This lecture

- How is transcription different from other encoding?
- Encoding from a physical document PoV
- What are the main elements necessary for transcription
How is transcription different?

- Transcription aims to:
  - make a primary source accessible
  - make a primary source comprehensible

- Which may entail:
  - adding or making use of additional information

- Transcription is:
  - selective
  - interpretative
  - subjective
  - dependent on editors’ choices
How is transcription different?

- Transcription happens in all types of editing, so how is this different?
  - Encoding for transcription foregrounds a different point of view

I am by birth a Genevese, and my family is one of the most distinguished of that republic. My ancestors had been for many years counsellors and syndics, and my father had filled several public situations with honour and reputation. He was respected by all who knew him for his integrity and indefatigable attention to public business. He passed his younger days perpetually occupied by the affairs of his country;
What is Transcription?

- A range of phenomena based on technologies and historical context:
  - Original layout information
  - Abbreviations or other ‘arcana’
  - Errors which invite correction or conjecture
  - Scribal additions, deletions, substitutions, restorations, transpositions
  - Damaged or illegible passages
  - Non-standard orthography which invites normalisation
  - ...

[Cartoon image of a monk stating 'PRINTER’S DOWN.']
Transcription: how to do it

- Inspect the original document or a surrogate (facsimile)
  - Identify areas in the document (text, graphics, etc)
  - Find the first line
  - Type the first line, identify special characters
  - Decide what kind of diplomatic method you are using; this will determine issues of lineation, and how much can be transcribed.
  - Record textual modifications/interventions
    - Highlighting
    - Additions
    - Deletions
    - Transpositions
    - Notes
  - Identify text-structures
  - Recognise writing activities
  - Identify named entities
Transcription: getting physical

- Can we record all this? Yes we can!
  - Inspect the original document or a surrogate (digital or otherwise, facsimile)
    - `<sourceDoc>` or `<facsimile>`
  - Identify areas in the document (text, graphics, etc)
    - `<surface>` or `<zone>`
  - Find the first line
    - `<line>`
  - Type the first line, identify special characters
    - `<g>` or `<hi>`
  - Record textual modifications/interventions
    - `<hi>`, `<mod>`, `<transpose>`, ...
  - Identify text-structures
    - `<div>`, `<head>`, `<p>`, `<list>`, ...
  - Recognise writing activities
    - `<add>`, `<del>`, `<subst>`, `<abbr>`, ...
  - Identify named entities
    - `<name>`, `<persName>`, `<placeName>`, ...
Transcription: structure

- `<TEI>`
  - `<teiHeader>`: Provides descriptive metadata of the original resource, possibly including a `<msDesc>`
  - `<facsimile>`: Organises a set of page images representing a set of `<surface>`s
  - `<sourceDoc>`: A non-interpretative representation of a document considered purely as a physical object
  - `<text>`: Contains a textual structured representation of the intellectual content (the text itself)
Transcription: module

- header
- transcr
- core
- text structure

TEI
Abbreviations

- An abbreviation is used in manuscript materials to shorten labour
- It uses significant marks to replace:
  - Single letters
  - Groups of letters
  - Whole words
  - Whole phrases

Recap: as we’ve seen, the core model already supports some simple elements to record abbreviations:
- `<choice>` → groups abbreviated and expanded readings
- `<abbr>` → abbreviated form
- `<expan>` → expanded form
Abbreviations

- **Suspension** —> only the first letter of a word or phrase, followed by a point
  - <abbr>p.</abbr> —> <expan>page</expan>
- **Contraction** —> the letters in the middle of a word are omitted
  - <abbr>Dr</abbr> —> <expan>Doctor</expan>
- **Brevigrap**h —> a single character representing two or more letters
  - <abbr>son</abbr> —> <expan>person</expan>
- **Superscript** —> characters representing various kinds of contractions
  - <abbr>Ma.</abbr><sup>tie</sup> —> <expan>Majestie</expan>
Abbreviations

- How do you view abbreviations?

Representation of a particular sequence of marks on the page (i.e., p with a bar through the descender, s, o n)

Representation of the particular sequence of marks it is believed to stand for (i.e., person)

We don’t have to decide, we can encode both!
Abbreviations

- How do you view abbreviations?

<am> — abbreviation marker — contains a sequence of letters or signs present in the abbreviation that are omitted or replaced in the expanded form

<ex> — editorial expansion — contains a sequence of letters added by the editor when expanding an abbreviation
Abbreviations

Ma.tie

<choice>
<abbr>
Ma<am rend="superscript">tie</am>
</abbr>
<expan>
Ma<jex>jestie</jex>
</expan>
</choice>
Corrections and emendations

- Apparent errors can be recorded:
  - in their original state —> <sic>
  - as corrected text —> <corr>
  - combined —> <choice> (+ <sic> and <corr>)

Recap: as we’ve seen, the core model already supports these elements!

