Do You TEI?

Survey Findings of Text Encoding Practices in Libraries

Michelle Dalmau, Digital Collections Services, IU Libraries
Fall 2013 Digital Library Brown Bag Series
20 November 2013
Twitter:  @mdalmau;  Hashtag:  #dlbb
TEI-C Special Interest Group: Libraries

• Conveners & Research Partners-in-Crime
  – Michelle Dalmau, Digital Collections Services, Indiana University Libraries

Twitter: @mdalmau #dlbb
Do You TEI?
A Survey of Text Encoding Practices in Libraries

http://www.surveymonkey.com/s/tei_in_libraries

BACKGROUND
Following on papers, presentations and discussions that resulted from the theme of the 2009 Conference and Members’ Meeting of the TEI Consortium, “Text Encoding in the Era of Mass Digitization,” the launch of the AccessTEI program in 2010, and the recent release of the “Best Practices for TEI in Libraries” in 2011, it behooves us—stewards of text encoding initiatives in academic libraries—to better understand if and how text encoding practices have changed as a result of mass digitization initiatives, declining budgets, and an increased emphasis on “productionizing” digital library services.

METHODODOLOGY
★ Online survey
★ Posted from November 4, 2012 to January 31, 2013
★ No more than 30 minutes to complete
★ Consists of study information, determination of eligibility, background info about the institution (type of library, size, attitudes) & text encoding practices
★ Personal information is not being collected in this survey
★ Recruiting librarians and libraries staff via listservs, social media, etc.
★ Seeking a non-probability sample of respondents following a quota sample of 200 respondents
★ Targeting communities of practice rather than individuals to minimize bias that may occur
★ Take our survey!

GOALS
★ Profile libraries and library staff engaged in text encoding
★ Determine values shaping text encoding practices in libraries
★ Understand text encoding partnerships in libraries and beyond
★ Identify text encoding services offered in libraries
★ Identify TEI-C membership benefits most important to libraries
★ Identify how the “Best Practices for TEI in Libraries” are used in libraries

BENCHMARKS
The following data points gathered from raw data sources or publications will be used to compare/contrast, and, if possible, correlate findings from the online survey:
★ TEI-C membership profile of library institutions from 2005–2012
★ Evolution/devolution of electronic text centers within libraries from as early as 2000 to present

Michelle Dalmau, Indiana University Digital Library Program, mdalmau@indiana.edu
Kevin S. Hawkins, MPublishing, University of Michigan, kevin.s.hawkins@ultraslavonic.info

2012 TEI Conference and Members' Meeting
Texas A & M University, College Station, Texas, November 9, 2012
Scope in brief

• Developed a SurveyMonkey survey with yes-no, multiple-choice, and free-response questions.
• Announced online on November 4, 2013 and closed on January 31, 2013.
• Survey participants had to answer “yes” to saying they work in a library.
• We encouraged responses from more than one person at the same institution.
• 138 began the survey; 112 “completed” it
Survey Structure

• Study Information
• Determination of Eligibility
• Background Information About the Institution (type of library, size, attitudes)
• Text Encoding Practices
  – Standards
  – Collaborations/Partnerships
  – Types of Text Encoding Projects
  – And more!
History and Hypothesis
Select milestones in the library digitization of text

- 1992: Founding of IATH at Virginia
- 1997: Founding of GDZ at SUB Göttingen
- 1999: Text Creation Partnership formed
- 2004: Google announced mass-digitization partnership with leading research libraries
- 2005: Open Content Alliance created
- 2008: Formation of HathiTrust
- 2010: TEI-C launches AccessTEI
The Question of all Questions

How has advocating for such wide-ranging library objectives —from digital access and preservation to digital literacy and scholarship; from supporting nonexpressive/nonconsumptive research practices to research practices rooted in the markup itself —informed the evolution or devolution of text encoding projects in libraries?
Profile of Survey Respondents
Who responded?

- Of the 112 respondents, we can see from IP addresses that:
  - 55 are clearly affiliated with an institution; 41 of which are unique institutions
  - 57 are unidentifiable due to off-site internet connections (via ISPs)
Indicate the type of library for which you work.

