Introduction to Course, Markup, and XML

James Cummings

August 2010
Introduction to the Course

Aims of Course

1. Examine the concept of markup and XML encoding
2. Provide hands-on experience in using TEI XML markup
3. Introduce the TEI scheme, its assumptions, and its organization
4. Survey the whole landscape of the TEI recommendations
5. Demonstrate how the TEI scheme may be customized to particular needs
6. Demonstrate some real world applications of the TEI scheme
7. Provide routes into more detailed information for exploration at your leisure
8. Provide opportunities for questions and discussions relating to your own encoding needs and priorities
Course Structure

The times of each lecture in the course aren’t written in stone and we will go as fast or as slow as students need, adapting as necessary. However the structure we will attempt to follow the structure below:

- Intro to the Course, Markup and XML
- Intro to TEI
- Intro to oXygen
- Ex1: oXygen and a basic TEI document
- Customising the TEI
- Ex2: Using Roma to customise the TEI
- Metadata: The TEI Header and Manuscript Description
- Ex3: Describing a Manuscript in TEI
- Names and Named Entities: people and places
- Ex4: Annotating people and places
- Linking texts and images: non hierarchic structures
- (Optional: Ex5: Marking-up facsimiles with the Image Markup Tool)
Course Materials

- All course materials including:
  - All slides from lectures
  - All exercises
  - All materials for the exercises

are available on the TEI @ Oxford website.

The url is: http://tei.oucs.ox.ac.uk/Oxford/2010-08-oxford/

All these materials are licensed with a Creative Commons Attribution license, which means they are freely available for re-use (though do let us know!)
After the workshop, if you have questions about:

- The workshop materials or teaching other workshops: tei@oucs.ox.ac.uk
- The TEI generally: TEI-L@listserv.brown.edu

If you mail the TEI-L mailing list it is better because:

- we’ll still try to answer as well as we would privately
- you get answers not only from us, but TEI experts around the world
- questions from those of all levels of ability stop the list becoming too technical
- everyone benefits from having the answers be public – and you benefit by reading (and sometimes answering!) others’ problems
In order to talk about texts, markup and encoding of texts, we need to understand what we mean by these basic concepts. When we talk about text encoding, what do we mean by a text? What is in a text and what assumptions do we make in reading them?
What’s in a text?

**The Scene:** A ship at sea; afterwards an uninhabited island.

**Act One**

Scene I. On a ship at sea; a tempestuous noise of thunder and lightning heard.

Enter a Shipmaster and a Boatswain.

Master. Boatswain!

Boats. Here, master; what cheer?

Master. Good! Speak to th' mariners; fall to 't yarely, or we run ourselves aground; bestir, bestir.

Enter Mariner.

Boats. Heigh, my hearts! cheerly, cheerly, my hearts! yare, yare! Take in the topsail. Tend to th' master's whistle. Blow till thou burst thy wind, if room enough.

Enter Alonso, Sebastian, Antonio, Ferdinand, Gonzalo, and Others.

Alon. Good boatswain, have care. Where's the master? Play the men.

Boats. I pray now, keep below.

Ant. Where is the master, boson?

Boats. Do you not hear him? You mar our labour; keep your cabins; you do

Cheerly, good hearts!—Out of our way, I say.

Gon. I have great comfort from this fellow. Methinks he hath no drowning mark upon him; his complexion is perfect galloways. Stand fast, good Fate, to his hanging; make the rope of his destiny our cable, for our own doth little advantage. If he be not born to be hang'd, our case is miserable.

[Exit.

Re-enter Boatswain.

Boats. Down with the topmast. Yare, lower, lower! Bring her to try wi' th' main-course. [A cry within] A plague upon this howling! They are louder than the weather or our office.

Re-enter Sebastian, Antonio, and Gonzalo.

Yet again! What do you here? Shall we give o'er, and drown? Have you a mind to sink?

Seb. A pox o' your throat, you bawling, blasphemous, incharitable dog!

Boats. Work you, then.

Ant. Hang, cur; hang, you whoreson, in-
What’s in a text (2)?

[Image of an ancient manuscript or document]
What's in a text (3)?

Hwæt wē Gār-Dēna in gēr-dagum
bēod-cyninga. brym gefrūnon,
hū ðā aþelingas ellen fremedon.

