**LIBRARY**
UNIVERSITY OF MICHIGAN

# TEI workshop session 9
# C H O I C E S

## You can encode (mark up) almost anything.

## But what should you encode?

### And to what depth?

### Your choices will be dictated by ...

- the nature of the material

- the character of your incoming data

- the amount of your funding

- the patience of your funders

- (how much time left till your retirement)

- the scale of the project (how many items)

- the scope and variety of the project

- the purpose of the project
  - desired functionality

  - expected (or guessed-at) audience

  - potential for repurposing

  - potential for sharing/reuse

- your own knowledge or ignorance

- the existence of standards
  - why to avoid them. They are:
    - complex, hard to use

    - not tailored to material

    - not supported by local expertise or compatible with local systems

    - not as good as what you can come up with yourself

    - etc.?

  - why to use them. You can
    - leverage community expertise

    - share data

    - share tools

    - entertain at least a faint hope of "sustainability"

  - library practice summarized in GUIDELINES. (http://www.tei-c.org/SIG/Libraries/teiinlibraries/)
    - a good starting-point

    - provide good suggestions rooted in actual practice

    - do not merely define but *apply* tags, with examples

    - offer five 'levels' of commitment:
      1. raw OCR marked off into pages, linked to page images

      2. = LEVEL 1 + chapter divisions and headings

      3. = LEVEL 2 + refinements. Text may (?) stand on its own.

      4. Better text (keyed or corrected), tagged enough to stand alone

      5. = LEVEL 4 + considerable manual intervention based on subject knowledge.

## Our own projects as examples:

  - (LEVEL I) Mass-digitization projects Making of America (http://quod.lib.umich.edu/m/moagrp/) and the Google-scanned books going into HathiTrust
    (http://www.hathitrust.org)

(http://www.hathitrust.org)

- (LEVEL IV) The Text Creation Partnership. (http://www.lib.umich.edu/tcp/docs/) See sample TCP file. (/files/departments/dpp/20101113-pfs/20101113-09b.xml)

- (LEVEL V) Middle English Dictionary (http://quod.lib.umich.edu/m/med) (and Compendium. (http://quod.lib.umich.edu/m/mec) See sample MED file. (/files/departments/dpp/20101113-pfs/20101113-09a.xml)

- (LEVEL IV) Knight's American Mechanical Dictionary (http://www-personal.umich.edu/~pfs/knight/index.html)

- (LEVEL V) The Faculty CV project (http://www-personal.umich.edu/~pfs/cvs/)

These differ in

- their adherence to standards

- their labor-intensity

- their longevity (the price of success?)

- their scale and scope

But share a common rationale:

- intelligible display

- intelligent navigation

- contextually useful search restrictions

- constraint by method and cost

- susceptability to incremental improvement

Page maintained by Paul Schaffner (/users/pfs)

Last modified: 11/22/2010

## Contact

## Copyright

**University of Michigan Library**

818 Hatcher Graduate Library South

913 S. University Avenue

913 S. University Avenue
Ann Arbor, MI 48109-1190
(734) 764-0400
contact-mlibrary@umich.edu

Give Feedback

Go To Mobile Site