- Academic Library: 82%
- Research Library: 9%
- Public Library: 4%
- National Library: 2%
- Special Library: 3%

Where is your institution located?

- North America: 75%
- Europe: 10%
- Asia: 2%
- No Response: 13%

What is the size of your academic institution based on student enrollment (or patrons served)?

- No Response: 22%
- Over 40,000: 13%
- 25,000-40,000: 20%
- 10,000-25,000: 24%
- 5,000-10,000: 14%
- Up to 5,000: 7%

(n=112)
What is the name of your unit or branch library?
(n=99 reported only one unit; n=9 reported more than one unit; n=4 no response)

- Technology: 32%
- General Library: 14%
- Special Collections: 11%
- Subject Library: 8%
- Publishing: 3%
- Public Services: 2%
- Preservation: 1%
- Digital Scholarship: 5%
- Collections: 4%
- Cataloging: 11%
- Archives: 6%
- Administration: 3%

Twitter: @mdalmau #dlbb
List the units or people (in terms of roles) in your library with which/whom you partner (n=58)

Twitter: @mdalmau #dlbb
TEI-C Membership
Is your institution a member of the TEI Consortium? (n=112)

- **Total Responses**
  - Yes: 39
  - No: 43
  - Unsure: 23
  - No Response: 23

- **Total Institutions**
  - Yes: 18
  - No: 20
  - Unsure: 7
  - No Response: 9

- **Total Unique Institutions**
  - Yes: 9
  - No: 6
  - Unsure: 2
  - No Response: 9

- **Total ISPs**
  - Yes: 21
  - No: 5
  - Unsure: 6
  - No Response: 14
TEI-C Member Institutions: 2005-2013
(missing 2012 data)

*Membership data (2005-2011) provided by Martin Mueller; Coded by Kevin Hawkins

Twitter: @mdalmau #dlbb
Would your institution become or remain a TEI Consortium member if doing so would allow for: (n=112)

- **Digitization discounts**: 25 responses, 87 no responses
- **Direct input into standards development**: 20 responses, 92 no responses
- **Community support**: 31 responses, 81 no responses
- **Training opportunities**: 43 responses, 69 no responses
- **Other**: 18 responses, 94 no responses

Twitter: @mdalmau #dlbb
Text Encoding Practices & Partnerships in Libraries
How many text encoding projects have you or members of your unit participated in over the years? (n=112)

- None: 4%
- No Response: 22%
- 1-5: 40%
- 6-10: 13%
- 11-15: 7%
- 16-20: 2%
- More than 30: 12%
- Twitter: @mdalmau #dlbb
In what ways do you or members of your unit support text encoding projects (select all that apply)? (n=112)

- Consulting: 44 responses, 68 no response
- Training: 44 responses, 68 no response
- Project management: 49 responses, 63 no response
- Establishing encoding workflows: 48 responses, 64 no response
- Text encoding/markup: 61 responses, 51 no response
- Web publishing: 43 responses, 69 no response
- Other: 7 responses, 105 no response

Twitter: @mdalmau #dlbb
Indicate how frequently you partner with the following people when undertaking a text encoding project:

- Almost Always/Often
- Sometimes
- Seldom/Never

<table>
<thead>
<tr>
<th>Professional Group</th>
<th>Almost Always/Often</th>
<th>Sometimes</th>
<th>Seldom/Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT Professionals</td>
<td>33</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Museum Professionals</td>
<td>4</td>
<td>58</td>
<td>23</td>
</tr>
<tr>
<td>Archivists</td>
<td>22</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td>Librarians &amp; Library Staff</td>
<td>37</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>Graduate Students</td>
<td>19</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>Faculty</td>
<td>33</td>
<td>17</td>
<td>19</td>
</tr>
</tbody>
</table>

Twitter: @mdalmau #dlbb
Indicate how you collaborate on text encoding projects with the following categories of people:

- Archivists
- Faculty
- Graduate Students
- IT Professionals
- Librarians & Library Staff
- Museum Professionals

