Journal of the Text Encoding Initiative

Issue 7 | November 2014
Reaching out, Opting in

# Learning the TEI in a Digital Environment

Stella Dee

OpenEdition Journals

**Electronic version**
URL: http://journals.openedition.org/jtei/968
DOI: 10.4000/jtei.968
ISSN: 2162-5603

**Publisher**
TEI Consortium

**Electronic reference**
Stella Dee, « Learning the TEI in a Digital Environment », *Journal of the Text Encoding Initiative* [Online], Issue 7 | November 2014, Online since 01 January 2014, connection on 19 April 2019. URL : http://journals.openedition.org/jtei/968 ; DOI : 10.4000/jtei.968

**Journal of the Text Encoding Initiative**

**Issue 7 | 2014**
**Reaching out, Opting in**

# Learning the TEI in a Digital Environment

**Stella Dee**

**Publisher**
TEI Consortium

**Electronic version**
URL: http://jtei.revues.org/968
ISSN: 2162-5603

**Electronic reference**

# *Learning the TEI in a Digital Environment*

Stella Dee

## 1. Introduction

1 An exchange on the TEI electronic mailing list sparked the research published in this article, when a community expert wrote that given "a graduate student in English who has heard about TEI and wants to dip her toes into it because she thinks it may be a better way of putting on the Web some 17th century poems ... Where does she go for help? Where in the TEI universe is the level of ubiquitous Grade I support?" (Martin Mueller, pers. comm, Jan. 18, 2013).[1] Having myself been very similar to this hypothetical student as an undergraduate, I researched available resources and conducted a survey designed to inform an answer to this question.

2 I was guided by previous similar work, including that of Burghart and Rehbein (2012), who conducted a survey investigating participation in the TEI community among practitioners of manuscript encoding. However, the present survey differed in its target audience, available to any interested party as opposed to a particular specialist community. Moreover, it focused on online instructional resources, in contrast to pedagogical resources more generally. The survey was designed in the context of recent work grappling with the place of pedagogical thinking in the digital humanities as a whole (Hirsch 2012; Mahony and Pierazzo 2012; Chickering and Ehrmann 1996), including how the TEI community might contribute to thinking surrounding teaching, learning, and new models of apprenticeship in the digital environment.

3 Respondents were generous with the depth of their feedback. And although the survey was focused on pedagogical resources, past and potential, the clearest message contained in the feedback voiced a slightly different need. In aggregate, the survey results communicated above all that both expert and novice users desired open TEI-encoded data repositories in which to publish their files. They felt a need for such repositories as much

as, if not more than, digital resources for learning the TEI. Of course, text repositories and pedagogical resources are not mutually exclusive; the survey confirmed that even now, many people learn the TEI inductively from already-encoded texts. However, the fact that so many people prioritized text repositories over didactic tutorials in a survey entitled "Learning the TEI—"which would presumably bias respondents in favor of the opposite order of priorities—suggests that the TEI Consortium should seriously consider how to meet this demand. I suggest some possible ways forward in the conclusion, but the question would benefit from wider community discussion.

## 2. Prior Resources

4 Although the majority of eventual TEI users receive initial instruction in person (Elena Pierazzo, pers. comm.), a number of digital resources currently provide support for scholars and researchers hoping to come to a better understanding of the TEI Guidelines and their usage, some of which were specifically addressed by the survey. These included static webpages of instructional material published by institutions such as Brown University[2] and the University of Oxford[3] as well as more comprehensive websites, including *TEI by Example* (Van den Branden, Terras, and Vanhoutte 2010), the *TEI Wiki*, and Teach Yourself TEI. The survey asked participants whether they were familiar with these resources, and if so, how useful they found them. There were no questions that explicitly addressed the resources prepared by academically affiliated individuals (Almas 2012; Mandell 2013; Chesley 2012; Spiro 2010),[4] and no questions concerning the customized resources that target specialist communities (Halsell 2013; Roueché and Flanders et al. 2014), leading to a possible bias in some of the results.

## 3. "Learning the TEI" Survey

5 The survey[5] was titled "Learning the TEI." Created using SurveyMonkey software, and live for approximately a month between late June and late August 2013, it collected 95 anonymous responses in total. It was offered only in English. As a consequence, those researchers without professional proficiency in English were probably excluded. The survey asked a combination of open and closed questions, targeting experts as well as novices in the field. Designed for gathering insight to help improve task-based online resources for working with the TEI, I wrote the questions to solicit "thickly descriptive" feedback wherever possible (Geertz 1973). The survey included questions focused on acquiring descriptions of the TEI to assess user understanding, technical competency of respondents, and the kinds of textual work currently being undertaken. I hoped that these questions might reach survey respondents who had never heard of the TEI but who were working intensively with texts. Because of this, nearly all questions were optional, so that those who had never heard of the technologies and communities mentioned were still able to respond. As a consequence, most of the survey results are not of statistical significance, and I would caution against understanding the results in a quantitative light.

6 This section is divided into four subsections, broadly categorized according to the different subjects about which the survey addressed. Section 3.1 provides an overview of the respondents—how they were reached, who they are, and how long they have been working in the field. Section 3.2 addresses the goals and aims respondents gave as reasons for using the TEI; section 3.3 summarizes their feedback on the available educational

resources. Section 3.4 captures responses concerning the future priorities of the TEI Consortium, opinions which inform the conclusions that follow.

## 3.1 Respondents

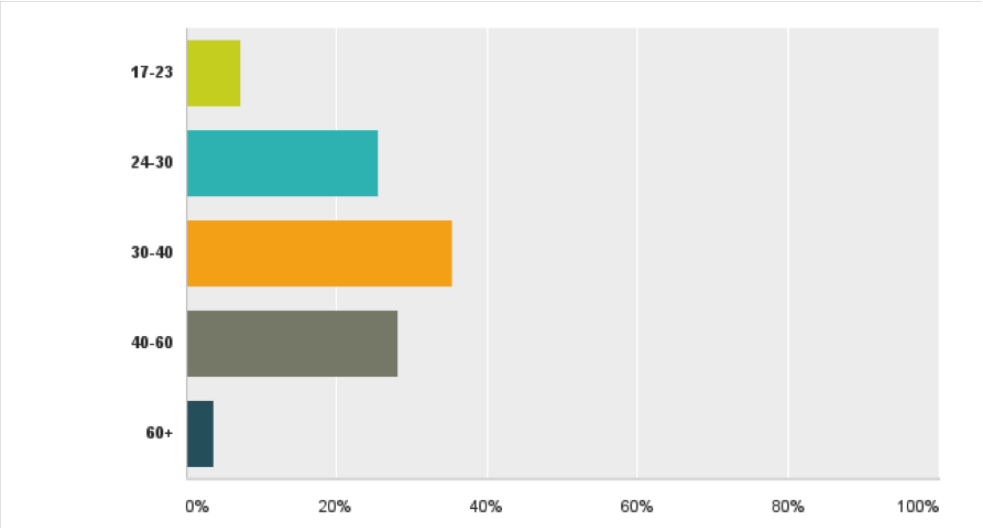
### 3.1.1 Dissemination

7 The survey was disseminated through several digital humanities mailing lists, including the TEI-L mailing list and the digital humanities mailing list run by the University of North Carolina at Chapel Hill. It was also disseminated over Facebook and Twitter. Nearly all of the questions were optional, in order to enable students and researchers who had never before heard of the TEI Guidelines or text encoding to participate, although the final number of inexperienced participants was not as high as I had hoped. Of the total responses gathered, 51 responded to the majority of questions.

### 3.1.2 Demographics

8 Demographic questions concerning the age, field, level of education, work, position, country, language, gender, and depth of engagement with the TEI community enabled an exploration of how research goals vary by individual or how they might cluster by context. The ages of the participants are summarized in figure 1.

