
This study seeks to analyze the feasibility of these projects that involve the Text Encoding Initiative Guidelines and standards for text encoding of scholarly material. Many libraries and other institutions are initiating digitization projects that involve text-markup. For this reason, such projects should be analyzed for feasibility and usefulness. Eighteen project managers and field professionals involved in TEI projects completed an online survey to assess the overall feasibility of their projects. Several factors were used in the assessment such as amount of funding, completion of project, satisfaction with the encoding, and user feedback.

Headings:

Project Management

Text Encoding Initiative

Information Systems -- Cost-Benefit Analysis

Feasibility

Electronic Data Archives -- Standards
PERPECTIVES ON TEXT ENCODING INITIATIVE PROJECTS FROM CURRENT DIGITAL LIBRARY MANAGERS

by James G. Dickie

A Master's paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Information Science.

Chapel Hill, North Carolina
April, 2008

Approved By:

_______________________________
Katherine M. Wisser
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Literature Review</td>
<td>4</td>
</tr>
<tr>
<td>Methodology</td>
<td>8</td>
</tr>
<tr>
<td>Results</td>
<td>10</td>
</tr>
<tr>
<td>Discussion of Results</td>
<td>21</td>
</tr>
<tr>
<td>Conclusion</td>
<td>27</td>
</tr>
<tr>
<td>Appendix</td>
<td>30</td>
</tr>
<tr>
<td>Bibliography</td>
<td>33</td>
</tr>
</tbody>
</table>
Introduction

Projects for digitizing manuscripts and historical artifacts for scholarly research are a current progressive move for archives and libraries. It is considered progressive since it is a move away from storing materials in print to a standardized encoded format of document storage and retrieval. In support of this concept, it has been said that digitization projects, “are now part and parcel of every day work in the humanities” (Hockey, 89). In addition to managing these projects, the text itself needs to be properly encoded to ensure proper access and analysis of the final product. Simply encoding documents facilitates analysis of scholarly text, but quality metadata and storage promote these features. One machine-readable standard was created for scholarly use by the Text Encoding Initiative and has since gathered attention in the field of digital humanities. Currently there are a wide variety of academic research institutions and digital libraries worldwide that use and support the TEI standard (Text Encoding Initiative).

The Text Encoding Initiative is a non-profit organization founded by librarians and archivists united with the plan to create metadata for scholarly text in digital formats (Giordano). In November 1987 the Association for Computers in the Humanities and the National Endowment for the Humanities sponsored a meeting at Vassar College to discuss issues of long term digital preservation for archives and libraries. Several institutions had complained of a lack of tools to develop scholarly text for research and analysis. Many were concerned about enterprise solutions for encoding
and preserving text that did not offer a consistent standard (Text Encoding Initiative).
The Text Encoding Initiative, formed after the Vassar College meeting in 1987, was
created to develop guidelines for encoding scholarly text. By 1994, the TEI Guidelines
(Then in its third edition, ‘P3’) were widely accepted and an international community of
scholars began editing and contributing to the guidelines.

Later in 2000 and 2001, the TEI Consortium was established as a community
that plans new ways for the TEI standard to assist in creating tools for scholarly
research. This established the Text Encoding Initiative as a not-for-profit organization
dedicated to making a user community for developing the TEI standard and spreading its
usage (Text Encoding Initiative). It was the Consortium that decided to move forward
and make the TEI Guidelines compatible with the eXtensible Markup Language (XML)
format. Since then, “TEI is internationally recognized as a critically important tool […]
for the long-term preservation of electronic data” (Text Encoding Initiative).

Within the articles discussing applications of the TEI standard, the majority
discuss the benefits of the TEI standard without supporting these theoretical benefits
with facts or opinions from administrators of digital storage and retrieval services. In
most cases, the administrators are the authors of articles and as such only give one
perspective on the use of the TEI standard use (Yeates, p.73). This project sought to
obtain opinions from professionals involved in TEI projects in order to fill this gap. In
this way it will complement the literature on individual projects by giving a broader
perspective. The
topics addressed in this study are why these projects are important to the institutions
that support them and how these projects are responsibly carried out.
This study seeks to create practical reasoning behind implementing tools for
digital scholarship. There is much to learn from current professionals about how agents
external to the digitization process – users, institutions, and funding sources – operate in
the real world environment. Smaller institutions unsure about going forward with a
digitization project may find it difficult to find literature that include practical opinions
on digitizing materials with the TEI standard. Proving TEI encoding projects to be
either beneficial or harmful in terms of financial feasibility and scholarly reward can
assist future digitization efforts.

Literature Review

Most of the literature available on the implementation of TEI focuses primarily
on case studies of projects and institutions implementing the TEI standard in the
digitization process. In addition, there are articles covering digitization of materials and
metadata creation standards. In both types of articles, the authors present problems in
the digitization process. These problems typically center on the quality of metadata,
project management, and institutional benefits that stem from digitization projects.

