What is the TEI?

The Text Encoding Initiative is an attempt to prevent the digital equivalent of the Tower of Babel. It proposes a community standard for representing digital texts in a way that is both powerful and responsive to the research needs of humanities scholars.

In practical terms, it is several things:

- it is a set of guidelines for encoding humanities texts using XML, the eXtensible Markup Language
- it is a standards organization that maintains and develops an XML encoding language for encoding humanities texts;
- it is an international consortium that exists to support this standards development work
- it is a community of projects and individuals who use the TEI Guidelines

Like many standardization efforts, the TEI faces the challenges inherent in such a project. How can the many disciplines and communities within the humanities domain find common ground in a single encoding language? How do we agree on the level of detail that is necessary or appropriate in describing our textual materials? How do we reconcile the advantages to be gained by consistency and agreement with the need for individual specialization? How do we handle the truly idiosyncratic and unexpected?

Unlike many standardization efforts, the TEI addresses these and similar questions by explicitly accommodating variation and debate within its technical framework. The TEI Guidelines are designed to be both modular and customizable, so that specific projects can choose the relevant portions of the TEI and ignore the rest, and can also if necessary create extensions of the TEI language to describe facets of the text which the TEI does not yet address. Because the TEI itself is complex, the customization process is not entirely trivial, but it is designed to be as straightforward as possible.

What is it used for?

For sheer quantity, the predominant use of the TEI Guidelines is in creating large digital library collections, such as those being developed at the University of Michigan's Humanities Text Initiative (for example, the American Verse Project), or the Electronic Text Center at the University of Virginia. Collections of this type provide access to large quantities of textual material, often focusing on rare or fragile materials which would otherwise be inaccessible. The text encoding in collections like these emphasizes features that will be of immediate, general use for searching and retrieval: information such as bibliographic data, basic text structure, and metadata such as subject keywords to help users find materials of interest.

The TEI Guidelines are also used by more specialized research projects to represent smaller, more thematically focused collections of texts, often organized around a single genre, author, period, or
country (or a combination of these: for instance, the Italian Women Writers Project at the University of Chicago). Projects of this type often use more detailed markup to represent particular features of the text that are relevant to the specific collection or important for the specific scholarly audience being served. For instance, a collection focusing on an author whose writings are an important source of information about famous contemporaries might usefully encode all references to names and works, perhaps including links to more detailed indexes of biographical or critical information. Similarly, an electronic edition of a particular author or work might well include a representation of variant readings, authorial revisions, editorial emendations, and similar editorial information. Some collections of this sort are intended to serve very specific research goals, such as linguistic analysis, or to serve as the basis for a dictionary, and in these cases the markup may be very highly specialized.

The uses described above are all some species of publication: the goal is to create a digital collection that can be used online by some public audience of greater or lesser extent—perhaps limited to a small community or to paying subscribers, perhaps open to the general public. More rarely, the TEI is also used by individuals to create digital representations of textual materials to support their own personal research, in forms which may or may not be published. Where the scope and purpose of the larger collections may be determined by audience and funding, in the case of an individual's work the constraints are more personal and professional: the encoded material might serve as a private research tool, or might develop into the equivalent of a digital monograph that represents the author's analysis of a set of texts. The use of the TEI Guidelines in these cases may be as detailed as the author finds useful—limited only by time, energy, imagination, and the constraints of usefulness.

**How do you learn about the TEI?**

The TEI web site is a good source of general information about the TEI, and is also the best place to find the TEI Guidelines themselves, as noted above. However, the TEI web site is only one of many sources of useful information about how to use the TEI and how to understand its significance.

For those who are interested in the TEI at a general level—those who need to understand what it is, what it does, and why it's important—there are a few different ways to begin. Although there is no single source that can give a complete picture, there is now a growing literature on the role of text encoding in scholarly work. The bibliography that accompanies this site attempts to list some of the sources we have found most useful. The WWP's Guide to Scholarly Text Encoding also includes a general overview of the central issues. Perhaps the best way to get an initial grounding in the TEI is to attend a workshop. The workshops in the WWP's 2007-2008 series, funded by the NEH, are specifically aimed at providing this kind of conceptual overview, coupled in some cases with hands-on experimentation, but aiming above all at examining how text encoding in general, and the TEI in particular, connect with basic scholarly concerns and practices. However, even a hands-on workshop that focuses on technical specifics can be a very good way to get a basic understanding of what is at stake. While the specific details of the encoding language may not be important for your work, they can give a fuller sense of how text encoding operates as a representational strategy, and of what it can and cannot do. Hands-on workshops are offered each year in the following programs:

- University of Victoria, Digital Humanities Summer Institute
- University of Virginia, Rare Book School
- International Congress on Medieval Studies at Kalamazoo (two introductory TEI sessions in
2007)

TEI workshops of various sorts are also held fairly regularly at Oxford University, Brown University, the University of Illinois at Urbana-Champaign, the University of Maryland, and other places as well. (If you know of additional events that should be listed here, please send email to WWP at brown dot edu.)