**Figure 1: Age distribution of respondents (82 total)**




9 Demographically speaking, of 82 respondents, 45 were men and 37 were women, with the majority falling between the ages of 30 and 40. English was the language spoken "most often" by respondents (50%, 38 total), closely followed by German (38%, 29 total). However, German and English were roughly equal in terms of language spoken "most fluently" by respondents, with 32 respondents speaking English most fluently and 30 speaking German, followed by five Italian, four French, one Chinese, one Greek, one Spanish, and one Japanese. Of the people who chose to list their location, all worked in North America, Europe, and New Zealand.

10 Participants conducted academic or research activities in a broad range of historic and modern languages, with diverse alphabets and temporal frames of use. In addition to the

languages spoken most fluently by respondents themselves, research languages included Latin, Maori, Coptic, Slovene, Old Icelandic, Arabic, Persian, Hebrew, Swedish, Finnish, Indonesian, Nxaʔamxcín, Hul'q'umi'num', Danish, Norwegian, Albanic, Occitan, Catalan, and Ge'ez. Seventy-four listed other programming languages with which they were comfortable working; 66 people responded that they had heard of the TEI prior to beginning the survey, while four said that they had not. The majority of respondents were generally comfortable with basic web technologies and commercial software.

Figure 2: Technological working proficiency (59 respondents)

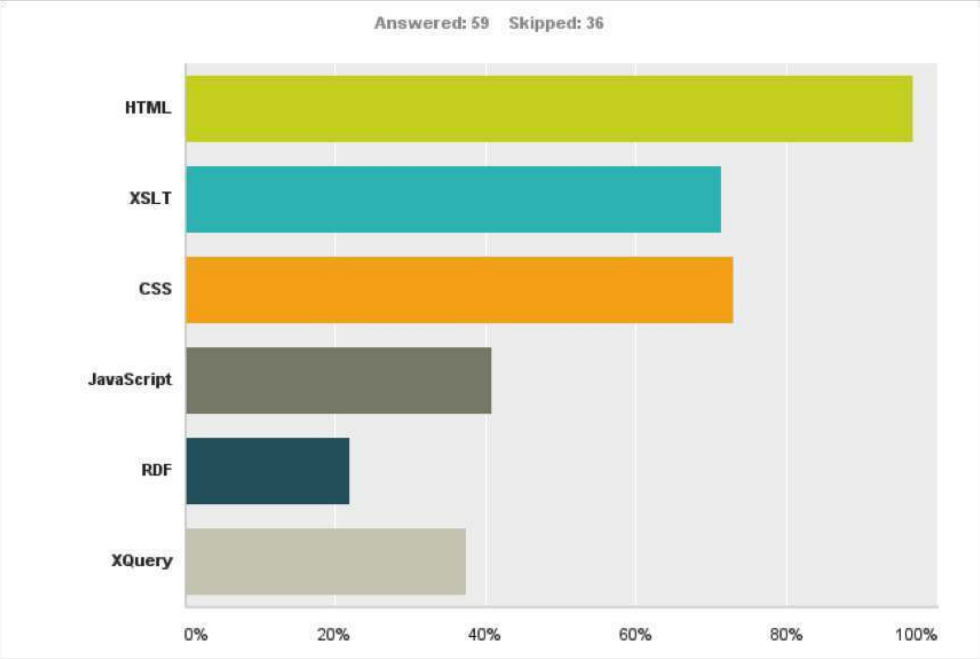
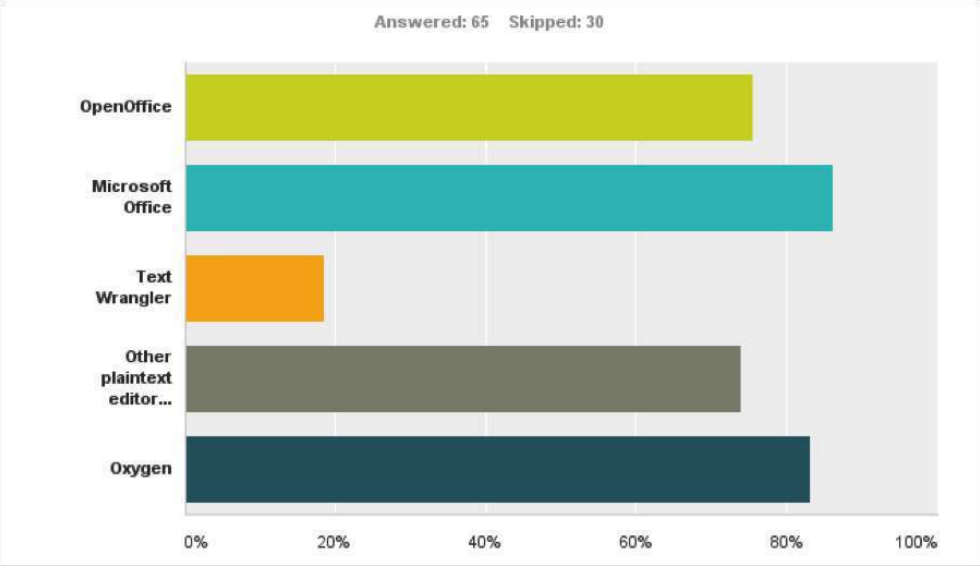


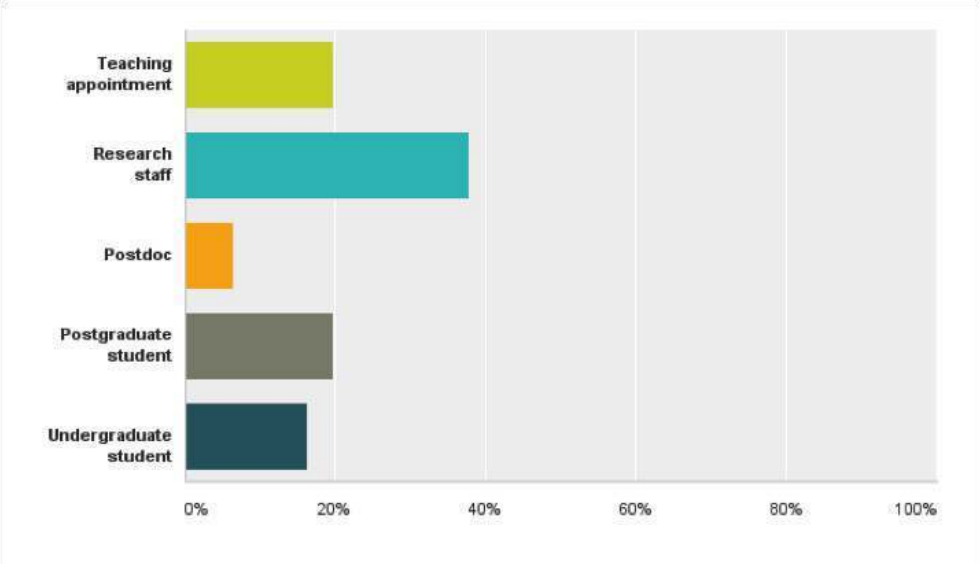

Figure 3: Software proficiency




11 Professionally, at least 60 respondents held their highest degree within a field of the humanities, although a number of respondents also held highest degrees in computer,

library, and information science disciplines. Comments generated by the question "What description most closely matches your current position?" suggested that library staff should have been explicitly offered as a choice among the closed options; five people described themselves in the comments as library-affiliated.

Figure 4: Answer to the question: "What description most closely matches your current position?" (61 respondents)




12 In short, the respondents were largely well-educated, middle-aged, listing English as the language they speak most often, although not necessarily as the language of their greatest fluency, and affiliated with an educational or cultural heritage institution in the Northern Hemisphere.

### 3.1.3 Community Involvement: On versus Off the TEI Mailing List

13 The majority of respondents were familiar with the TEI community, although only a minority identified themselves as active participants. Numerically, 61% of respondents were on the TEI mailing list and 30% of respondents, or around 40% of those on the list, had asked a question via the mailing list.