Oft Scyld Scēfing sceapenā þrēatum,
5 monegum mægðum meodo-setla oftēah;
egsode Eorl[e], sy∂ōan ðērest wearð
fēasceaf funden; hē ðæs frōfre gebād:

wēox under wolcnun, weorō-myndum þāh,
oðþæt him æghwylc þāra ymb-sittendra

ofe hrōn-rāde hyrān scolde,
A MONSEIGNEUR

Le Reverendissime Cardinal du Bellay,

S.

EV le Personnage, que tu joues au Spectacle de toute l’Europe, voyre de tout le Monde en ce grand Theatre Romain, veu tant d’affaires, et telz, que seul quasi tu sustiens à l’Honneur du Sacré Collège! pecheroy’-je pas (comme dit le Pindare Latin) contre le bien publicq’, si par longues paroles j’empechoy’ le temps, que tu donnes au service de ton Prince, au profit de la Patrie, et à l’accroissement de ton immortelle renommée? Epliye doncques quelque heurte de ce peu de re- lax, que tu prens pour respirer soubz le pesant faiz des affaires francoyses (charge vrayement digne de si robustes epaules, non moins que le Ciel de celles du grand Hercule), ma Muse a pris la hardiesse d’entrer au sacré Cabinet de tes saintes et studieuses oc-
Are these images of the same thing?

A MONSEIGNEUR

Le Peverefindissime Cardinal du Bellay, S.

Veu le personnage que tu joues au spectacle de toute l'Europe, voire de tout le monde, en ce grand théâtre romain: veu tant d'affaires et telz. que seul

Joachim du Bellay

Défense et illustration de la langue françoyse (1549)

La Défence, et Illustration de la Langue Françoys.

L'auteur prie les lecteurs d'effrayer leur jugement jusques à la fin du livre, et ne le condamner sans avoir premièrement bien vu, et examiné ses raisons.

Épitre à Monseigneur le réverendissime cardinal du Bellay S.

Vu le personnage que tu joues au spectacle de toute l'Europe, voire de tout le monde, en ce grand Théâtre Romain, veu tant d'affaires, et tels que seul quai tu soutiens à l'Honneur du sacré Collège, pècherais-je pas (comme dit le Pindare Latin) contre le bien public, si par longues paroles l'empeschoy le tens que tu donnes au service de ton Prince, au profit de la Patrie, et à l'accroissement de ton immortelle renommée?

Épist le douteux quelque heure de ce peu de relâchant, que tu prends pour respirer le peu de respiration que tu fais, si des affaires françoyses (chargeayement digne de si robustes épaules, non moins que le Ciel de celles du grand Hércule) ma Muse a pris la hardiesse d'entrer au sacré Cabinet de tes sainctes, et studieuses occupations; et la entre tant de

11/39
A text is not a document

Where is the text?

- in the shape of letters and their layout?
- in the original from which this copy derives?
- in the stories we read into it? or in its author’s intentions?

A "document" is something that exists in the world, which we can **digitize**.

A "text" is an abstraction, created by or for a community of readers, which we can **encode**.
A text is not a document

Where is the text?

- in the shape of letters and their layout?
- in the original from which this copy derives?
- in the stories we read into it? or in its author’s intentions?

A "document" is something that exists in the world, which we can digitize.

A "text" is an abstraction, created by or for a community of readers, which we can encode.
A text is not a document

Where is the text?

- in the shape of letters and their layout?
- in the original from which this copy derives?
- in the stories we read into it? or in its author’s intentions?

A "document" is something that exists in the world, which we can *digitize*.

A "text" is an abstraction, created by or for a community of readers, which we can *encode*. 
A text is not a document

Where is the text?

- in the shape of letters and their layout?
- in the original from which this copy derives?
- in the stories we read into it? or in its author’s intentions?

A "document" is something that exists in the world, which we can digitize.

A "text" is an abstraction, created by or for a community of readers, which we can encode.
Encoding of texts

- A text is more than a sequence of encoded glyphs or lexical tokens
  - It has a *structure* and a *communicative function*
  - It also has multiple possible *readings*
- Encoding, or markup, is a way of making these things explicit

Only that which is explicit can be reliably processed
What’s the point of markup?

- To make explicit (to a machine) what is implicit (to a person)
- To add value by supplying multiple annotations
- To facilitate re-use of the same material
  - in different formats
  - in different contexts
  - by different users
What’s the point of markup?

- To make explicit (to a machine) what is implicit (to a person)
- To add value by supplying multiple annotations
- To facilitate re-use of the same material
  - in different formats
  - in different contexts
  - by different users
What’s the point of markup?