Greenstein in his article titled: “On digital Library Standards: From Yours and Mine to
Ours” writes, “these libraries [deploying digitization projects] face very real large-scale
investments and the need to make a commitment to a broad range of methodologies that
will evolve into complex operational services” (Greenstein, para. 11).

Digitization has yet to develop into a refined process. Text encoding projects
can present several problems for an institution. Deegan outlines two main constraints
for digitization projects: time and money (Deegan, 361). Kahn agrees with Deegan
about the core concerns for digital library projects but asserts that managing and
training staff are problems, as well (Kahn, 41). With the amount of investment going into digitization, it should be evaluated whether or not digital library projects produce results. The authors of these studies outlined only what could be potential desires for institutions rather than obtaining actual opinions and data from the field about the digitization of documents using the TEI standard.

One of the desired results for digitization projects should be proper description of original materials using current metadata standards. Such standards are outlined by the Text Encoding Initiative Guidelines, currently in the P5 format. It is not enough that materials be transferred from print to digital format; they must also have proper metadata applied to them for retrieval in a database. Baca asserts that such standards can be achieved through controlled vocabularies and careful selection of metadata standards (Baca, 52). It is hypothesized in Baca's study that information in descriptive formats can create “crosswalks” into other topics that readers initially did not search for. There would also be the “flexibility of digital information”: the ability for electronic text to be searched, have its contents counted and analyzed, and having all references properly tagged for cross-referential analysis (Hockey, 90). Whether the TEI standard offers the possibility to mark-up texts in such a way to provide intuitive “crosswalks” between knowledge concepts within several different texts remains a topic to be explored.

Providing metadata for digitized documents is a cornerstone of the guidelines behind the TEI standard. Greenstein has written several articles on the need for quality metadata in encoded humanities text. He finds that many institutions create their own “homegrown” solutions to metadata regardless of the presence of existing standards.
Whether this causes problems for scholars or not remains to be seen, but when institutions deviate from a standard, there are large theoretical conflicts. Following this line of thinking, Yeates discusses the need for libraries, archives, and museums to bind together to offer materials online and become one with the “global digital space based on the Internet and other digital networks” (Yeates, Sec. 2). If primary source materials are to be digitized and hosted for efficient online retrieval, information is required to be marked up in a standard format for users. Regardless, several institutions implement their own metadata schema, as brought up by Greenstein (Greenstein, para. 5). These studies do not necessarily link to projects using the TEI standard. Instead, they provide a survey of planned schema developed by a variety of users.

Several methods for capturing the essence of an object repeatedly and in a recognizable format exist in the TEI standard. Giordano explains the use of the header section as a method of encapsulating metadata for a TEI-standard document: the header “describes both bibliographic and non-bibliographic information and supports, in addition to the identification and retrieval of an encoded text, the machine analysis of encoded text” (Giordano, 391). A document using TEI serves as a traditional description of the encoded document in addition to having metadata within the document body. TEI provides features to mark up the meaning of the encoded manuscript, play, transcript, or other surrogate artifact. TEI makes the document as a whole, “intelligent because the tags concern meaning rather than display” (Nellhaus, 258). As an encoding standard, TEI qualifies as a tool for generating the “crosswalks” within documents.
With the trend of implementing “digital spaces” and the importance of preserving artifacts in a useful format, the positive claims to the TEI standard should be addressed on a practical level. One researcher voices concern: “Is the rush to digitize simply a reaction to the funding climate, or is there added value in creating digital instances of existing archival collections?” (Holz, 30). While digitization projects have been remarkably successful, they require a huge amount of planning and resources. For the Mark Twain project examined by Holz, several committees had to be outsourced in order to achieve a final product of a quality, accessible source of scholarly data on Mark Twain (Holz, 42). Text encoded in TEI was considered “ultimately superior” to their “analog alternatives” (Holz, 42). Not all institutions will be able to accomplish encoding, however, if these projects require the expertise, funds, and contacts that the Mark Twain Project (MTP), described as an “elaborate, complicated, and cooperative project” (Holz, 42). Such an objective question can not be answered through looking at one institution's success, but rather through a survey of several institutions.

Grasso et al. tested their digitization project to empirically research the cost-benefit analysis of digitization projects for libraries. To achieve this, they categorized costs for their Portland library digitization project into two phases: pre-digitization costs (procurement of materials, quality assurance, research, cataloging) and processing costs (scanning, storing data, encoding) (Grasso et al., p. 55). While a good addition to the research on library cost-benefits for digitization, Grasso et al. did not cover other institutions, instead focusing on their own unique instance (Grasso et al., p. 64). Their conclusions did not present much obvious detail other than “costs for digitization can vary widely” (Grasso et al., p. 62).
Currently few articles exist that discuss applications of the TEI standard, other than articles written by institutions about their own digitization programs. These case studies imply that digitizing materials provides benefits for document retrieval and storage. Nellhaus explains that the TEI standard is a set of guidelines developed by and for those concerned with digitization and preservation of documents (Nellhaus, 265). It follows that the institutions deploying digital library projects choose the TEI standard for its structure and design because it is tailored to scholarly use of encoded documents. At its core, the TEI standard offers tags for almost every genre of literature and historical manuscript (Nellhaus, 266). There is a theoretical potential in using the TEI standard, but not a fully analyzed basis for using it. This study seeks to fill the gap of research by finding opinions on using TEI in the field and the overall satisfaction with the standard.