14 The community of respondents absent from the TEI mailing list, numbering 28 total, was predominantly German, with 58% claiming German as the language they speak most fluently, and young, with 44% falling between the ages of 24–30. A majority were undergraduates (39%), with postgraduates being a close second at roughly 35%, followed by teaching appointments and research staff tied at around 13%. Eighty-five percent had heard of the TEI prior to beginning the survey. The author's current position within a German department of computer science likely influenced the distribution of geographic location and professional pursuits of this cohort.

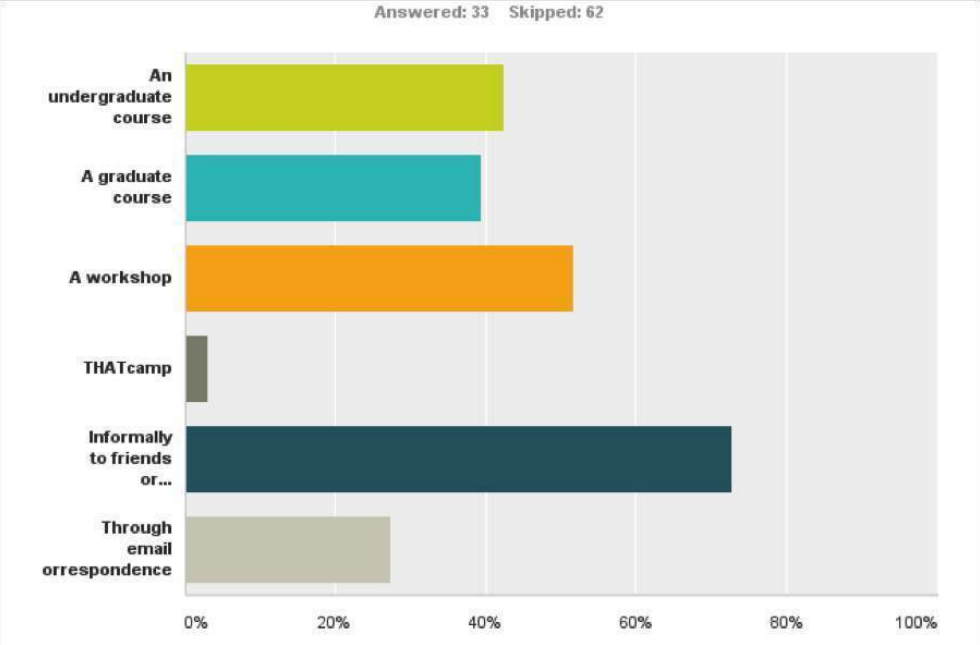
15 Despite not being on the mailing list, seven of the people above had reached out to a member of the TEI community for mentorship or guidance. Six of those people met that mentor in person, finding them through university courses, seminars, or conferences. Only one person found a mentor digitally, via a university website. This was in stark contrast to the demographic on the mailing list, the majority of whom displayed a distinct bias towards seeking advice digitally; it supports Dr. Pierazzo's point (pers.

comm.) that the majority of current TEI introductory support takes place face-to-face. Furthermore, the interactions between those outside the mailing list with those in the TEI community makes for a distinct minority of respondents without any connection whatsoever to the Consortium.

### 3.1.4 Teachers and Students

16 Thirty-three participants had taught the TEI and 33 had not. All of the former responded when asked for the forums in which they have taught it, with nearly all having taught informally to friends or colleagues. The only digital form of instruction mentioned explicitly, "through email correspondence," received a minority of responses.

Figure 5: Among those who have taught the TEI, the capacity(ies) in which they've taught it

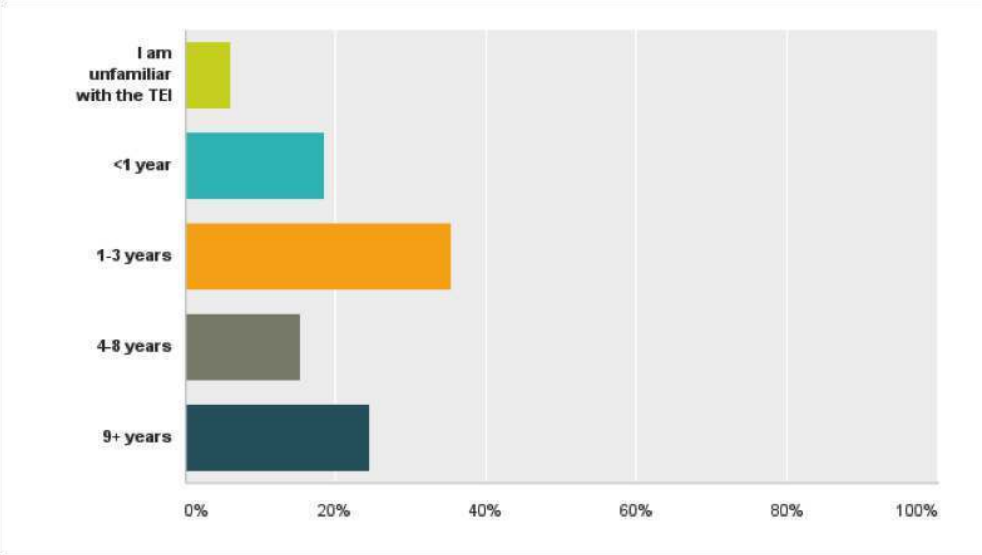



17 figure 5 further reinforces the largely interpersonal nature of current TEI instruction.

### 3.1.5 Experts and Novices

18 The majority of respondents were neither experts nor novices, but rather somewhere in between. This can be seen in figure 6, which demonstrates that the majority of respondents had been working with the TEI between one and three years.

Figure 6: Answer to the question: "How long have you been using TEI in your work?" (65 responses total)




19 I chose to classify as experts the 26 users who had been using the TEI for more than four years; 38% of this group had been using the TEI for 4–8 years, while 62% of these experts had been using the TEI for nine or more years. Sixty-five percent had known about the TEI for longer than eight years. The expert community was older in age, with roughly 38% (10 people) between the ages of 30–40, 38% between the ages of 40–60, 12% 24–30, and 12% older than 60. This group was 62% male, and 52% listed English as the language they speak most fluently, while 30% listed German and the rest either French or Italian. Seventy percent of this group held a position as research staff. Eighty-five percent of this group was on the TEI mailing list; 77% (20 people) had asked a question through the TEI mailing list, 90% (18) of whom had received a satisfactory answer. Seventy-two percent had reached out to another member of the community for guidance, and 91% had found contact information for that person through the mailing list. Eighty-five percent had taught the TEI before, and 72% had taught the TEI informally.