My favourite food is *pasghetti*
Regularisation

Coopers Hill.

If there be Poets, which did never dreame
Upon Parnassus, nor did tast the streame
Of Helicon, we justly may suppose,
Modifications

- There are several ways of recording modifications in a text:
  - `<mod>` → general modifications, without any specific interpretation
    - `@type` attribute can be used for further specification
  - `<add>` → addition to the text
  - `<del>` → characters or sequence of characters marked as deleted from the text
  - `<subst>` → groups additions and deletions as a single intervention
  - `<supplied>` → editorially supplied text
Modifications

not restrain
<mod>
  <del>us.</del>
  <add>our enquiries</add>
</mod>
I revolved these circum

not restrain
<subst>
  <del>us.</del>
  <add>our enquiries</add>
</subst>
I revolved these circum
Partially legible text

- The text may be only partially legible, due to:
  - cancellation;
  - damage;
  - other environmental factors.
- `<unclear>` marks the portion of the text that can be read
  - `@reason` records the cause of uncertainty
  - `@resp` attributes responsibility for the offered reading
  - `@cert` records the degree of certainty for the offered reading
  - `@agent` categorises the cause of any damage
Omitted or damaged material

- The text may be **completely illegible** due to:
  - thorough cancellation;
  - damage;
  - other environmental factors.

- Or the editor might decide to omit a certain portion of the text
  - `<gap>` → records the existence of text that cannot be read or has been purposefully left out
  - `<damage>` → records the existence of an area of physical damage to the source

*<damage> element often contains <gap>s or <unclear> sections of text*
And I need not bee ashamed to owne
<damage agent="rip">
  <supplied cert="high">the</supplied>
  <unclear>m.</unclear>
</damage>
Transcription: structure

<TEI>
  <teiHeader>
  <facsimile>
  <sourceDoc>
  <text>
Facsimile

- A digital facsimile contains images of the pages of a source document.
- A valid (but not very useful) TEI document may contain only:
  - Metadata
  - Images of the source document
- A valid (and much more useful) TEI document may also contain:
  - Metadata
  - Images of the source document
  - A transcription of those pages
- `<facsimile>` —> contains the representation of the source as a set of images
- `<graphic>` —> indicates the location of each image with `@url` attribute
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- METADATA! -->
  </teiHeader>
  <facsimile>
    <graphic url="A1r.png"/>
    <graphic url="A1v.png"/>
    <graphic url="A2r.png"/>
    <graphic url="A2v.png"/>
    <!-- and so on -->
  </facsimile>
</TEI>
Facsimile and transcription

- We can combine a facsimile with a transcription:
  - If the transcription is regarded as text in its own right, and organised independently of its physical manifestation, we use the `<text>` element
  - If the transcription is meant to prioritise its physical manifestation, we use the `<sourceDoc>` element
Facsimile and transcription

- For simple cases in which one image corresponds to one page transcription:
  - Facsimiles may be referenced directly in <text> by using:
    - `<pb/>` —> the milestone page break element in conjunction with
    - `@facs` the facsimile attribute.

```xml
<text>
  <pb facs="Alr.png"/>
  <!-- transcription of Alr goes here -->
  <pb facs="Alv.png"/>
  <!-- transcription of Alv goes here -->
</text>
```

- However, this approach assumes that
  - only one image exists per page of text
  - the transcribed page corresponds exactly with the image
- Not a very scalable approach
Facsimile and transcription

- A more sustainable approach is to define `<surface>`s and `<zone>`s with `<facsimile>`s
- `<surface>` → defines a written surface as a two-dimensional coordinate space
  - i.e., pages, openings, etc.
- `<zone>` → defines a single area on a `<surface>` using coordinates
  - @points → define the `<zone>` using point-pairs to build the text area (x, y)
  - @ulx, @uly, @lrx, @lry → define the upper left and lower right corners of a rectangle
  - You can automate the definition of these coordinates by using additional tools, for example:
    - Oxygen facsimile addon
    - Image Markup Tool
Facsimile and transcription

<facsimile>
  <surface xml:id="A1r"><graphic url="A1r.png"/></surface>
  <surface xml:id="A1v"><graphic url="A1v.png"/></surface>
  <surface xml:id="A2r"><graphic url="A2r.png"/></surface>
  <surface xml:id="A2v"><graphic url="A2v.png"/></surface>
</facsimile>