**Methodology**

A survey is the best method for obtaining professional opinions about the feasibility and benefits of implementing TEI from a wide range of locations and institutions. Fourteen survey questions were generated that asked professionals working with the TEI standard their opinions on financial and merit-based achievements with projects using TEI and feasibility with accomplishing the projects they have been managing or are currently managing. The survey was broken into three parts: five open-ended and choice questions based on the parameters of the participant’s project; eight questions were framed as Likert-scale questions to assess perceptions of the experience with the project(s); and a final section with two open-ended questions about user and
project manager satisfaction. Each question except for the final two open-ended questions allowed the participant to provide an answer and to optionally add additional comments or other information.

The survey was administered online to reach the global community of TEI-based digitization projects. This eliminated the need for participants to reveal their identities and required the least amount of their time and effort to participate. An open-source software package based on the PHP language called “PHP Surveyor version 1.01” was used for the online survey. Participant anonymity was a primary concern taken into consideration for the study. Ethically, participant institutions and specific projects need to be anonymous to insure participant attitudes and positions on questions do not have consequences within the community. All participants remain anonymous and no identification as to who answered survey questions was received. An e-mail served as a consent form and was sent to the international TEI-L Discussion List, the official e-mail discussion list for professionals using TEI and those interested in TEI. Included in the e-mail was information about the survey, the study, and a link to the online survey.

The survey was posted online for a two month period. Two participants were unable to answer the questions in the given time due to business conflicts. In both cases participants were allowed to submit copies of their answers in Microsoft Word format. The answers to these two late participants were then added to the reported data and included in the data analysis. The identities of both of these late participants were removed.
A total of eighteen responses were received. Due to the evaluative and qualitative nature of the answers, traditional statistical tests were not used in the analysis. Instead, the intention has been to study the answers by comparing them against other (anonymous) participant’s answers. From this analysis, general trends and directions in project management attitudes and opinions for digitization projects involving the TEI standard were generated.

Results

The first section of the survey consisted of five questions that sought to identify the scope of participant's projects similar to the Grasso et al. cost-benefit analysis. Obtaining sufficient funding and the size of collections to be digitized are major variables for digitization projects. The questions in this section were intended to retrieve data about financial support and scope of the participant collection(s) involved in order to evaluate the range of projects using TEI currently in production.

For the first preliminary question, the highest amount received for a participant's project was 500,000 GBP or 991,415.97 USD (using conversion rate: '1 USD = 0.504108484 GBP'). The lowest reported amount was a project that received restricted funding: “$67,000 over three years.” On average, the fixed values that were given were between 500,000 and 900,000 USD. 27.8% (5 of 18) of participants only provided partial information, using the free text field to give a relative rather than a fixed amount or to provide an explanation rather than a straight-forward response. This is most likely due to participants wishing not to divulge information about their funding. One reason
given in the open text field for this kind of response was: “We have a core TEI infrastructure which supports a large number of projects. Funding was partly a matter of software acquisition, but even more so the commitment of ongoing staff time over many years, so I can't come up with an initial cost.” Another reason for not reporting a fixed amount was “internal Funding for experimenting purposes.” Those not reporting a fixed amount seemed to indicate that the digitization program is in an experimental phase with limited funding. An additional 27.8% of participants (5 of 18) left the value field blank or reported “N/A”, most likely due to the same reason as those who gave no fixed values for funding.

Table 1: Participants with Ongoing Funding for their Projects (n = 18)

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the funding for your project(s) ongoing?</td>
<td>6 (33.3%)</td>
<td>3 (16.7%)</td>
<td>9 (50.0%)</td>
</tr>
</tbody>
</table>

Examples of Comments:
“Project is complete”
“planned, but not yet decided on. Decision is due December”
“It runs out at the end of 2009”
“The internal departmental funds are ongoing, but there isn't an endless supply for this project.”
“Indirectly. Our initial project has been completed. However, we have included text encoding as a component of several other completed and ongoing projects.”

As funding is a primary problem found in the literature on cost-benefit analyses of implementing digitization, one of the preliminary questions participants were asked
was whether their funding is ongoing or limited (see Table 1). 33.3% (6 of 18) of participants answered with a direct “Yes” and 16.7% (3 of 18) of participants answered with a direct “No.” The remaining 50.0% (9 of 18) of participants gave more detailed answers with one participant leaving the field blank (no answer). For seven detailed answers, the participants indicated they were unsure due to postponed funding decisions and limited funding reaching its end. 11.1% of the projects (2 of 18) were completed and had their funding terminated at the time of the survey. On the surface this suggests that the majority of participant collections are ongoing projects with no clear termination.