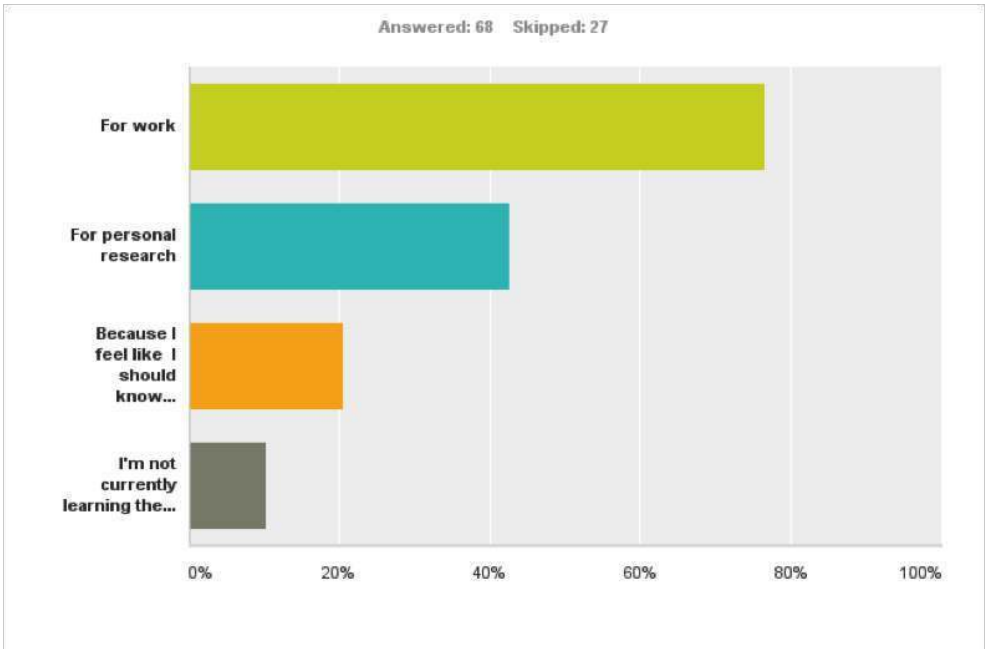
20 Although people who had never before heard of the TEI were invited to participate in the survey in the hope that their qualitative responses might help inform more effective explanations of the TEI and its attraction for beginning users, this cohort was ultimately too small for their responses to be particularly useful. Only three respondents had never before heard of the Guidelines; one worked as research staff, one held a postdoctoral position, and one was an undergraduate student. One commented more explicitly that they were a curator. Therefore, although some of the desired audience was reached, the questions were not written well enough to draw out the helpful information that these respondents might have been able to provide.

21 Therefore, in the sections below I classify those who had heard of the TEI for less than one year within the novice cohort. This included seven respondents, five in the humanities, one in computer science, and one in information systems. Five were currently working in the United States and two in Germany. As mentioned above, due to the small sample size of this group, their feedback serves as a suggestive case study rather than a statistical assertion.

## 3.2 Goals and Aims of TEI Usage

22 The survey sought to elicit descriptions of the specific purposes for which people use the TEI, in order to inform more task-based approaches to pedagogical resources. Respondents were also asked more broadly why they chose to learn the TEI, and why they continue to use it. The majority of overall users had learned the TEI for work (see figure 7).

Figure 7: Answer to the question: "Why did you choose to learn the TEI, or why are you considering learning it?"




23 Also interesting is the large number of users who turn to the TEI for what they consider "personal research." Of course, both "work" and "personal research" are somewhat vague and open to interpretation, so the above graph can only serve as a rough guide to users' attitudes toward their work with the TEI.

24 The goals of the expert cohort were relatively consistent when asked, "What work, research, or publication do you hope that usage of the TEI could help you to complete?" In response, many gave general answers involving digital editions, although given that no chance was offered to be more specific, it remains unclear just how much methodology is shared by those who claim to be working on digital editions. Furthermore, unlike many of the novice and student respondents, the experts often described their work in the plural, unbound from a specific project. They said, "Digital libraries, annotated language corpora, other language resources;" "I am involved in numerous research projects that involve TEI, including a number of digital scholarly editions and tools for working with TEI-encoded texts;" "Almost all my work is TEI related (teaching it, encoding to produce digital editions, etc.);" "manuscript catalogues, digital editions, research papers, presentations/talks;" "various digital editions;" "digital editions, mostly;" "Creating digital editions Creating digital collections [sic]." Sixty-two percent intended to transform their data with an independently-developed stylesheet, 24% with a TEI stylesheet, and only 10% plan to leave them as TEI XML.

25 The respondents from the mailing list gave less consistent answers than the expert users, yet still appeared to be using the TEI primarily in service of interoperability and the construction of digital editions. When asked, "What do you hope to achieve through using TEI markup?," most answers centered around the themes of "well documented, stable, interoperable," or else a project-based answer that often mentioned digital editions. The same percentage as the experts, 62%, planned to transform their documents with an independently developed stylesheet, while only 5%, two people, planned to leave them as TEI XML.

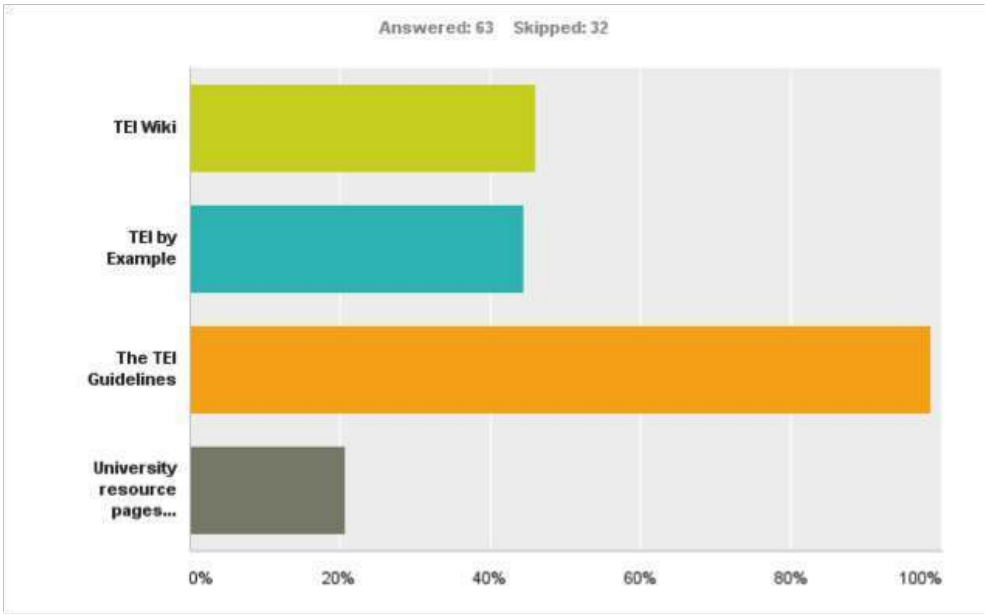
26 Compared to the cohort on the mailing list, the respondents not on the mailing list responded with more specific projects when asked, "What work, research, or publication do you hope that usage of the TEI could help you to complete?" Rather than speaking broadly in terms of "digital editions," this group offered answers such as "Project of my Institute [sic];" "publishing my master thesis [sic];" "a critical edition of a 19c diary;" "I am experimenting with using it to make digital editions of texts in dialect;" "Supporting various faculty projects—most recent is musical notation on a 14th C chant manuscript;" and "Making library collections available for research and teaching." This suggests the need of the more peripherally involved for reference corpora that allow them to quickly find genre-based answers to highly specific questions.

27 When those not on the mailing list were asked, "What do you hope to achieve through using TEI markup?," one person expressed some skepticism, writing, "I am waiting for convincing arguments." Many others listed some explanation of interoperability: "greater compatibility;" "other users/researchers should be able to use my work easily;" "Interoperability;" "Create universally readable files with unambiguous structure, metadata and content." Or editions: "to create a version of our text suitable for both print and online publication;" "a useful online publication;" "I want to create a usable edition of a text with character identification and annotations. Maybe versioning in the future." In striking contrast to those already on the TEI mailing list, 42% planned to leave their files as TEI XML, with only 26% planning to transform them with an independently developed stylesheet. The higher proportion of computer scientists in this subset may have influenced this response; leaving files as raw TEI XML might indicate projects that prioritize the manipulation of XML data over its display.

