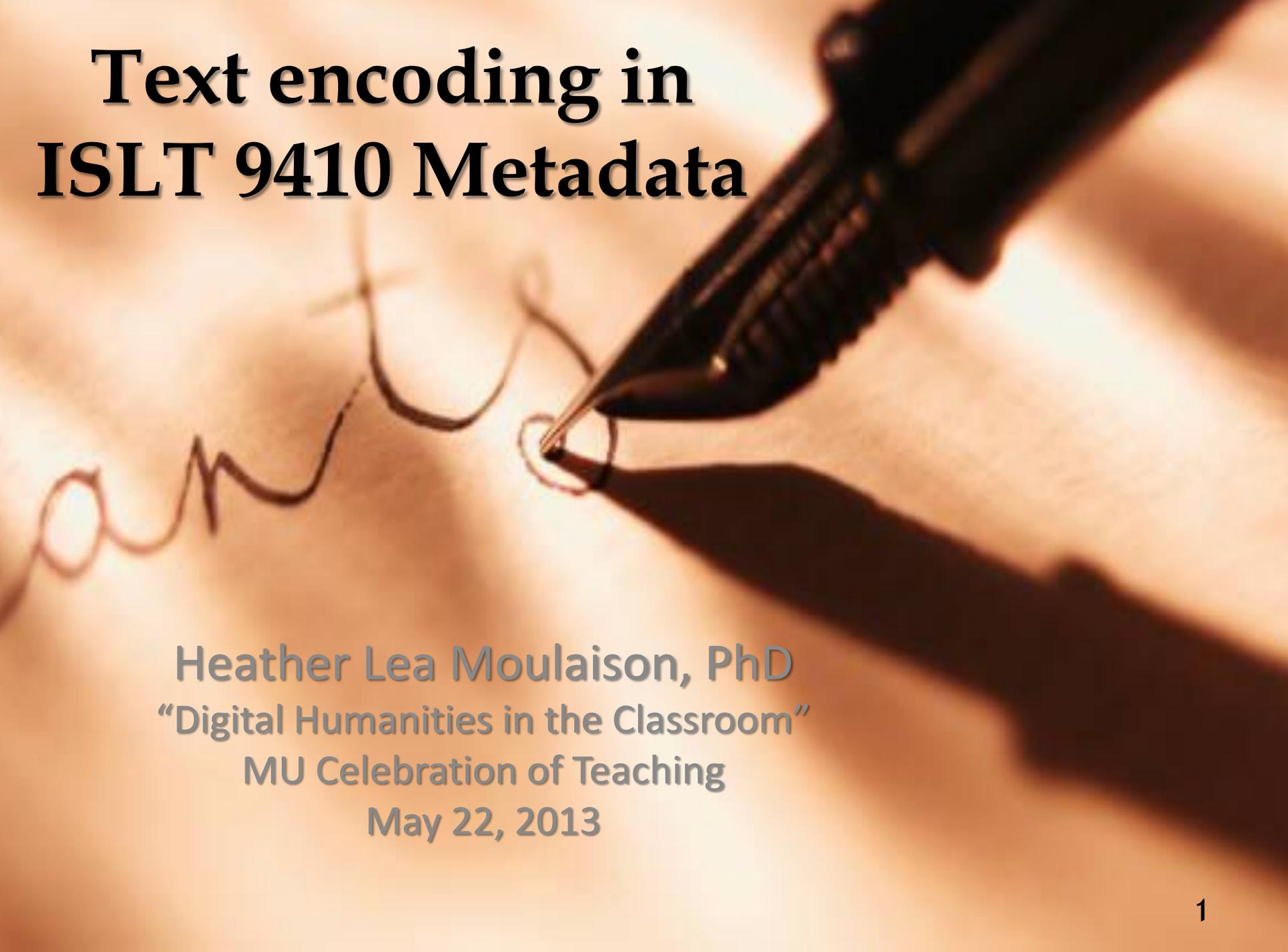


Text encoding in ISLT 9410 Metadata



Heather Lea Moulaison, PhD
“Digital Humanities in the Classroom”
MU Celebration of Teaching
May 22, 2013

What is text encoding?