Table 2: Projects with TEI as a central part of the collection(s) (n = 18)

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are TEI documents a central part of your collection(s)?</td>
<td>11 (61.1%)</td>
<td>7 (38.9%)</td>
</tr>
</tbody>
</table>

The prominence of TEI encoded text in a collection was asked in question three. If a TEI project is successful in encoding and user support, it would likely follow that TEI documents are a focus of a collection. The results demonstrate that TEI documents are a central part of a digital project with a majority of participants (see Table 2). A majority of participants have TEI documents as a central part of their collection, with just over a third reporting that TEI documents only serve a portion of their collection. Two of these participants expanded upon why TEI documents are not a central part of the collection. One referenced “page images” as being “equally important”, most likely meaning that their online presence offers both encoded text as well as JPEG or TIFF
images of documents. Users of these participant projects may only wish to view an image of a document and not perform any textual analysis. In this case, providing scholarly mark-up for a document would be unnecessary. The other participant explained, “The overwhelming majority of our TEI documents are based on page images rather than on full encoding.” For this project, documents encoded with the TEI standard are therefore sometimes not fully used or they are used equally along with images of pages, audio text, and other multimedia.

Table 3: Size of Participant Collections (n = 18)

<table>
<thead>
<tr>
<th>Collection Size</th>
<th>Number Participants</th>
<th>Percentage Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small (~100 documents in a single collection)</td>
<td>9</td>
<td>50.0%</td>
</tr>
<tr>
<td>Medium (~1000 or more documents spanning more than just one genre/department/era)</td>
<td>6</td>
<td>33.3%</td>
</tr>
<tr>
<td>Large (Several thousand documents spanning multiple departments, projects, and/or eras)</td>
<td>3</td>
<td>16.7%</td>
</tr>
</tbody>
</table>

For the question referring to the size of the participant's collection, the results showed that most projects are either currently starting up or are small in scope (see Table 3). The largest group of participant collections fit the “Small (~100 documents in a single collection)” category, which comprises half of participants. The next largest category was “Medium (~1000 or more documents spanning more than just one
genre/department/era)” with a third describing themselves as such. A small group - 16.7% of participants (3 of 18) - reported having a “Large (several thousand documents spanning multiple departments, projects, and/or eras)” project size.

**Table 4: Projects in completion (n = 18)**

<table>
<thead>
<tr>
<th>Yes</th>
<th>Completed in a Scheduled Time (Almost Complete)</th>
<th>Continual (Continue until Funding Stops)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 (16.7%)</td>
<td>8 (44.4%)</td>
<td>7 (38.9%)</td>
</tr>
</tbody>
</table>

When asked about whether the project or projects is completed, only a small portion of participants gave a definite “yes,” with one reporting that the process is only “partially completed.” Due to the range of answers given for estimated time until completion, two large groups emerged from the participant data (see Table 4). One group (44.4%, 8 of 18) of participants answered with relative figures of how their projects will be completed either in a scheduled time or an indefinite time. Answers for estimated time until completion included: “6 months”, “one to two years”, “Spring”, “December 2008”, and the year “2010.” Another large group (38.9%, 4 of 18) of participants expanded on their answer by indicating that as long as funding continues, they will do their best to continue their project(s) indefinitely.

The preliminary questions revealed demographic factors of participant projects. Looking at the data, most participants work with small collection sizes, have relatively large budgets with restrictions, and have ongoing projects that will continue on into the future. The next seven questions were evaluative in nature and were intended to retrieve
the participant's relative opinions on certain topics in order to support the data collected in the preliminary questions. These topics centered on the principal concerns of most TEI projects as outlined by the literature: document encoding, funding, and public recognition. Participants were given an opinion statement with five scalar options: “Strongly Agree”, “Disagree”, “Neutral”, “Agree”, and “Strongly Disagree.” For each question, an optional text field was available for participants to explain or expand upon their opinion.

Table 5: Document Encoding Satisfaction (n = 18)

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The materials for your collection have been encoded and preserved towards your own initial goals.</td>
<td>7 (38.9%)</td>
<td>10 (55.5%)</td>
<td>1 (5.6%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Having the materials encoded with the TEI standard is better than relying solely on a printed version.</td>
<td>16 (88.9%)</td>
<td>2 (11.1%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Documents encoded with the TEI standard are used more than their printed versions.</td>
<td>5 (27.8%)</td>
<td>5 (27.8%)</td>
<td>6 (33.3%)</td>
<td>1 (5.6%)</td>
<td>0 (0.0%)</td>
</tr>
</tbody>
</table>

When addressing about the satisfaction with the extent of the encoding of materials text, 38.9% (7 of 18) answered “Strongly Agree,” 56% (10 of 18) answered “Agree,” and one answered “Neutral” (see Table 5). For the “Neutral” answer, this explanation was given:

“The level of encoding is part of the experiment and the goal of the project is to meet the needs and goals of the faculty invested in the project. Understanding their needs is one of the activities of the project. While that
sounds like an easy thing to do, we've found that the lack of a common foundational understanding of our different disciplines (Digital Library and 17th century History) has caused us to invest a lot of time in this process. We're still uncertain what level of encoding is the best initial investment to serve the faculty's needs.”