For those who need a more detailed understanding of the actual encoding process, the above workshops are a good start, but they are not sufficient on their own. Learning the TEI is like learning a language—it has a fairly extensive vocabulary and a complex range of usage. An introductory workshop will give you a good understanding of what the language is for and of its basic terms, but you need to follow it with both practice and a detailed exploration of how the language is really used. Having a project you can work on—a set of documents that interest you and will motivate your researches—is a great help. To gain a detailed knowledge of the TEI Guidelines as you encode, there are several things you can do:

- Read the TEI Guidelines, starting with the basic chapters (chapters 6 through 10), and extending to the chapters on more specialized topics. The most immediately relevant for most encoding work are chapter 14 (which covers linking and the alignment of parallel texts), 15 (which covers methods of representing analytical and interpretive perspectives on the text), 17 (which covers the representation of uncertainty and statements of responsibility), 18 (which covers the transcription of primary sources and of manuscripts in particular), 19 (which covers the encoding of critical apparatus), and 20 (which covers the detailed encoding of names, places, and dates).
- Read and post questions to the TEI listserv, and search the TEI-L archives for discussions of topics that interest you. TEI-L contains a very detailed record of a huge range of encoding issues, some abstruse and some very basic; the range of opinions and approaches can give you a valuable sense of how different kinds of projects use the TEI.
- Look at the work of actual text encoding projects. Many projects have documentation and some have exceptionally good documentation that explains the rationale behind their encoding decisions and the criteria by which they recognize and encode specific textual features. We list some exemplary documentation on our readings and resources page. Ask questions; most encoding projects are happy to help those who are new to the field and glad to find that their experience is valuable. Many are also happy to share sample encoded texts, which can be a very useful way of getting a more detailed view of the encoding landscape.

What kind of TEI knowledge is useful, and for whom?

The modern academy is full of digital research tools and online collections, and humanities scholars and students increasingly depend on digital materials—primary sources, research databases, online journals, and the like. Among these collections, the ones which take the editing and scholarly integrity of the digital text most seriously are increasingly using TEI. Large-scale initiatives such as Early English Books Online Text Creation Partnership are using the TEI as their underlying encoding model, and the major digital editions now being produced typically use TEI because it offers the most nuanced and rigorous approach to describing the complexities of the edited text currently available. Some understanding of text encoding in general, and of the TEI in particular, is thus becoming the
equivalent of an understanding of standard editorial practices a few decades ago: helpful for anyone who works closely with texts, and essential for anyone whose research depends on a critical awareness of how scholarly sources are produced and consumed.

Although comparatively few scholars currently are actively engaged in creating digital materials of this type, that may change in the near future. Digital publication is becoming more common, even though there are still strong institutional obstacles (such as the still scanty formal recognition of digital publication for tenure and review). As these obstacles diminish, and as the tools for digital publication become more widely available and easier to use, scholars may find that involvement in digital publications of one sort or another—digital editions and anthologies, digital monographs, collaborative digital research projects—is both feasible and attractive. As a consequence of this shift, they may also find that the scholarly discourse is increasingly conducted in digital modes. To whatever extent this becomes true—perhaps slowly and only partially—those who want to retain a full engagement with their colleagues' work and with the research materials available will benefit from an understanding of the underlying concepts and technologies involved.

It should be emphasized, though, that this understanding need not be at a technical level. The most significant concepts of text encoding, from a scholarly standpoint, are not the technical details but rather the broader ideas about modelling textual information, representational strategies, and editorial method: in fact, the same domain that has been the province of scholarly editing for centuries. What needs to be grasped is how these ideas translate into the digital medium, and what changes when they do.

For scholars who are planning a more active involvement in a TEI project—possibly as an advisor or director—these broad concepts are crucial but in addition it will be important to understand how the TEI specifically represents textual information: what kinds of features it handles most naturally, what features require additional work or customization, what kinds of customizations are appropriate and how the customization process works. These issues are important because they have an impact on the guidance of the project and on basic practical matters such as what kinds of staff are needed and how much work they will need to do.

Some scholars may also be interested in experimenting with TEI encoding on their own, perhaps to begin an edition of a particular text, or to create a digital text for their own research purposes. The easiest way to get started in such a case is probably to take a workshop, but you can also use the materials that are available here to get oriented.

Clearly those who have technical responsibility for designing and developing TEI publications need to know far more. In addition to an overall familiarity with the TEI language and its range of usage, they need a strong understanding of how the TEI customization works and of the tools that are used to manipulate TEI documents and schemas. Information on these topics is available from the TEI, both at the [TEI web site](http://www.tei-c.org) and at the [SourceForge](http://sourceforge.net) site where the TEI maintains the TEI Guidelines. The TEI listserv (TEI-L@listserv.brown.edu) is also a good place to consult with experts who are willing to help.

[an error occurred while processing this directive]