- To make explicit (to a machine) what is implicit (to a person)
- To add value by supplying multiple annotations
- To facilitate re-use of the same material
  - in different formats
  - in different contexts
  - by different users
What’s the point of markup?

- To make explicit (to a machine) what is implicit (to a person)
- To add value by supplying multiple annotations
- To facilitate re-use of the same material
  - in different formats
  - in different contexts
  - by different users
In the beginning there was \textit{procedural} markup
\begin{verbatim}
  RED INK ON; print balance; RED INK OFF
\end{verbatim}
which being generalised became \textit{descriptive} markup
\begin{verbatim}
  <balance type='overdrawn'>some numbers</balance>
\end{verbatim}
also known as \textit{encoding} or \textit{annotation}

descriptive markup allows for easier re-use of data
Some more definitions

- Markup makes explicit the distinctions we want to make when processing a string of bytes.
- Markup is a way of naming and characterizing the parts of a text in a formalized way.
- It’s (usually) more useful to markup what we think things are than what they look like.
Separation of form and content

- Presentational markup cares more about fonts and layout than meaning.
- Descriptive markup says what things are, and leaves the rendition of them for a separate step.
- Separating the form of something from its content makes its re-use more flexible.
- It also allows easy changes of presentation across a large number of documents.
Separation of form and content

- Presentational markup cares more about fonts and layout than meaning
- Descriptive markup says what things are, and leaves the rendition of them for a separate step
- Separating the form of something from its content makes its re-use more flexible
- It also allows easy changes of presentation across a large number of documents
Separation of form and content

- Presentational markup cares more about fonts and layout than meaning
- Descriptive markup says what things are, and leaves the rendition of them for a separate step
- Separating the form of something from its content makes its re-use more flexible
- It also allows easy changes of presentation across a large number of documents
Separation of form and content

- Presentational markup cares more about fonts and layout than meaning
- Descriptive markup says what things are, and leaves the rendition of them for a separate step
- Separating the form of something from its content makes its re-use more flexible
- It also allows easy changes of presentation across a large number of documents
Separation of form and content

- Presentational markup cares more about fonts and layout than meaning
- Descriptive markup says what things are, and leaves the rendition of them for a separate step
- Separating the form of something from its content makes its re-use more flexible
- It also allows easy changes of presentation across a large number of documents
Markup as a scholarly activity

- The application of markup to a document can be an intellectual activity.
- In deciding what markup to apply, and how this represents the original, one is undertaking the task of an editor.
- There is (almost) no such thing as neutral markup – all of it involves interpretation.
- Markup can assist in answering research questions, and the deciding what markup is needed to enable such questions to be answered can be a research activity in itself.
- Good textual encoding is never as easy or quick as people would believe.
- Detailed document analysis is needed before encoding for the resulting markup to be useful.
The application of markup to a document can be an intellectual activity

In deciding what markup to apply, and how this represents the original, one is undertaking the task of an editor

There is (almost) no such thing as neutral markup – all of it involves interpretation

Markup can assist in answering research questions, and the deciding what markup is needed to enable such questions to be answered can be a research activity in itself

Good textual encoding is never as easy or quick as people would believe

Detailed document analysis is needed before encoding for the resulting markup to be useful
Markup as a scholarly activity

- The application of markup to a document can be an intellectual activity.
- In deciding what markup to apply, and how this represents the original, one is undertaking the task of an editor.
- There is (almost) no such thing as neutral markup – all of it involves interpretation.
- Markup can assist in answering research questions, and the deciding what markup is needed to enable such questions to be answered can be a research activity in itself.
- Good textual encoding is never as easy or quick as people would believe.
- Detailed document analysis is needed before encoding for the resulting markup to be useful.
Markup as a scholarly activity

- The application of markup to a document can be an intellectual activity.
- In deciding what markup to apply, and how this represents the original, one is undertaking the task of an editor.
- There is (almost) no such thing as neutral markup – all of it involves interpretation.
- Markup can assist in answering research questions, and the deciding what markup is needed to enable such questions to be answered can be a research activity in itself.
- Good textual encoding is never as easy or quick as people would believe.
- Detailed document analysis is needed before encoding for the resulting markup to be useful.
Markup as a scholarly activity

- The application of markup to a document can be an intellectual activity.
- In deciding what markup to apply, and how this represents the original, one is undertaking the task of an editor.
- There is (almost) no such thing as neutral markup – all of it involves interpretation.
- Markup can assist in answering research questions, and the deciding what markup is needed to enable such questions to be answered can be a research activity in itself.
- Good textual encoding is never as easy or quick as people would believe.
- Detailed document analysis is needed before encoding for the resulting markup to be useful.
Markup as a scholarly activity

