Overview

This workshop is really about how to represent research objects
--In a sense, not even really about how to represent them in digital form, although that is our practical emphasis.
--but rather, what are the salient informational features of the research materials that concern us, and how do we *understand* those features, how do we derive meaning from them? What aspects of their nature do we need to express in order to have a useful representation of the object?

The TEI is not so much an attempt to answer those questions, as it is a tool for expressing our answers.

So there are things many of you already know about the TEI, but for the sake of making sure that we are all on the same page, we'd like to talk a bit about the TEI and also (as background) about the kind of digital representation the TEI provides.

The TEI exists to make it possible for us to represent research materials in a way that is:
--sustainable (standards-based)
--sharable (provides conditions for mutual intelligibility)
--and above all, analytically rich (a very large vocabulary for describing and analysing humanities content)

But in addition, it focuses our attention in a very particular way on the translation of observations about texts into formal information
--we look at a book and we can read it, handle it, study it
--we take a digital image of the book and we can read that, disseminate it
--but if we do a TEI encoding of the book other things become legible: for instance, the array of genres it contains; the people it addresses and names; the way it quotes and cites other works, the ways it has been revised

These things become legible in a different way: not just things we can see when we look at the page ("oh, look, it's a poem") but things we can observe and work with at a distance and at scale: for instance, we can look at the interplay of poems within prose contexts (at the start of chapters in Middlemarch; in dialogue in Charlotte Smith's children's works; in personal narratives in Lady Mary Wroth's Urania).

The reason this is possible is because languages like the TEI serve as a kind of intermediary between the brute stuff of the world, and the formal analytics of a research process.
--through TEI, we select the parts of the text that are of interest to us
--we describe them using a set of formal descriptors (which we will be looking at closely in a moment)
--and to some extent we regulate and constrain our descriptions according to our expectations about genre and document structure.
So although it looks as if what we're doing here is *reproducing* a text, in fact what we're doing is modeling it, very selectively and very strategically.

--we're creating a proxy, a surrogate, sort of like a plastic model of an atom or a DNA strand: with selective, useful omissions and distortions

This model then serves as the basis for all the things we want to do with our text in the digital world: publish it, print it, visualize it, analyse it.

Our success in these activities depends directly on the quality of the model, and on how well it matches our research aims with respect to the documents we're working with.

So: during the course of this workshop we'll be learning the core of the TEI, and you'll be experimenting with the TEI on your own materials.

--basic encoding this morning

--more advanced encoding tomorrow morning

---tomorrow afternoon, we'll take a closer look at some of your projects and the questions that arise from them.

What I'd like to do now is ask you about the kinds of text modeling you're doing:

1. What kinds of information have the highest research value for your work with your texts? what are you modeling?

2. What are you omitting? (are there other people who find that omission problematic?)

*****

So just to make sure we're all on the same page, a little background on the TEI and on what it is

1. The TEI is a markup language.

2. It is developed and maintained by an international consortium, and it is free and open-source; the community that uses the TEI also has responsibility for developing and expanding the TEI (and this includes all of us here)

3. It has been around for over 20 years and is still growing and developing; it is both a standard and an ongoing research effort.

So before we start looking at TEI samples, let's take a quick step backwards and say a few words about XML and how it fits in...
I said TEI is a markup language; what is XML? XML is a metalanguage: a system for defining markup languages....

Slides on XML