### Web Publishing
- Archivists: 11
- Faculty: 22
- Graduate Students: 11
- IT Professionals: 40
- Librarians & Library Staff: 24
- Museum Professionals: 0

### Text Encoding/Markup
- Archivists: 19
- Faculty: 30
- Graduate Students: 30
- IT Professionals: 14
- Librarians & Library Staff: 35
- Museum Professionals: 4

### Establishing Encoding Workflows & Support
- Archivists: 19
- Faculty: 31
- Graduate Students: 15
- IT Professionals: 18
- Librarians & Library Staff: 35
- Museum Professionals: 3

### Project Management
- Archivists: 12
- Faculty: 24
- Graduate Students: 9
- IT Professionals: 16
- Librarians & Library Staff: 33
- Museum Professionals: 2

### Training
- Archivists: 14
- Faculty: 22
- Graduate Students: 25
- IT Professionals: 7
- Librarians & Library Staff: 30
- Museum Professionals: 3

### Consulting
- Archivists: 26
- Faculty: 42
- Graduate Students: 16
- IT Professionals: 19
- Librarians & Library Staff: 33
- Museum Professionals: 8
Rank the nature of your text encoding projects by "dragging and dropping" each option into place (1 is most common, 8 is least common)

- Rare Books & Manuscripts
- Archival Materials
- Faculty or Librarian Digital Research Projects

Twitter: @mdalmau #dlbb
What type of materials do you digitize and encode? Indicate frequency.

- **Manuscripts / Handwritten Materials**
  - Almost Always/Often: 35
  - Sometimes: 16
  - Seldom/Never: 15

- **Printed Books**
  - Almost Always/Often: 35
  - Sometimes: 14
  - Seldom/Never: 17

- **Transcriptions of Audio or Video**
  - Almost Always/Often: 6
  - Sometimes: 7
  - Seldom/Never: 43

- **Born-Digital Works**
  - Almost Always/Often: 9
  - Sometimes: 12
  - Seldom/Never: 38

- **Catalogs of Manuscripts or Other Items**
  - Almost Always/Often: 5
  - Sometimes: 7
  - Seldom/Never: 42

- **Newspapers**
  - Almost Always/Often: 5
  - Sometimes: 11
  - Seldom/Never: 44

- **Serials**
  - Almost Always/Often: 9
  - Sometimes: 13
  - Seldom/Never: 40

Twitter: @mdalmau #dlbb
Indicate the frequency in which materials from the following time periods are encoded:

- Almost Always/Often
- Sometimes
- Seldom/Never

### 19th century and after
- Almost Always/Often: 43
- Sometimes: 15
- Seldom/Never: 8

### 18th century
- Almost Always/Often: 23
- Sometimes: 17
- Seldom/Never: 17

### 17th century
- Almost Always/Often: 18
- Sometimes: 11
- Seldom/Never: 23

### 16th century
- Almost Always/Often: 14
- Sometimes: 9
- Seldom/Never: 27

### 15th century and before
- Almost Always/Often: 11
- Sometimes: 6
- Seldom/Never: 29

Twitter: @mdalmau #dlbb
Describe the types and frequency of encoding projects you undertake based on the following levels of encoding:

- **Almost Always/Often**
- **Sometimes**
- **Seldom/Never**

### Basic reformatting of text
- Almost Always/Often: 27
- Sometimes: 7
- Seldom/Never: 19

### Mid-level structural encoding
- Almost Always/Often: 36
- Sometimes: 14
- Seldom/Never: 16

### Richer encoding for content analysis
- Almost Always/Often: 22
- Sometimes: 18
- Seldom/Never: 22

### Scholarly encoding projects
- Almost Always/Often: 25
- Sometimes: 15
- Seldom/Never: 22

Twitter: @mdalmau #dlbb
Level of Encoding v. Types of Materials (top two items)

- Basic Reformatting of Text (Q89)
- Mid-level Structural Encoding (Q90)
- Richer Encoding for Content Analysis (Q91)
- Scholarly Encoding Projects (Q92)

Twitter: @mdalmau #dlbb
Text Encoding
Interests & Attitudes in Libraries
How would you rate the level of interest in text encoding by members of your library as whole? (n=112)