In this case, having an encoding format that supports the content being digitized is of comfort to the participant. The data, however, shows that the overwhelming majority of participants agree with using TEI as the standard for their project.

A similar consensus occurred for the next question. Participants agreed heavily that digital versions of documents were as good if not better than their print counterparts, with 100% positive opinion (see Table 5). Specifically, 88.9% (16 of 18) participants answered with “Strongly Agree”, while 11.1% (2 of 18) answered “Agree.” One participant who answered “Agree” explained: “...The printed version is not full-text searchable. The ability to align structure among the different books is the key to this project....” Another expanded reason was for preservation: “For this particular project (transcripts of taped interviews), yes, because the original materials are A/V materials which may not withstand time.”

It appears that users might not share the same approval of digital documents over print documents (see Table 5). The question regarding whether digitally encoded versions of documents were used more than their print counterparts had both positive and negative results. Just over a half of the participants (approximately 55.5%) agreed that users take advantage of digital material over print material. One third (33.3%, 6 of 18) of participants, however, answered “Neutral”, with several explaining that the matter is something of minimal concern for now. One participant in this group
countered with: “digital will replace print” in the future. The one participant that answered “Disagree” did not provide an explanation for their opinion.

Table 6: Opinions on Funding for TEI Standard Projects (n = 18)

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting funding for your most recent project has been problematic.</td>
<td>3 (16.7%)</td>
<td>2 (11.1%)</td>
<td>6 (33.3%)</td>
<td>6 (33.3%)</td>
<td>1 (5.6%)</td>
</tr>
<tr>
<td>There was a suitable amount of staff present to accomplish the tasks set forth by your most recent project.</td>
<td>3 (16.7%)</td>
<td>7 (38.9%)</td>
<td>2 (11.1%)</td>
<td>5 (27.8%)</td>
<td>1 (5.6%)</td>
</tr>
</tbody>
</table>

Participants were across the scale when asked if funding was an obstacle for their projects. Just over a fourth of participants answered that funding is an obstacle (see Table 6). Some gave negative remarks, such as: funding is a problem “unless you have very good friends in high places.” A third of participants answered with “Neutral”, with one such participant explaining that “the project is an experiment”, a “collaboration” among departments that receives contributions. An equal third of participants answered with “Disagree,” with one participant remarking that their funding has been due to “good fortune.” The one participant who answered “Strongly Disagree” did not provide an explanation for this opinion.

Similarly, answers for whether the participant projects had enough staff were distributed across the scale. Over a half of participants felt their staffing was adequate, with a total of 55.6% (10 of 18) agreeing (see Table 6). One participant explained that they were provided with both XML encoders and researchers in philology. 11.1% (2 of 18) of participants answered with “Neutral.” 27.8% (5 of 18) of participants answered
with “Disagree” and one answered “Strongly Disagree.” From this data, it can be assumed that some participants have better luck in staffing, while others may be forced to train newcomers or outsource work.

Table 7: Opinions on Public Recognition from TEI Projects (n = 18)

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>There has been more academic recognition for your institution due to projects involving the TEI standard.</td>
<td>3 (16.7%)</td>
<td>13 (72.2%)</td>
<td>1 (5.6%)</td>
<td>1 (5.6%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Your institution has received at least some material gain (money, funding grants, additions to the faculty) as a result of the project(s) involving the TEI standard.</td>
<td>4 (22.2%)</td>
<td>8 (44.4%)</td>
<td>2 (11.1%)</td>
<td>3 (16.7%)</td>
<td>1 (5.6%)</td>
</tr>
</tbody>
</table>

Recognition is an important factor for digital projects since it can provide academic honors for the players involved and ease financial barriers through monetary gain. Answers for whether there is academic recognition for the participant's institution due to projects involving the TEI standard were concentrated in “Strongly Agree” and “Agree” (see Table 7). 16.7% (3 of 18) of participants answered with “Strongly Agree” and 72% (13 of 18) answered “Agree.” One participant answering “Agree” commented that some academics feel “‘online' and 'academic' are contradictions”, but that this opinion is rapidly going away. The participants answering “Neutral” (1 of 18) and “Disagree” (1 of 18) provided similar statements saying that they expect to never get any positive recognition from the project.
The positive response for academic recognition among participants does not correlate with recognition from donors and fund givers in every case. Question twelve of the survey dove into whether the participant institution has received at least some material gain (money, funding grants, additions to the faculty) as a result of project(s) involving the TEI standard. A total of twelve of the eighteen participants (66.7%) feel that this is true (see Table 7). 22.2% (4 of 18) of participants answered with “Strongly Agree” and 44.4% (8 of 18) of participants answered “Agree.” Two comments illustrate participant's experience and positive attitude: stating that several research grants have resulted from the project and that receiving grant money has become easier. 11.1% (2 of 18) of participants answered with “Neutral”, but did not provide any explanation for the reasoning behind their opinion. The remaining participants answered with “Disagree”, with one participant explaining that their goals have been with the “intellectual significance” of the materials instead of possible monetary gain. The participant that answered with “Strongly Disagree” provided no explanation.