- The application of markup to a document can be an intellectual activity
- In deciding what markup to apply, and how this represents the original, one is undertaking the task of an editor
- There is (almost) no such thing as neutral markup – all of it involves interpretation
- Markup can assist in answering research questions, and the deciding what markup is needed to enable such questions to be answered can be a research activity in itself
- Good textual encoding is never as easy or quick as people would believe
- Detailed document analysis is needed before encoding for the resulting markup to be useful
Markup as a scholarly activity

- The application of markup to a document can be an intellectual activity.
- In deciding what markup to apply, and how this represents the original, one is undertaking the task of an editor.
- There is (almost) no such thing as neutral markup – all of it involves interpretation.
- Markup can assist in answering research questions, and the deciding what markup is needed to enable such questions to be answered can be a research activity in itself.
- Good textual encoding is never as easy or quick as people would believe.
- Detailed document analysis is needed before encoding for the resulting markup to be useful.
What does markup capture?

Compare

```xml
<hi rend="dropcap">H</hi>&amp;WYN;ÆT WE GARDE <lb/>na in gear-dagum þeod-cyninga <lb/>þrym gefrunon, hu ða æþelingas <lb/>ellen fremedon. oft scyld scefing sceaph <add>na</add> <lb/>þreatum, moneg<expan>um</expan> mægþum meodo-setl <add>a</add> <lb/>of<damage> <desc>blot</desc> </damage>teah ...
```

and

```xml
<lg>
 <l>Hwæt! we Gar-dena in gear-dagum</l>
 <l>þeod-cyninga þrym gefrunon,</l>
 <l>hu ða æþelingas ellen fremedon,</l>
</lg>

<lg>
 <l>Oft Scyld Scefing sceaphena þreatum,</l>
 <l>monegum mægþum meodo-setla ofteah;</l>
 <l>egsode Eorle, syððan ærest wearþ</l>
 <l>feasceafth funden...</l>
</lg>
```
A useful mental exercise

Imagine you are going to markup several thousand pages of complex material....

- Which features are you going to markup?
- Why are you choosing to markup this feature?
- How reliably and consistently can you do this?

Now, imagine your budget has been halved. Repeat the exercise!
Some alphabet soup

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGML</td>
<td>Standard Generalized Markup Language</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
<tr>
<td>DTD</td>
<td>Document Type Definition (or Declaration)</td>
</tr>
<tr>
<td>CSS</td>
<td>Cascading Style Sheet</td>
</tr>
<tr>
<td>Xpath</td>
<td>XML Path Language</td>
</tr>
<tr>
<td>XSLT</td>
<td>eXtensible Stylesheet Language - Transformations</td>
</tr>
<tr>
<td>XQuery</td>
<td>XML Querying</td>
</tr>
<tr>
<td>RELAXNG</td>
<td>Regular Expression Language for XML (New Generation)</td>
</tr>
</tbody>
</table>

Oh, and then there’s also **TEI**, the *Text Encoding Initiative*
Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML also now plays an indispensable role in the exchange of a wide variety of data on the Web and elsewhere.
XML: what it is and why you should care

- XML is **structured data** represented as strings of text
- XML looks like HTML, except that:-
  - XML is **extensible**
  - XML must be **well-formed**
  - XML can be **validated**
- XML is application-, platform-, and vendor- independent
- XML empowers the **content provider** and facilitates data integration
XML terminology

An XML document may contain:-

- elements, possibly bearing attributes
- processing instructions
- comments
- entity references
- marked sections (CDATA, IGNORE, INCLUDE)

An XML document must be well-formed and may be valid
XML terminology Example

```xml
<?xml version="1.0" ?>
<root>
  <element attribute="value"> content </element>
  <!-- comment -->
</root>
```
The rules of the XML Game

- An XML document represents a (kind of) tree
- It has a single root and many nodes
- Each node can be
  - a subtree
  - a single element (possibly bearing some attributes)
  - a string of character data
- Each element has a name or generic identifier
The rules of the XML Game

- An XML document represents a (kind of) tree
- It has a single root and many nodes
- Each node can be
  - a subtree
  - a single element (possibly bearing some attributes)
  - a string of character data
- Each element has a name or generic identifier
The rules of the XML Game