<text>
  <pb facs="#A1r"/>
  <!-- transcription of A1r goes here -->
  <pb facs="#A1v"/>
  <!-- transcription of A1v goes here -->
</text>
Facsimile and transcription

- Defining <zone>s is particularly useful for:
  - Encoding text that is further signified by its position on the page
  - Encoding MSS that have different blocks of composition
  - Encoding text in non-standard positions
Facsimile and transcription

As it hath beene diverse times acted by his Highness se-

princes in the Cittie of London : as also in the two V-

niversities of Cambridge and Oxford, and els-where
Transcription: structure

<TEI>

<teiHeader>

<facsimile>

<sourceDoc>

<text>
Documentary transcription

- `<sourceDoc>` contains a documentary, embedded or faithful transcription of a single document, prioritising its physical disposition and/or its genetic history
  - Like `<facsimile>`, a `<sourceDoc>` usually contains one or more `<surface>` elements, each with potentially multiple `<zone>`s or `<line>`s
    - `<line>` —> contains the transcription of a topographical line
- Some editorial markup is allowed (`<add>`, `<del>`, `<unclear>`, etc.) but any interpretative markup should be avoided
Documentary transcription

```xml
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- METADATA! -->
  </teiHeader>
  <sourceDoc>
    <surface>
      <zone>
        <line>
          <!-- transcription of first line goes here -->
        </line>
        <line>
          <!-- transcription of second line goes here -->
        </line>
        <!-- and so on -->
      </zone>
    </surface>
  </sourceDoc>
</TEI>
```
Genetic Editing

- Genetic editing is concerned primarily with the order of composition.
- There are several elements in the TEI that can be used for genetic editing, such as:
  - `<mod>` → a familiar one, general modification element
  - `<metamark>` → any kind of mark designed to guide the reading of a document
    - (i.e., arrows, lines, index symbols, etc.)
  - `<retrace>` → any writing that has been rewritten or otherwise fixed or reinforced
    - (i.e., for example tracing in pen what had been written in pencil)
  - `<undo>` and `<redo>` → textual modifications that have been reversed and/or reinstated
  - `<transpose>` and `<transposeGrp>` → groups of text that have their order shifted
Genetic editing

<line>
not restrain
<mod change="#stage1">
<del>us.</del>
<add>our enquiries</add>
</mod>
I revolved these circum
</line>
Metamark

- Any kind of mark that guides the reading rather than being part of the text
  - Arrows, crosses, asterisks, other symbols
- More specificity can be added by using the attributes:
  - @function → to specify the function of the metamark
    - transposition, deletion, insertion, status, etc.
  - @target → identifies one or more elements to which the function indicated by the metamark applies
I am that halfgrown boy, fallen asleep, 

The tears of foolish passion yet undried 

Upon my cheeks. 

I pass through the travels and fortunes of those years as become old, 

Each in its due order comes and goes, 

As thus a message for me comes. 

The entered - Yes.
<undo> and <redo>

- An alteration that either has been
  - cancelled → <undo>
    - @target attribute points to the elements which are to be reverted
  - reinstated → <redo>
    - @target attribute points to the elements which are to be reasserted
<undo> and <redo>

This is just some sample text, we need a real example.

This is just some sample text, we need a real example.
Transpositions

- Passages that should be moved to a different position
- Often associated with metamarks indicating the change
- `<transpose>` element indicates a single textual transposition of at least two pointers (`<ptr>`), recording the order in which the elements should be read
- `<transpose>` must be included within a `<listTranspose>` element:
  - `<listTranspose>` groups a list of `<transpose>` elements occurring in the document. This list can be:
    - Embedded in the transcription
    - Part of the `<teiHeader>`, under `<profileDesc>`
Transpositions

2 This is the first line of my poem
1 And this is the second line
Or is it?

This is the first line of my poem
And this is the second line
Or is it?
Genetic editing — recording the process

- Genetic editing is concerned with order of composition
- Ascertaining that order is the editor’s interpretative act
- The editor can record their ‘best guess’ by indicating different stages of composition:
  - <listChanges> —> groups a list of revision phases
    - @order attribute can record whether the order of changes is significant (if known) or not (if unknown)
  - <change> —> describes a single revision phase
    - @xml:id —> uniquely identifies that phase
  - As the list of changes is metadata, it is part of <profileDesc><creation> in the <teiHeader>
Transpositions

This is the first line of my poem.

And this is the second line.

Or is it? Yes it is.
I want to know more!

- Chapter 11 of the TEI guidelines for more about representation of primary sources