The two remaining questions labeled “Open Questions” were intended to allow the participants a space to briefly expand on their opinions on user feedback and overall satisfaction with their project(s). User feedback is an important issue as the TEI standard is implemented in projects delivering an end product (encoded document) to a user (scholar). If users are not satisfied with what is being produced, then it may be a sign that the TEI standard is not fulfilling participant’s project specifications. The first open question regarding user feedback indicated that the majority of participants perceive user feedback as positive. 50.0% (9 of 18) participants answered that they either do not track user responses or they do not track user responses currently. One
such participant explained that while they do not track user responses, they still
informally receive overwhelmingly positive compliments about their digital library
materials. 33.3% (6 of 18) of participants answered that they do track user responses,
and that these responses were all positive. One such participant indicated that they
receive “no negative feedback” on their project.

The last question asked whether participants felt that TEI projects are considered
to be feasible and worth the effort put into them. Answers were mostly positive. 50.0%
(9 of 18) of participants answered similarly that projects are indeed feasible. For this
group of participants, using the TEI standard provides a good “payoff”, even though
using the standard may be “time consuming” or require “proper staff” to complete.
27.8% (5 of 18) of participants were hesitant to claim that the TEI standard can be
attributed to project feasibility. Most agreed that feasibility depends upon the project in
question, with one participant giving the following elegant explanation:

“That depends entirely on the project in question. I've seen excellent
ones and lousy ones and all points in between. Using TEI will not in itself
salvage an ill-conceived project. But I do think that using TEI will enhance a
well-conceived one. In feasibility terms, where TEI really helps is that it
provides a tested encoding methodology that is well-documented but not
inflexible, and access to an online community of experienced and imaginative
text-encoders who are generally keen to share their own expertise gratis.”

Some of the participants (16.7% - 3 of 18) were somewhat undecided on the issue at
hand. Another gave a somewhat off-hand answer:

“In our institution, we have not yet built up a lot of experience in doing
TEI projects within the Digital Library – our concern is the expense of providing
quality encoded-text projects. We are trying to balance image access with text
access. The standard is very reliable as far as maintaining quality data over time.
We will be trying several pilots over the next few years.”
While this does not clearly indicate that using the TEI accounts for the feasibility of their projects, the participant did indicate that future projects will use the standard.

**Discussion of Results**

Taking the results as representative of the participant population of digitization projects, Deegan and Kahn's arguments that money and time are primary issues are confirmed by these participants. More than half (55.6%) of the participant responses for the first question (asking for approximate values for funding) were left either blank or provided only relative values. This is probably due to restrictions attached to the funding for participant projects. Actual values ranged all over, indicating that projects are receiving a wide array of funds. In addition, one third of participants felt that their projects are considered experiments or fledgling projects that have limited scope and therefore a limited budget. That participant answers to funding range dramatically can possibly mean that the amount of money their projects receive depends on the particular project and the institution involved.

Related to these financial restrictions, participants seem to have varying priorities for project completion. Some participants answered that their projects were in completion and others answered with relative end dates for their projects. Added together, this means that a half of participants have projects with definite end dates, and the other half have possible ongoing projects, or at least funding that has no definite termination. Project management for these digital projects is therefore divided between planned stages for projects and projects that are intended to go on for indefinite periods. Grasso et al. mention that some institutions have to spend a great deal of time and
resources on cataloging different and wide-ranging materials from several different departments (Grasso et al., 56). When this happens, project completion can be pushed back significantly. Funds would then have to be acquired, making money a larger issue. This assumption is supported by an answer that indicated that their project is “a collaboration among departments.”

A pattern that emerges between money and completion time may be that larger collections or larger digitization projects tend to present stumbling blocks for projects. Based on participant’s answers for the size, completion, and budget of their projects, the scope of the average participant project is a small collection of documents without a planned end date, and with a planned budget, regardless of funding limitations. Half of participants agreed that their collection should be counted as small, or less than one hundred digitized documents. This may indicate that the majority of participant collections are starting, since most of those answering 'Small' also gave future dates for the completion of their project. Those answering 'Medium' or 'Large' also tended to have larger budgets and federated projects with multiple departments.

A small collection of encoded text is the digital project scope that is preferable to participants. This is explained by the scalar question on whether funding is an obstacle. There was no consensus or pattern among participants for funding obstacles. Participant projects that are in the 'Neutral' field for funding obstacles appear to be experimental or starting out. This is also attributable to those participants having trouble with funding. Therefore there may be a pattern among the participants of implementing small start-up collections with complex contracts to their funding over larger projects. If
a majority of participant projects are starting up, it would explain the high number of small collections and ongoing funding for participant projects.

That the participants continue their projects regardless of staffing and financial problems may be due to participant idealism. All participants answered positively for the question regarding print versus digital documents. This should be expected, as all participants will be advocates for digitization in their institutions. Some interesting reasons were brought up, such as “aligning structure among books” being something possible in the digital world and not the print world. As Greenstein mentions, however, it is important to make sure metadata is properly encoded into each document (Greenstein, para. 6). For this reason, it was important to ask participants their opinion on whether the TEI standard allowed them to encode documents to their satisfaction.

