Sukovic (January 2002). “Beyond the Scriptorium: The Role of the Library in Text Encoding.” D-Lib Magazine 8:1, 1-8.**

# What Is Text Encoding?

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)

# What Text Encoding Is NOT

- 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](#)

# Example of Choices in Text-Encoding

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>`

See: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html> (p.74-75)

# Text Encoding Allows Users To...

- 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

Web site: <http://www.tei-c.org/>

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 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

# Our Book Viewer



CASE WESTERN RESERVE  
UNIVERSITY  
EST. 1826

Kelvin Smith Library  
Digital Case Book Viewer

## Digital Case

- About
- Policies
- Forms
- 3D Image Wall
- Contribute Content
- RSS Builder
- Browse by Subject
- Newsletter
- Help
- Reports
- Fact Sheet
- Deposit Agreement
- KSL Home

### Contact Us

**Email:**  
digitalcase@case.edu  
**Phone:**  
216.368.6513



Back to Digital Case About This Book Help Download PDF / TEI  / 2  Search

- Image View
- Full Text View
- Compare View

page [1]

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 [Car], Saml. Cowles, Reuben Wood, John W. Willey, J.J. Andrews, Jon N. [Allen] and Saml. [Meriweather]; of merchants, Nathan Perry, corner of Superior and Walton Streets, P.M. Weddell, corner of Superior and Band streets and Joel Scranton, on the South side of Superior Street. Herschel Foot had a small book store, corner of Superior Street and the Public Square. John Blair and W.M. Giddings had each a Forwarding and Commission Sore 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 offices of state and by a large concourse of other citizens. A procession was formed which marked 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 dated 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, E. M. [Segres].

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

[KSL Home](#) | [BlackBoard](#) | [Site Map](#) | [Privacy](#) | [Contact Us](#) | [Ohio](#) | [Libraries of Case](#)  
Kelvin Smith Library | 11055 Euclid Avenue | Cleveland, OH 44106-7151 | 216-368-3506

This web site and all of its contents are ©2007 The CWRU. All Rights Reserved.

# 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**

# Learning TEI

"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 the  
penitentiary from whence on account of his noble  
conduct towards the sick during the prevalence  
of Cholera in that institution in 1849 he was  
liberated by the president, Gen. Taylor.

Ashley had died there in 1838. "After various  
"vicissitudes they started in 1842 a second  
"scheme of financiering, in which, if they had  
"succeeded, they would have rivalled the board  
"of 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

<pb facs="wrhsms06169-00006.jp2" n="[6]"/>

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><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

# Learning TEI

## Challenges as a Transcriber

- **Minimal Time Allotment**
- **Evolution of Institutional Practice**
- **Limited Opportunity for Continued Education**
- **Encoded Transcription as part of Whole**

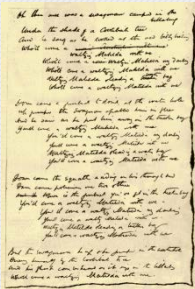
# General Issues To / Challenges with Text Encoding Projects

# Text Encoding Manuscripts

## Handwriting Identification

+ The articles of association provided that when one third of the parties in interest should wish to have their lands set off in a body it was to be done. This request was not made but before the first draft <sup>eighteen</sup> ~~several~~ large ~~owners~~ proprietors representing \$84,536.35 dollars (one party is \$84,595.85) determined to form a Sub Company to be called the "Ohio Land Company". They believed they could manage their business more economically than the Company of 1795, although the <sup>constitution</sup> ~~principles~~ of the two.