## 3.3 Educational Resource Usage and Feedback

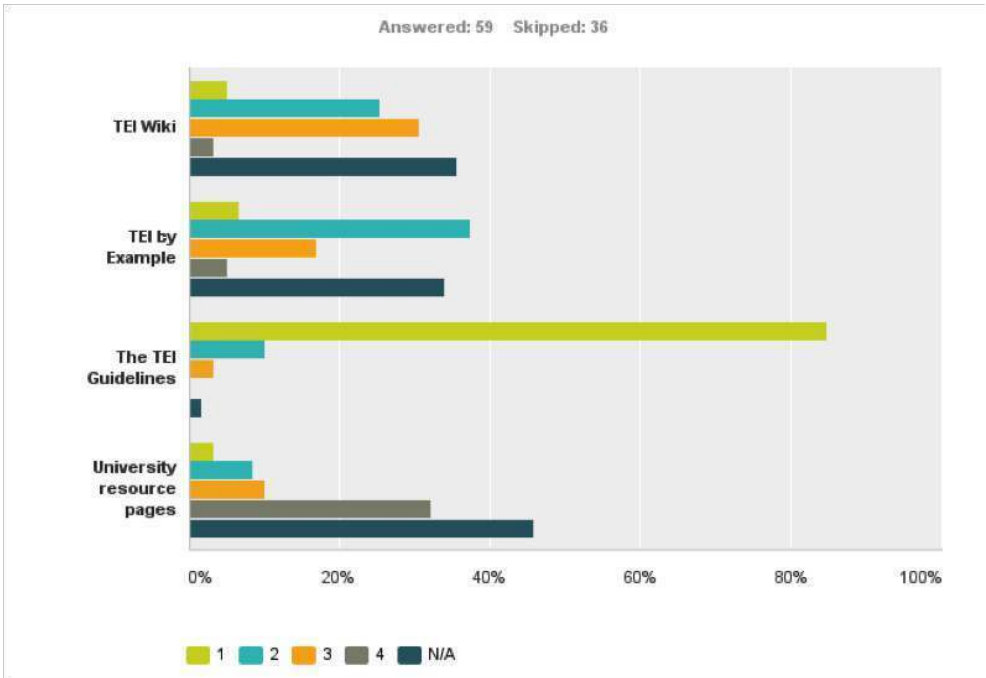
28 The survey requested feedback on extant resources, in order to assess the degree to which current resources are known and accessed, as well as to think about future directions in resource development. Results suggested that the closer the affiliation between the explanatory resources and the TEI Consortium, the more intensively they appeared to be used (figure 8).

Figure 8: Resource usage




29 When all the respondents were asked to rank the resources with which they were familiar in order of helpfulness, an overwhelming majority ranked the Guidelines themselves as the most helpful resource, followed by *TEI by Example*:

Figure 9: Resources ranked by helpfulness




30 However, the numbers suggest that the mailing list itself might be the most effective mechanism for Grade I support in a digital environment, and should have been listed as an option in the question above; of the 43 respondents on the TEI mailing list, 65% of them had asked a question via the list, and 81% of those received a satisfactory answer. Twenty-two, or just over 50% of those on the mailing list, had reached out to a member of

the TEI community for mentorship or guidance. An overwhelming 93% of those had found the contact information for that person via the TEI mailing list, compared to 36% who found the information via a university web page and 7% who found the information via a personal website. This suggests that great deal of knowledge exchange happens through the mailing list itself.

31 A general trend showed more experienced and older TEI users to be self-taught, in contrast to a younger generation who appeared to have received more personal instruction. This most likely speaks not to changing preferences, but rather reflects the increased opportunities for personal instruction in TEI XML. Of those who had been working with the TEI for less than one year, 57% said that they had learned through course attendance, 43% were self-taught, and 29% said their learning was task-based. Only one person identified "online" as their primary educational resource. In contrast, 67% of those on the mailing list said that they were self-taught and 50% said they had attended a course. And in significant contrast, among the expert cohort as defined in section 3.1, a full 88% identified as self-taught, with 54% saying that some of their instruction was "task-based." Less than half of the expert users had ever participated in a formal course, and only 31% had made use of online resources. Sixty-four percent of the respondents not on the mailing list said they were self-taught with respect to the TEI, with 52% making use of a course, 48% describing their instruction as "task-based," and 20% describing it as online. The large numbers of users who achieve success in teaching themselves the TEI Guidelines support the claim that open TEI text repositories could serve as a tool of inductive learning for many researchers.

### 3.3.1 The Teachers' Perspective

32 Among the 29 persons on the TEI mailing list who have taught the TEI, 20 responded when asked to list the most common student questions. Many of these questions concerned post-processing; for example, "1. Students ask questions about how to publish TEI documents on the Web. 2. They ask questions about how to encode to get the document to look a certain way, which of course is not really the point of TEI." "What can you do with the TEI files once ready? Where are the processing tools?" "What does it look like in the end?" "... How can I produce an output if I encoded a text using TEI?" "More examples please! Now I have my TEI document, what can I do with it?" "What happens if I am not TEI conform [sic]?" "Usually the hardest part for them to learn is that TEI is not describing the eventual layout of the web display and that that will come through later processing." "Why go through all these pains, when you can save your text as HTML or PDF and put them on the Internet and Google will find them?" "Why a person/group/institution [sic] should use TEI as opposed to something else? What happens to the TEI files after they are finished?" "Why are we doing this? What's the difference between TEI and EAD?" "Cool, but what can I do with this?" Students seem to face the most difficulty understanding why they should use TEI at all, and additionally that TEI is not a layout format. The former suggests the need for explanatory resources of the variety developed by Mandell and Chesley; the latter the need for TEI XML introductory resources to be coupled with transformation technologies such as XSLT in the manner of Bridget Almas's NEH Tutorial. As will be discussed later in this subsection, certain users find that the absence of integrated XSLT resources on the *TEI by Example* site falls short of their needs.

### 3.3.2 The Learners' Perspective

33 Of those who had been working with the TEI for less than one year, two were on the TEI mailing list and five were not. Only one had asked a question via the TEI mailing list. Two had reached out to a member of the community for guidance, and both had met their mentors in person, one through a summer school TEI training and one through a university course. Five had used the TEI website; two had not. Online tools they claimed to have used included OxGarage, Roma, *TEI by Example*, and the TEI Guidelines, as well as the university-hosted resources published by the Brown University Women Writers Project (Bauman and Flanders 2013) and Humboldt University in Berlin. Within the resources, these respondents mostly reported problems with seeking and finding the information that they need for their particular project.

34 Of the four respondents who declared their affiliation with an academic and/or research institution, all said that their institution did *not* offer instruction in the TEI. In response to the question, "How easily would you be able to find the funding necessary to travel anywhere within your home continent for the purposes of learning the TEI?," five said that it would be possible, but not easy. One said it would be very difficult, and none said it would be very easy. This confirmed the hypothesis that novices are less likely to have access to the funding required for transportation than experts, and suggests that digital pedagogical resources would be useful. Six said they were proficient with HTML, and three said that they would like to learn CSS, suggesting that novices do not necessarily have the proficiency with basic Web tools necessary to work with their files in the manner that they would prefer. The majority hoped to transform their project files with an independently developed stylesheet, and none planned to leave them as TEI XML, confirming the need for resources demonstrating the various workflows in which TEI XML plays a role. Overall, the users still learning the TEI were not as vocal in their comments as those teaching the TEI. As a consequence, the feedback from the teachers paints a more vivid picture of learner challenges.