- Extremely Interested: 1%
- Very Interested: 9%
- Moderately Interested: 32%
- Slightly Interested: 35%
- Not at all Interested: 14%
- Not Applicable: 4%
- No Response: 5%

Twitter: @mdalmau #dlbb
How would you rate your library’s administrative support for text encoding projects today? (n=112)

- Extremely Supportive: 4%
- Very Supportive: 19%
- Moderately Supportive: 33%
- Slightly Supportive: 21%
- Not at all Supportive: 12%
- Not Applicable: 5%
- No Response: 6%
### Administrative Support (Q5) v. Library Overall Interest (Q6)

#### Q5 & Q6 Crosstabulation

<table>
<thead>
<tr>
<th>Q5</th>
<th>Extremely Interested</th>
<th>Moderately Interested</th>
<th>N/A</th>
<th>Not at all Interested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Supportive</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Moderately Supportive</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Not at all Supportive</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Slightly Supportive</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>Very Supportive</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Total: 6 1 36 4 16

#### Q5 & Q6 Crosstabulation (Detailed)

<table>
<thead>
<tr>
<th>Q5</th>
<th>Slightly Interested</th>
<th>Very Interested</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Supportive</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Moderately Supportive</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Not at all Supportive</td>
<td>13</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>Slightly Supportive</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Very Supportive</td>
<td>5</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

Total: 39 10 112

Twitter: @mdalmau #dlbb
In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today? (n=63)
In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today?

Coded Positive
In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today?

Coded Negative
In a few sentences, could you describe how you see the state of and attitudes toward text encoding in your library today?

Coded Neutral

Twitter: @mdalmau #dlbb
What’s Next?
You Ask.
Martin Mueller, Northwestern University, TEI-Nudge Proposal
http://cscdc.northwestern.edu/blog/?p=872
A New Part of Your Digital Humanities Toolkit

Tapas is the TEI Archiving, Publishing and Access Service for scholars and other creators of TEI data who need a place to publish their materials in different forms and ensure it remains accessible over time. Tapas is also for anyone interested in reading and exploring TEI data, and communicating with those that share that interest. The goal of Tapas is to provide TEI publishing and repository services at low cost to those who lack institutional resources: faculty, students, librarians, archivists, teachers, and anyone else with TEI data who wants to store, share, and publish it. Tapas seeks to achieve these goals in a collaborative, open, and community-driven way using open-source tools.

We are building Tapas now — contact us to get involved!

News

Tapas Panel at Rome TEI Conference, October 2013

November 01, 2013

As the Tapas project approaches its soft launch in early 2014, members of Tapas and the TEI board of directors held a panel presentation at the annual TEI conference in Rome on October 5, 2013 to report on progress and seek feedback from members of the TEI community.

(read more)

jTEI article on Tapas published

July 26, 2013

Those who would like to find out more about the development of the Tapas service will be interested in a recently published article in the Journal of the Text Encoding Initiative: “Tapas: Building a TEI Publishing and Repository Service” (http://jtei.revues.org/788).

(read more)
A TEI Project

Best Practices for TEI in Libraries

Author: TEI SIG on Libraries
Editors: Kevin Hawkins, Michelle Dalmau, and Syd Bauman
2011-10
Version 3.0 (October 2011)

Table of contents

1. Introduction
2. Relationship to TEI Title
3. General Recommendations
   3.1. Standards and Local Practice
   3.2. Transcription
      3.2.1. Punctuation
      3.2.2. Hyphenation
   3.3. Filenames
3.4. URIs
3.5. Textual Divisions
3.6. Page Breaks
3.7. Linking Between Encoded Text and Images of Source Documents
3.8. General Guidelines for Attribute Usage
   3.8.1. type
   3.8.2. n
   3.8.3. key and ref
   3.8.4. rend and rendition
   3.8.5. xml:lang
4. Structure of a TEI Document
   4.1. The TEI Header

Best Practices for TEI in Libraries
http://purl.org/TEI/teiinlibraries
Questions?
Thanks!