# 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

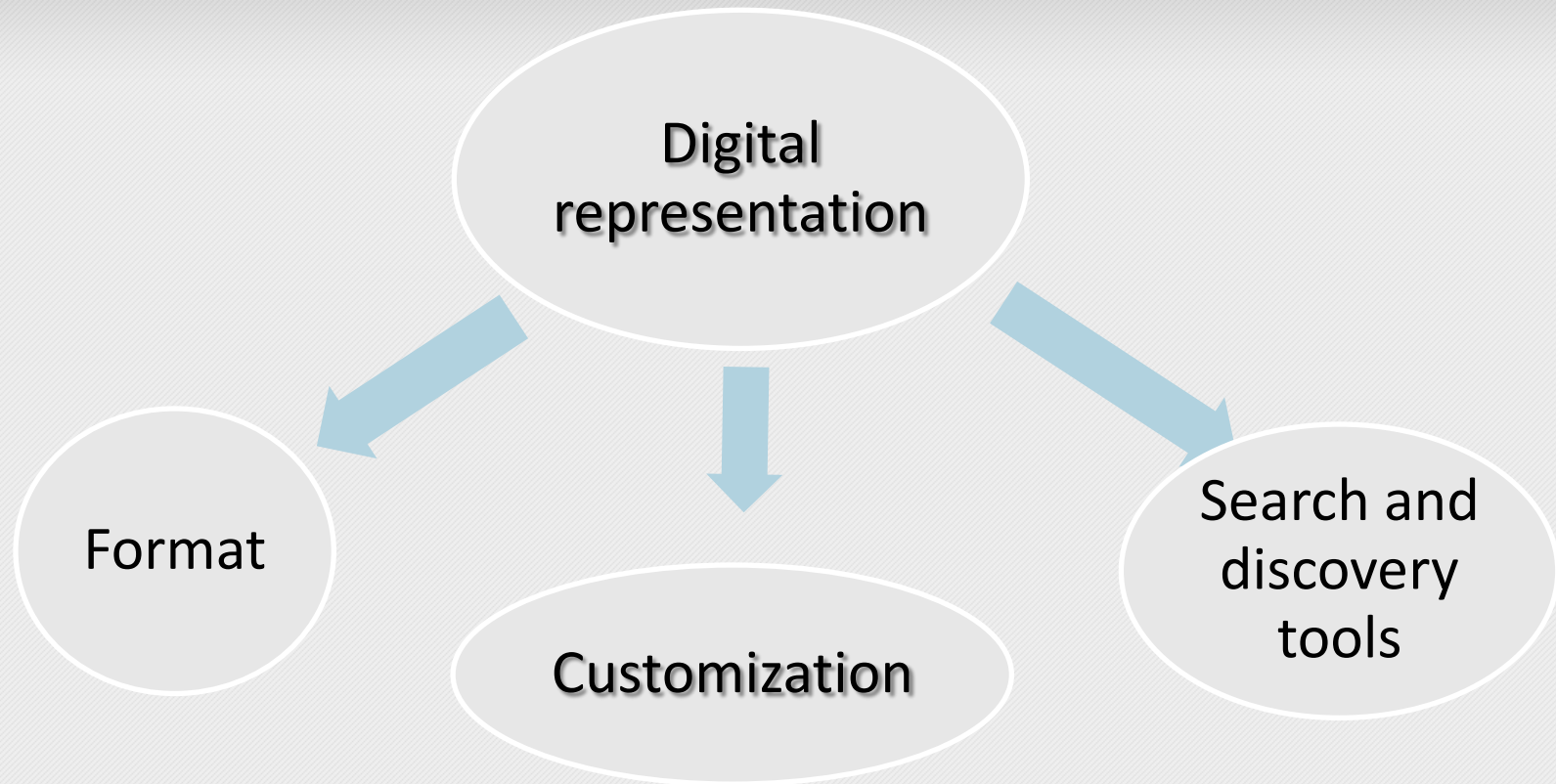
## Evaluation

- Value
- Users
- Alternative access

## Access

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

# Digital representations



# Digital representation (1)

~~Answer me, year of repulses~~ After <sup>a certain</sup> 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 <sup>a</sup> ~~O~~-year out of  
you, O year? (themes for ironical <sup>sarcastic</sup>  
laughter?)

What are the ~~ballads~~ & proofs to  
be finally made <sup>shown</sup> ~~of~~ out of you?

Are they not indeed to be shown  
with pride, as by poets <sup>bards descended</sup> of  
~~mine own~~ <sup>from me</sup> ? by my children?

# Digital representation (2)

original text + line endings facsimile translation notes artworks

1r:1  
[Den Haag, 29 september 187]2.