### 3.3.3 Comments on the TEI Website and Digital Guidelines

35 Even those within the TEI community and expert users of the TEI Guidelines recognize the shortcomings of the digitized Guidelines and TEI website. When those on the TEI mailing list were asked to rate the clarity of the TEI website as very clear, somewhat clear, or not at all clear, 69% rated the website as "somewhat clear." When asked to comment on the utility of the available resources, many on the mailing list focused their complaints on difficulties using the Guidelines to find what they need, especially when they are unsure of their own needs. For example, respondents said, "Finding what you need can be problematic. Lack of links to chapters that describe elements next [to] some element definitions;" "Searching for how elements are nested (what goes where) [can be a frustrating feature]. Looking for examples are helpful via the Guidelines though it seems like you can go around in circles sometimes;" "... it is challenging simply in that there is so much information that not all of it pertains to my work. It is challenging to sort through it all and find what I need." The above shortcomings in searchability are only increased by the unfriendly user interface, in the words of one respondent: "The design, navigation, and general usability of the TEI Guidelines is poor. It is a complex document and making it easy to use is a difficult task, but it has never been properly attempted, in my opinion." Moreover, participants articulated a desire for examples and

contextualization incorporated into the Guidelines themselves. “In some of the guidelines, you can find how an element is used. But why don’t the examples show you the element that contains that element? It would be nice to be able to click through to an example and then navigate the tree of sample text;” “The Guidelines are pretty awful to consult but remain the best resource for comprehensiveness. One thing that would greatly help would be if in the examples in the Elements section more of a context were included for the examples, as sometimes you want to see where the element is used and not simply what it might contain.” In general, the comments of experienced users reflect problems with searchability, interface, and contextualization within the Guidelines.

36 Those not on the mailing list also demonstrated minimal enthusiasm for the current organization and interface of the Guidelines. Twenty-one of the twenty-eight people not on the mailing list had used the TEI website, with two describing the content as “not at all clear,” fourteen as “somewhat clear,” and four as “very clear.” The one person who commented on this question echoed the need for better searchability, writing “there are just too many information on the web site [sic].” A response to a later question by a member of this cohort also implied the need for searchability: “... the TEI guidelines [sic] online provide an overwhelming amount of information, which makes it sometimes a bit difficult to find what you are looking for.” When asked to evaluate features of online resources, comments seemed to suggest the need for more integrated editorialization by experts in the field; resources “Succeeded by linking to examples, failed by not simplifying the reference for beginners.” “Very detailed, there’s a lot to read>too time-consuming ... Important features clearly marked>very helpful.” “TEI by example very useful for thinking about which tags we need for our project. The XSLT resources are pretty unhelpful, though;” “The TEI Guidelines are very comprehensive and provide good examples—but no ‘best practices’ in areas where there is more than one solution. Unfortunately *TEI By Example* did not deliver any better in that respect;” “Not enough variety of examples, case studies—I have sought out examples of projects that use TEI but it can be difficult to identify. Or maybe my problem is that I don’t know how to identify if a project uses TEI or not;” “Useful as a reference resource. Learnt mostly by induction from examples;” “The modular approach TEI takes to including/not including elements is very confusing to the novice or person who’s familiar with working with smaller XML schemas that don’t take a modular approach. The guidelines and the wiki don’t make this clear once you’re deep in the documentation; they assume you already understand this.” All of the above comments which state or imply the need for expert guidance might reflect the absence of the respondents from the TEI-L mailing list, where these kinds of “best practices” are continuously discussed, and perhaps suggest a future need for the TEI-L mailing list archives to be re-formatted for greater accessibility.

## 3.4 Future Priorities of the TEI Consortium

37 When respondents were asked to prioritize future directions of the TEI Consortium, the aggregate results show a clear desire for both pedagogical resources and open text repositories (figure 10 - figure 12).

Figure 10: Answers to the question: “Do you believe the TEI should invest more money in TEI instruction?”

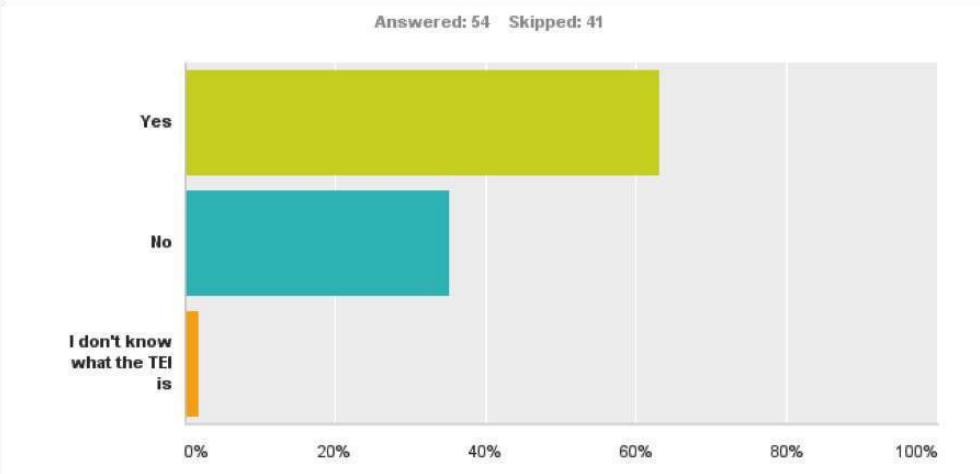



Figure 11: Answers to the question: “Do you believe that the TEI should invest more money in helping those with TEI-encoded files to publish their files?”

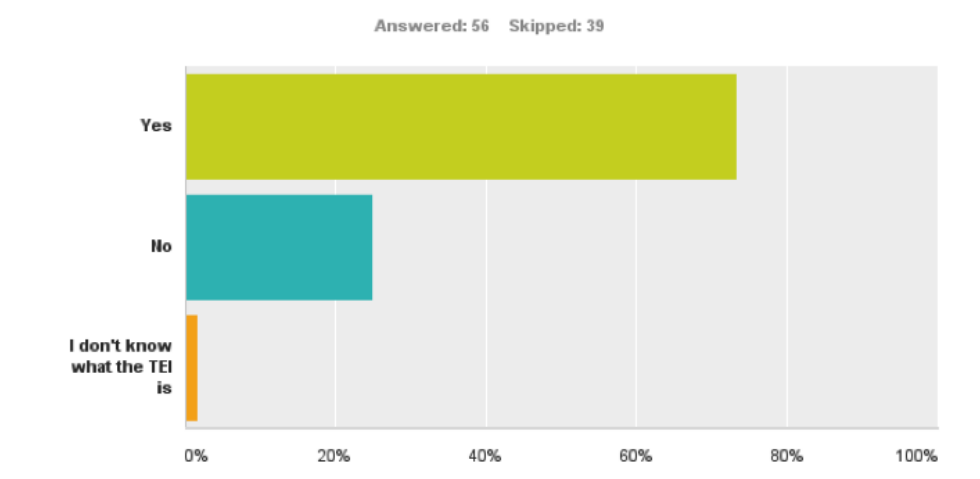
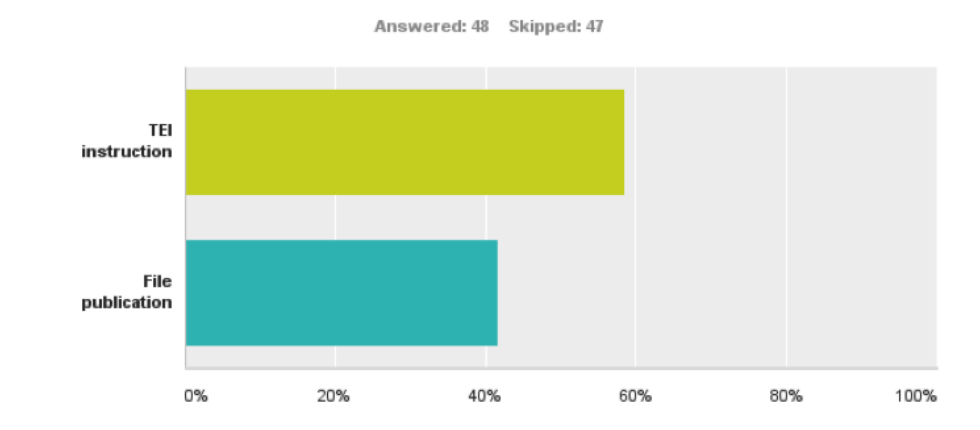
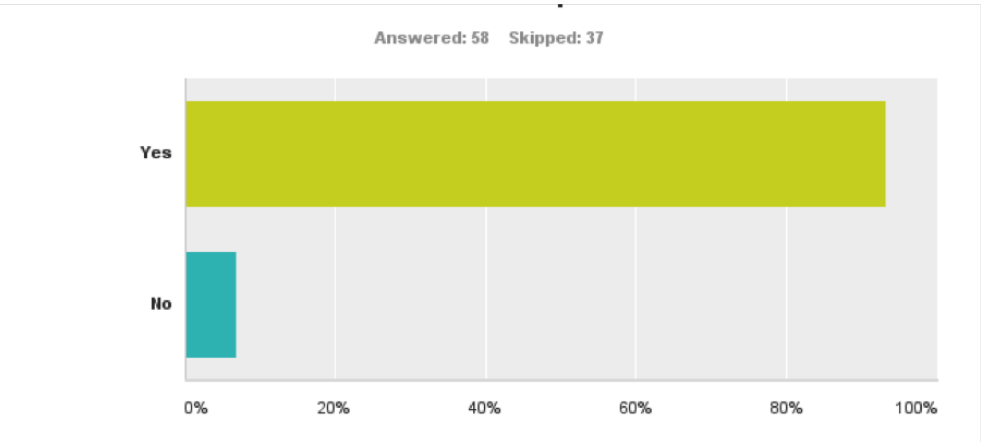