Nearly all participants felt that the encoding of documents was to their satisfaction. A majority of participants answered that the staff at hand were enough to finish their project in time. The high number of staff for most participant projects may be a result of TEI projects being on university campuses, where student work abounds in libraries. However, few participants answered “Strongly Agree” for this question, most likely because even when supplied with staff, they are usually students that require training and therefore take time away from the project. Outside of academia, hiring staff specialized in the TEI standard may be problematic, possibly making up the negative side of participant answers for this question.

Over half of the participants answered that their TEI documents are a central part of their collection. Almost two-thirds feel that other materials are important for their collection. Others in the literature of digitization have argued for the benefits of
digitized documents, but they did not consider the weight of other electronic resources.

In addition, an image of a document may count as more important to a user than a marked-up version of the text. For example, this may be true for medieval documents, where the image of the pages has artistic or structural value to a related scholar. In addition, simplicity may be the best method for some projects, as Holz mentions in his article (Holz, 32). Just over a half of participants consider digital documents to be used more than their print counterparts. Despite one participant who disagreed, the remaining participants feel neutral about the use of digital over print. This may be due to lack of user feedback or simply that users do not actively recognize the importance of TEI over using other resources offered by the participant institution.

The clear reward for participants is academic recognition and financial gain for digital library projects. A clear majority of participants felt that they receive academic recognition and/or honors for their work, a group which included small projects as well as large projects. With the exception of two participants, all participants feel they are receiving funding, grants, and attention in their field for digitizing documents. There is therefore a sense of honor for working on digitization, at least within the world of the participant institutions. This kind of recognition from scholars and faculty does not always lead to financial gain. A smaller majority (66.7%) feel that funding grants and other forms of material gain are a result of their digitization efforts. Just as agreement among participants about the utility of digital versus print documents may not equate with user satisfaction, academic recognition among participants for digital collections may not equate to attention among the wider community and third parties for donations and other material gain.
Regardless of the amount of attention, financial, academic or otherwise, there is consistent positive feedback for participant projects. This question revealed that even when user feedback is not tracked, it is most likely positive. This can be due to participant projects only being contacted by scholars and faculty who are knowledgeable of the resources provided by the participant’s library or institution. Participant TEI collections are not always the central collection and marked-up documents are not always used more often than digital surrogates. Therefore, there may be a population of traditional scholars who do not give feedback, either because they do not wish to take the time or are not familiar with the digital library initiative within their institution. Those users that do give feedback have positive comments, which may mean that in time digital initiatives will gain more attention. Holz’s analysis of the Mark Twain project supports this concept (Holz, 35).

Data collected in the survey supports the concept that digitization using the TEI standard is feasible for participant institutions. Participants had varied answers for feasibility with half feeling that such projects are currently feasible. Among the answers given supporting the use of the TEI standard were several technical reasons. For example, with XML, documents are encoded in a permanent yet flexible mark-up format. Also, new technologies such as XSLT and browser support for XML were identified by one participant as reasons for using TEI as a mark-up standard. However, this can be applied to many other mark-up standards, such as TEXTCLASS and customized mark-up DTDs and W3 Schema. The flexibility and customization that the TEI standard allows seems to be the reasons for its success.
Participants giving less positive responses to this question mention alternatives to using the TEI standard. One participant gave an alternative to text-encoding itself, saying: “TEI is one of many solutions to a project and is certainly preferred when text-encoding is required, but a digital project may be successful without text encoding.” As mentioned before, some projects may involve only page images of text or artifacts. This means that some participant digitization projects wish to avoid the time and effort of encoding documents in TEI for a simpler route. They may involve only TIFF images and OCR data, for example. Alternatives such as this may be necessary for those participants implementing small projects with limited funding. A participant made the distinction that “page-image” projects are feasible with using the TEI standard, but “full-encoding” for text can be expensive. This participant explained that: “one must choose projects according to their potential impact, and not assume that a fully-encoded edition is the most cost-effective choice for any text.” It can not be assessed from this survey alone whether an image archive is as effective for end-users as a fully encoded archive of documents and should be a focus of future research.

Some responses indicated that while projects involving encoding with the TEI standard are feasible, these successful projects come with requirements for success. The central requirements include experienced staff and the ability to customize the mark-up for a project. Interestingly, the answers for having proper staff were spread out, but almost half of participants answered that having “knowledgeable staff” or at least a “supportive community” for using the TEI standard is necessary. When discussing knowledge of TEI, participants mentioned being able to plan the encoding strategy in advance, as in a “tested encoding methodology.” Being able to flexibly integrate TEI
with “XML/XSLT/Unicode and ... browser support” is one example of using a customized TEI encoding strategy. A successful project, therefore, is dependent on adequate knowledge of the tools being used to progressively encode each unique document.