[Waarde] Theo,  
Dank voor je brief, het deed 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 t'huis kwam.  
Wij hebben prettige dagen samen gehad, en tusschen de  
droppeltjes door [a](#) toch nog al eens gewandeld & het een en  
ander gezien.  
Wat vreesselijk weer, je zult het wel *benauwd* hebben op je  
wandelingen naar Ois[ter]wijk. [1](#) Gisteren is het  
harddraverij geweest ter gelegenheid van de  
tentoonstelling, [2](#) maar de illuminatie & het vuurwerk zijn  
uit gesteld, om het slechte weer, [3](#) het is dus maar goed dat  
je niet gebleven zijt om die te zien. Groeten van de familie  
Haanebeek [4](#) & Roos. [5](#) Steeds

je liefh.  
Vincent

001

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

original text + line endings facsimile translation notes artworks

1r:1  
1' [Den Haag, 29 september 187]2.

2' [Waarde] Theo,  
3 Dank voor je brief,  
4 het deed mij genoegen dat je  
5 weer goed aangekomen zijt.  
6 Ik heb je de eerste dagen ge-  
7 mist & het was mij vreemd  
8 je niet te vinden als ik s'mid-  
9 dags t'huis kwam.  
10 Wij hebben prettige dagen sa-  
11 men gehad, en tusschen de  
12 droppeltjes door [a](#) toch nog al  
13 eens gewandeld & het een en  
14 ander gezien.  
15 Wat vreesselijk weer, je zult  
16 het wel *benauwd* hebben  
17 op je wandelingen naar  
18' Ois[ter]wijk. [1](#) Gisteren is het hard-  
19 draverij geweest ter gelegenheid van

# Digital representation (3)-1



[Home](#)

[Site Tools](#)

[Project Information](#)

Search: enter term, hit

[BROWSE MANUSCRIPTS](#)

## The Chymistry of Isaac Newton

[Introduction](#) | [Normalized Transcription](#) | [Diplomatic Transcription](#)

[Page Images](#) | [Manuscript Information](#)

**Keynes MS. 52, King's College Library, Cambridge University**

<1r> Page Image

Sir George Ripley his Epistle to K. Edward unfolded. 1

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



# Digital representation (3)-2



Home

Site Tools

Project Information

Search: enter term, hit return

BROWSE MANUSCRIPTS

## The Chymistry of Isaac Newton

[Introduction](#) | [Normalized Transcription](#) | [Diplomatic Transcription](#)

[Page Images](#) | [Manuscript Information](#)

**Keynes MS. 52, King's College Library, Cambridge University**

<1r> Page Image

S<sup>f</sup> 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<sup>ch</sup> in it is both learnedly described & artificially veiled, yet as y<sup>e</sup> Author himself testifieth y<sup>t</sup> in this ep hee doth unty the maine knot, soe I can & doe testify w<sup>th</sup> him y<sup>t</sup> there is nothing desirable for the true attaining of the mastery both in y<sup>e</sup> Theory & Practique of w<sup>ch</sup> it is not in this short ep fully taught. This y<sup>n</sup> I intend as a corollary or rather a key to all y<sup>t</sup> I <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

# Discussion

## 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?**

# References

- A Companion to Digital Literary Studies:  
<http://www.digitalhumanities.org/companion/DLS/>
- Pichler, Alois (Aug. 1995). Transcriptions, Texts and Interpretation. *Austrian Ludwig Wittgenstein Society*, pp. 690-695
- Burnard, Lou, Katherine O'Brien O'Keeffe, and John Unsworth, eds. (2006). *Electronic Textual Editing*. New York: MLA
- Hoover, David (Summer 2007). The end of the irrelevant text: Electronic texts, linguistics, and literary theory. *DHQ: Digital Humanities Quarterly* 1:2, 1-19.
- Tuohy, Conal (Oct. 2007). Topic maps and TEI – using topic maps as a tool for presenting TEI documents. Conference: TEI Day in Kyoto 2006.
- Piez, Wendell (Summer 2008). Something called “Digital Humanities.” *DHQ: Digital Humanities Quarterly* 2:1, 1-5.
- Wheeler, Robert, ed. (2000). *Visions of the Western Reserve: Public and private documents of Northeastern Ohio, 1750-1860*. Columbus: Ohio State University Press.
- Dalmau, M. and Melanie Schlosser (2010). Challenges of serials text encoding in the spirit of scholarly communication. *Library Hi Tech* 28:3, 345-359.