Figure 12: Answers to the question: “Given a limited budget, what should be the highest priority of the TEI community for further development over the next five years?”



38 Participants were also asked to reveal their depth of proprietary feeling with respect to their data, and the overall numbers suggested that researchers are actually eager to release their files (figure 13).

Figure 13: Answers to the question: "Would you be willing to release your TEI-encoded files into an open web of data?"




39 Moreover, this willingness to release files only increases with seniority in the community.

40 For example, of the four who have been working with the TEI for less than a year, only two would be willing to release their files into an open web of data. In contrast, 100% of the expert users would be willing to release their TEI-encoded files into an open web of data, with one commenting that "I think we might open up new research questions about daily life in the past if sufficient numbers of scholars published TEI encoded files or files that could be transformed into TEI compatible data." Similarly, 92% of respondents from the TEI mailing list, or 35 people, would be willing to release their files into an open web of data.[6] Those who would not be willing cite copyright issues and concerns about commercial use, as well as individual proficiency. Some respondents commented: "not sure anyone would find it useful until I gain a greater proficiency. Might be more embarrassing to me than helpful to anyone else." "I'm ready to do it, but honestly I find it's a pain for two reasons: -there is very little hope an actual Humanist scholar will re-use my TEI encoded file instead of using the interface I've developed to access this file [sic]- = the ones who will likely look at the TEI code are nitpicky TEI buffs, who will have ni [sic] business with the content but comment on my use and abuse of the TEI elements and attributes 'til they're blue in the face." However, the general trend of the data suggests that an overwhelming majority would be willing to release their files, and that experts were even more ready and willing to do so than novices.

### 3.4.1 On the mailing list versus off the mailing list

41 Of the 41 people on the mailing list, in response to the question "Do you believe that the TEI should invest more money in TEI instruction?," 24 said yes and 14 said no. When asked "Do you believe that the TEI should invest more money in helping those with TEI-encoded files to publish their files?," 26 said yes and 11 said no. In response to "Given a limited budget, what should be the highest priority of the TEI community for further development over the next five years?," 16 people (55%) marked the answer "TEI instruction" and 13 (45%) marked "file publication." One comment expressed a desire for

"... instruction on file publication, rather than only TEI encoding. It is easier to learn TEI in a self-taught manner using the Guidelines than it is to learn the workflow around TEI, from encoding and transforming and visualizing to publishing in a self-taught way, because that involves a lot more technologies and tools." Another participant explained that they desire "tutorials on project level (blueprints for exemplary server and editor setup, collaboration management, encoding guidelines); tools that allow facsimile integration; further development of generic stylesheets, (partial) transformation scenarios." And similarly, "tools in general. It's time to make the TEI actually useful to individual Humanists with no digital background, and no support from DH centres." Overall, responses from the community of the TEI mailing list adhered to the general trends of the survey as a whole, but even more assertively, and coupled with detailed and helpful comments.

42 A majority (63%) of those outside the mailing list believed that the TEI Consortium should invest more money in TEI instruction, yet even more (79%) believed the TEI should invest more money in helping those with TEI-encoded files to publish their files. In response to the question, "Given a limited budget, what should be the highest priority of the TEI community for further development over the next five years?," however, 68% said TEI instruction compared to 32% for file publication. Whether or not this is an accurate gauge of opinion or whether it was influenced by the survey being named "Learning the TEI" is difficult to determine. In response to the question, "Are there any other thoughts on the use, perception, community or instruction of the TEI, text encoding, or computational skill development for academic humanists that you would like to share?," members of this group wrote: "I'm from china and studying right now in germany. Development of TEI in china is still poor. I hope after my study in germany I can do something in chinese and let the chinese people know, there is an awesome thing called TEI :D;" "TEI should be made as easy as possible;" "raise awareness and rid the world of EAD;" "It should be less complex by separating concerns. The throw everything into one huge schema is completely outdated and self-defeating. Where, after 20 years, are the large searchable repositories of TEI documents?" As with the respondents from the mailing list, the plea for files and case studies resounded throughout the comments.

### 3.4.2 Experts and Novices

43 The experts and novices were more widely divergent from each other compared to the groups on and off the mailing list, most probably because several experienced users choose to remain off the mailing list. The group of expert users was ambivalent with regard to whether or not the TEI should invest more money in TEI instruction, with half saying yes and half saying no. However, 70% believed that the TEI should invest more money in helping those with TEI-encoded files to publish their files. When asked to prioritize the two, 63% believed that file publication should be higher priority than TEI instruction. Interestingly, the desire to learn RDF visible both within the cohort on the mailing list and among the general population was even more pronounced among the expert users, suggestive of a more general trend in the field towards increasing use of linked data, a trend which should perhaps be addressed when considering the creation of future resources.

44 In contrast, three of those who had been using the TEI for less than a year believed that the TEI should invest more money in TEI education, while one did not. One commented that they should "at the very least, invest in a kind of advertising campaign to get

institutions and schools to offer instruction." Even more respondents believed that the TEI should invest more money in enabling file publication; six responded in the affirmative when asked directly. However, when asked to prioritize the two, TEI instruction gained a lead, with four saying that it should take budgetary priority over file publication during the next five years. One respondent commented that the Consortium should "more specifically think about developing tools that allow for easy creation of TEI files," echoing the prior comment regarding "awareness-raising."

# 4. Conclusions

45 The survey results suggest that the improvement of Grade I support for learning the TEI Guidelines in a digital environment is deeply tied to the current needs of experts in the field. In other words, the same resources that will benefit experts—massive open corpora of TEI-encoded text and improvement of the navigability of the digitized TEI Guidelines—will also benefit learners, through establishing a source for a compendium of examples suitable for inductive learning and through enabling users to efficiently find the sections of the Guidelines that serve their purposes.

## 4.1 The TEI Website and Digital Guidelines

46 The survey results suggest that the TEI Consortium should invest serious energy in the organization and searchability of the digitized TEI Guidelines. This most probably requires intensive consideration of usability and technical architecture that is outside the scope of this paper and the expertise of this author to address. However, at minimum the Guidelines should be searchable by genre of encoded text and by language.[7]

## 4.2 File Repositories

47 The data indicate the need for vast and rich repositories of published TEI-conformant files and their schemas. In addition to requiring computationally tractable ways of checking for conformancy, this will most probably require the establishment of a system for distributed review by experts, similar to the model under development by the Perseids Project.[8] This framework would allow scholars and domain experts to gain recognition in the field for new forms of digital micropublication. Once reviewed and approved, all files in such a repository would need to be searchable by tags that they contain, by language, and by genre, like the TEI Guidelines. Moreover, the file repository should include, if at all possible, access to those stylesheets or links to those editions, sites, or databases in which the files were used. Learners would then be able to inductively explore how files are marked up, transformed, and incorporated into larger projects.