CONCLUSION

Data collected from the survey indicate that the participants remain positive about the outlook of the TEI standard in digitization projects. Despite voiced difficulties in the participant’s current and past projects, participants of the study remain faithful to the use of TEI to digitize scholarly material. This does not indicate that the TEI standard is widely accepted in participant institutions or among the participant community. Indeed, the road to wide-spread implementation of TEI may be a long one, with more digital projects implementing the TEI standard forcing acceptance of XML and TEI in browsers, enterprise databases, and XML editing software. As the use of TEI grows, so will specialized support staff for using the TEI standard in XML. This should be a positive influence for institutions and professionals preparing digitization projects.

In time, users of marked-up documents may appreciate the benefits of having such a standard in place as TEI. It is not necessary for those using the documents of the participant projects to have knowledge of the TEI standard. Rather, they enjoy the material more than the way it is encoded or presented, as one participant pointed out. Future discussion should focus on user perspectives of encoded digital library material and whether they actually utilize the tools encoded mark-up provides. Thus the popularity of use for the TEI standard will come from users enjoying the added benefits
of marked-up text and not from the popularity of the standard itself. Participants most likely work hard to keep the encoding process invisible to the end user.

As indicated by the results of this study, participants have received positive user feedback from their digitization efforts. Participants have also received a significant amount of recognition from the academic community. Successfully completing a TEI project has led to grant proposals and additional funding. That a small group of participant institutions do not fund their digitization projects continually means that a majority of participant institutions feel that the digital library is worth funding. These factors may denote that TEI projects play a positive role in storage and dissemination of legacy material and will likely continue to do so. Those considering being a part of the digitization and TEI community have chosen a good time to join.

The main role of a digitization project is providing material in a pragmatic, usable, and accessible format. Secondary to this primary role of digital libraries are the issues examined in this paper. The Text Encoding Initiative Guidelines may in time be the final judge as to what metadata are recorded for digitized documents. As more institutions rely on digital surrogates to replace permanent print collections, having a standardized, usable and accurate format to store legacy material will be important. Already the participants of this study have devoted their allocated resources and time into implementing metadata standards according to the TEI Guidelines. Through assessing their project scope, opinions, and remarks on user feedback and feasibility, it is more clear that digital collection projects are feasible in the real world and may provide material benefits and benefits to the end-user when completed. For those initiating a digital collection project, having enough funding, a willing institution that
will continually fund a project, staff familiar with TEI standards and XML, and technology supporting XML are considerations to take in. From a real world perspective, digitization implementing the TEI standards is a wise choice for institutions.
Appendix

Survey Questions (Including participant instructions):

Thank you for participating in this survey. This survey measures your professional viewpoint of the status of your TEI digital project. It is assumed that you are a project manager or supervisor hosting digitally transcribed and encoded documents, some or all of which use the TEI format. The questions presented here are partly yes and no and scaled questions. They are organized into preliminary questions about the funding and parameters of your project and specific questions as to the benefits resulting from the TEI project.

*Project scope questions:*

1. How much (approximately) was the initial funding of your project?

2. Is the funding for your project(s) on-going?

3. Are the TEI documents the central part of your collection or are they only a component of the collection?

4. What is the relative size of your digital collection that uses the TEI standard (choose below)?
   - Small (~100 documents in a single collection)
   - Medium (1000 or more documents or a collection spanning more than just want genre/department/era)
   - Large (Several thousand documents spanning multiple departments, projects, and/or eras)

5. Is the project in completion? If it is not completed, please provide an estimate for the time until completion.

*Evaluative Questions.*

*Please rate the degree to which you agree or disagree with the following statements. Feel free to elaborate on any of these questions by typing in the text box provided for the question:*

6. The materials for your collection have been encoded and preserved towards your own initial goals.

   Strongly disagree    Disagree    Neutral    Agree    Strongly agree
7. Having the materials encoded with the TEI standard is better than relying solely on a printed version.

Strongly disagree    Disagree    Neutral    Agree    Strongly agree

8. Documents encoded with the TEI standard are used more than their printed versions.

Strongly disagree    Disagree    Neutral    Agree    Strongly agree

9. Getting funding for your most recent project has been problematic.

Strongly disagree    Disagree    Neutral    Agree    Strongly agree

10. There was a suitable amount of staff present to accomplish the tasks set forth by your most recent project.

Strongly disagree    Disagree    Neutral    Agree    Strongly agree

11. There has been more academic recognition for your institution due to projects involving the TEI standard.

Strongly disagree    Disagree    Neutral    Agree    Strongly agree

12. Your institution has received at least some material gain (money, funding grants, additions to the faculty) as a result of the project(s) involving the TEI standard.

Strongly disagree    Disagree    Neutral    Agree    Strongly agree

Open questions:

13. Are you tracking user behavior and feedback for use of documents encoded with the TEI standard? If so, has this user feedback been positive or negative, generally?
14. What is your overall impression of projects involving encoding documents with the TEI standard in terms of feasibility?

*Thank you for your time!*
Bibliography


*British Library: Research and Development Department. BLRD Report, 6075, 117-128.*