- HTML permits documents to be marked up for the web.
 - Browsers understand how to render content based on the way it's marked up
- Digital humanities *text encoding* goes well beyond simple markup for display!

```
<!DOCTYPE html>
<html>
<body>

<h1>My First Heading</h1>

<p>My first paragraph.</p>

</body>
</html>
```



What is TEI?

- Text Encoding Initiative (TEI)
 - Trusted standard for the representation of texts in digital form.
 - Used primarily in the humanities, social sciences, and linguistics
 - The product of a group first brought together in the 1980s
 - Release of current version(TEI Guidelines, P3) in 1994
 - “The TEI is a very extensive encoding language and is intended to support very complex encoding of very complex documents” -- http://www.tei-c.org/About/faq.xml#body.1_div.3_div.1



Projects using TEI

- African American Women Writers of the 19th Century
- African Languages Lexicon Project (ALLEX)
- Alex Catalogue of Electronic Texts
- American Memory from the Library of Congress
- American Verse Project
- Ancient Inscriptions of the Northern Black Sea
- Aphrodisias in Late Antiquity (2004)
- Archimedes Palimpsest Project ATLAS : ATLA Series
- Autour des Res Gestae Divi Augusti
- Base de Français Médiéval – Old French Corpus
- Berardier.org: Édition électronique de Bérardier de Bataut, Essai sur le récit (1776)
- Berlin Intellectuals 1800-1830
- Boccaccio's Decameron
- Etc...

<http://www.tei-c.org/Activities/Projects/>

Benefits of TEI

- If carefully encoded and consistently applied, TEI permits documents to be:
 - Searched (using keywords!) through the search engines of the repositories where the files are housed
 - Especially important if the original was a manuscript (i.e. hand-written)
 - Permits certain *structures* to be searched (the stage directions of plays that are marked up)
 - Integrated with other documents, in other systems, or along with other standards potentially



Specifics of text encoding for DH

- Here is a typical use case scenario for the DH, and it happens to be what we did in class, too:
 - An archival manuscript, a diary, was the basis for the encoding project
 - Archival quality photos of the manuscript were used
 - Students received two pages and were asked to transcribe the content
 - Now the hand-written manuscript was searchable
 - Students then marked up the text using TEI



tionary history of America, but at 9 I was busy
memorising the Declaration of Independence
under a mother's promise of an American flag
when it should have done it. I never ~~learned~~^{memorized}
~~farther than its~~ ~~do~~ its assertions of the mere
facts of British oppression, but I never forgot its
great principles of the ^{natural} equality of all men in
~~the~~ civil and political rights, rights to which all
men are born and which are not something to
be earned by ^{or overruling} killing other men.

Even then, I remember, I asked myself
how men could declare such ideal truths and

Benefits to TEI in ISLT 9410

- TEI is an XML language, so students gained experience using an XML editor (free version) used by many professionals when encoding
 - TEI records were output as HTML and as ePub documents using the XML editor's XSLT capability
- The diary entries are being studied by a doc student in Arts & Sciences at MU
 - This work, it is hoped, will benefit him in his close study of the diaries' contents
 - These TEI records might, eventually, be housed in a repository and made available



Complements to text encoding in ISLT 9410

In a later class project to build a digital library, group members:

- Carried out an assessment of user needs
- Decided on functional requirements for metadata records
- Investigated standards used in similar projects, studied best practices and application profiles, and made sure their choices permitted interoperability and preservation

All the above factors influence choice of metadata schema(s), level of markup used, and aspects of the system in which records are housed and made accessible, and would be necessary to consider if doing their own digital humanities projects.