## 4.3 Integrated Resources

48 While initiatives such as TAPAS, TEICHI, and CWRC-Writer[9] have begun to address to different aspects of these needs (Flanders and Hamlin 2013; Pape, Schöch, and Wegner 2013; Crane 2010), there has yet to be a deeply comprehensive resource intimately linked to the TEI Guidelines themselves. New technical infrastructure should support workflows

that allow users to enter the genre with which they are working in a search engine connected to the TEI Guidelines (e.g., *poetry*), find a list of relevant tags with explanations of their functions, and from those tags find projects and files that make use of those tags; for example, a search that retrieves all TEI-conformant files using an `<l>` tag, and allows the user to search the projects that created these files.

49 This vision may be a long way off, and should certainly be modified by community expertise, changing needs, and computational realities. However, this is the kind of organized, integrated, and open plan that the TEI community, both present and potential, seems to be calling for.

## BIBLIOGRAPHY

Almas, Bridget. 2012. *NEH Institute Working with Text in a Digital Age: Digital Edition Demonstration and Sample Code.* Tufts University, July. Accessed July 11, 2013. https://github.com/balmas/tei-digital-age

Bauman, Syd, and Julia Flanders. 2013. *Resources for Teaching and Learning Text Encoding*. Brown University Women Writers Project. Accessed April 10, 2013. http://www.wwp.brown.edu/outreach/resources.html.

Burghart, Marjorie. 2013. "TEI Cheatsheets." *TEI Wiki*. Accessed September 9, 2013. http://wiki.tei-c.org/index.php/TEI_Cheatsheets.

Burghart, Marjorie, and Malte Rehbein. 2012. "The Present and Future of the TEI Community for Manuscript Encoding." *Journal of the Text Encoding Initiative* 2. Accessed September 9, 2013. http://jtei.revues.org/372. doi:10.4000/jtei.372.

Chesley, Amelia. 2012. "TEI: An Overview." Youtube video, 10:12, posted March 2. https://www.youtube.com/watch?v=R6iiIFrWvmU.

Chickering, Arthur W., and Stephen C. Ehrmann. 1996. "Implementing the Seven Principles: Technology as Lever." *AAHE Bulletin* 49(2): 3–6. Updated version accessed September 9, 2013. http://www.tltgroup.org/programs/seven.html.

Crane, Gregory. 2010. "Give us Editors! Re-inventing the Edition and Re-thinking the Humanities." OpenStax CNX, May 13. http://cnx.org/content/m34316/latest/.

Flanders, Julia, and Scott Hamlin. 2013. "TAPAS: Building a TEI Publishing and Repository Service." *Journal of the Text Encoding Initiative* 5. http://jtei.revues.org/788. doi:10.4000/jtei.788.

Geertz, Clifford. 1973. *The Interpretation of Cultures: Selected Essays*. New York, NY: Perseus Books Group.

Halsell, Lindsey A. 2013. "TEI and So Can You: Corpus Linguistics and the Text Encoding Initiative." *PNLA Quarterly* 77(3): 63–70.

Hirsch, Brett D. 2012. "</Parentheses>: Digital Humanities and the Place of Pedagogy." *Introduction to Digital Humanities Pedagogy*, 3–30. Cambridge: Open Book Publishers. http://

www.openbookpublishers.com/product/161/digital-humanities-pedagogy--practices--principles-and-politics.

Mahony, Simon, and Elena Pierazzo. 2012. "Teaching Skills or Teaching Methodology?" In *Digital Humanities Pedagogy*, edited by Brett D. Hirsch, 215–25. Cambridge: Open Book Publishers. http://www.openbookpublishers.com/product/161/digital-humanities-pedagogy--practices--principles-and-politics.

Mandell, Laura. 2013. *Introduction to Digital Textual Editing: An UNOFFICIAL Guide to the Value of TEI*. Slidecast posted June 30, 2013. http://www.slideshare.net/mandellc/tei-and-xslt-23711832.

Pape, Sebastian, Christof Schöch, and Lutz Wegner. 2012. "TEICHI and the Tools Paradox: Developing a Publishing Framework for Digital Editions." *Journal of the Text Encoding Initiative* 2. http://jtei.revues.org/432. doi:10.4000/jtei.432.

Roueché, Charlotte, and Julia Flanders. 2014. "Introduction for Epigraphers." *EpiDoc Guidelines: Ancient Documents in TEI XML*. Last updated February 9, 2014. http://www.stoa.org/epidoc/gl/dev/intro-eps.html.

Spiro, Lisa. 2010. *Digital Humanities Education*. Zotero Group. Accessed September 9, 2013. https://www.zotero.org/groups/digital_humanities_education/.

Van den Branden, Ron, Melissa Terras, and Edward Vanhoutte. 2010. *TEI by Example*. Last updated July 9, 2010. http://www.teibyexample.org/.

## NOTES

**1.** I understand the term *Grade I support*, used in Martin Mueller's email, to mean those tools and resources that meet the needs of the hypothetical graduate student framed as the use-case here.

**2.** TEI-L Home Page, accessed September 7, 2013, http://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L.

**3.** "Getting Started Using TEI," accessed September 7, 2013, http://tei.oucs.ox.ac.uk/GettingStarted/html/.

**4.** Bridget Almas published a tutorial entitled *Working with Text in a Digital Age* which walks the user through the creation of a digital edition for the NEH Institute—the resources can be found here: https://github.com/balmas/tei-digital-age; Laura Mandell posted a Slideshare *Introduction to Digital Textual Edition: An UNOFFICIAL Guide to the Value of TEI* that can be found here: http://www.slideshare.net/mandellc/tei-and-xslt-23711832; Lisa Spiro's *Digital Humanities Education* Zotero Group is here: https://www.zotero.org/groups/digital_humanities_education/; and Amanda Chesley's introduction can be found on Youtube: https://www.youtube.com/watch?v=R6iiIFrWvmU.

**5.** This survey was the research for a master's dissertation at King's College, London, under the supervision of Dr. Elena Pierazzo.

**6.** Please remember that there is some redundancy among these cohorts; nevertheless, the point stands.

**7.** Marjorie Burghart (2013) took a first step towards alternative ways of navigating the Guidelines with the "TEI Cheatsheets," http://wiki.tei-c.org/index.php/TEI_Cheatsheets, targeted at humanities scholars already familiar with the conceptual purpose of the Guidelines and seeking only the right tag to fit a specific situation.

**8.** For more information, see Perseids: A Collaborative Editing Platform for Source Documents in Classics (a project of the Perseus Digital Library), http://sites.tufts.edu/perseids.

**9.** "Welcome to CWRC Writer," CWRC-Writer Help, accessed September 7, 2013, https://sites.google.com/site/cwrcwriterhelp/.

---

## ABSTRACT

This article provides a brief overview of currently-available digital resources for learning to understand and use the TEI Guidelines. It reflects on and analyzes these resources and their audience through the results of a survey intended to inform future support from the TEI Consortium for novice users. Increasing numbers of students look online for self-directed and task-based tutorials, and increasing numbers of scholars in the humanities recognize the TEI Guidelines as a standard tool for publication and analysis. In this context, the author designed the survey presented in this paper to solicit qualitative feedback from both experienced and aspiring practitioners in the field concerning their skills, needs, and goals, pedagogical as well as technical. The article suggests revising and expanding TEI community resources, proposing possibilities for their new form and functionality.

## INDEX

**Keywords:** pedagogy, student resources, survey, file publication

## AUTHOR

### STELLA DEE

Stella Dee is currently a Research Associate with the Open Philology Project at the University of Leipzig. The research of this article was conducted while studying for a Masters degree in digital humanities at King's College, London.