- An XML document represents a (kind of) tree
- It has a single root and many nodes
  - Each node can be
    - a subtree
    - a single element (possibly bearing some attributes)
    - a string of character data
- Each element has a name or generic identifier
The rules of the XML Game

- An XML document represents a (kind of) tree
- It has a single root and many nodes
- Each node can be
  - a subtree
  - a single element (possibly bearing some attributes)
  - a string of character data
- Each element has a name or generic identifier
The rules of the XML Game

- An XML document represents a (kind of) tree
- It has a single root and many nodes
- Each node can be
  - a subtree
    - a single element (possibly bearing some attributes)
    - a string of character data
- Each element has a name or generic identifier
The rules of the XML Game

- An XML document represents a (kind of) tree
- It has a single root and many nodes
- Each node can be
  - a subtree
  - a single element (possibly bearing some attributes)
  - a string of character data
- Each element has a name or generic identifier
The rules of the XML Game

- An XML document represents a (kind of) tree
- It has a single root and many nodes
- Each node can be
  - a subtree
  - a single element (possibly bearing some attributes)
  - a string of character data
- Each element has a name or generic identifier
An XML document represents a (kind of) tree

- It has a single root and many nodes
- Each node can be
  - a subtree
  - a single element (possibly bearing some attributes)
  - a string of character data

- Each element has a name or generic identifier
Representing an XML tree

- An XML document is encoded as a linear string of characters.
- It begins with a special **processing instruction**.
- Element occurrences are marked by **start-** and **end-tags**.
- The characters `<` and `&` are Magic and must always be "escaped" if you want to use them as themselves.
- Comments are delimited by `<!- - and -->
- **CDATA sections** are delimited by `<![CDATA[ and ]]>`.
- Attribute name/value pairs are supplied on the start-tag and may be given in any order.
- Entity references are delimited by `&` and `;`.
Parts of an XML document

```xml
<?xml version="1.0"?>
<greetings xmlns="http://www.example.org/greetings">
  <hello xmlns="http://www.example.org/greetings" type="sarcastic">
    hello world!
  </hello>
</greetings>
```

- The XML declaration
- Namespace declarations
- The root element of the document itself
- Other elements and content
- Attribute and value
Parts of an XML document

- The XML declaration
  - Namespace declarations
  - The root element of the document itself
  - Other elements and content
  - Attribute and value

```xml
<?xml version="1.0"?>
<greetings xmlns="http://www.example.org/greetings">
  <hello xmlns="http://www.example.org/greetings" type="sarcastic">hello world!</hello>
</greetings>
```
Parts of an XML document

- The XML declaration
- Namespace declarations
  - The root element of the document itself
  - Other elements and content
  - Attribute and value

```xml
<?xml version="1.0"?>
<greetings xmlns="http://www.example.org/greetings">
  <hello xmlns="http://www.example.org/greetings" type="sarcastic">hello world!</hello>
</greetings>
```
Parts of an XML document

```
<?xml version="1.0"?>
<greetings xmlns="http://www.example.org/greetings">
  <hello xmlns="http://www.example.org/greetings" type="sarcastic">hello world!</hello>
</greetings>
```

- The XML declaration
- Namespace declarations
- The root element of the document itself
- Other elements and content
- Attribute and value
Parts of an XML document

```
<?xml version="1.0"?><
greetings xmlns="http://www.example.org/greetings">
  <hello xmlns="http://www.example.org/greetings" type="sarcastic">hello world!</hello>
</greetings>
```

- The XML declaration
- Namespace declarations
- The root element of the document itself
- Other elements and content
- Attribute and value
Parts of an XML document

- The XML declaration
- Namespace declarations
- The root element of the document itself
- Other elements and content
- Attribute and value

```xml
<?xml version="1.0"?>
<greetings xmlns="http://www.example.org/greetings">
  <hello xmlns="http://www.example.org/greetings" type="sarcastic">hello world!</hello>
</greetings>
```
The XML declaration

An XML document must begin with an XML declaration which does three things:

- specifies that this is an XML document
- specifies which version of the XML standard it follows
- specifies which character encoding the document uses

```xml
<?xml version="1.0" ?>
<?xml version="1.0" encoding="iso-8859-1" ?>
```

The default, and recommended, encoding is ‘UTF-8’ (Unicode)
Namespace declarations

All TEI documents are declared within the TEI namespace: `<TEI xmlns="http://www.tei-c.org/ns/1.0"> ... </TEI>`

XML documents can include elements declared in different namespaces.

- A namespace declaration associates a namespace prefix with an external URI-like identifier.
- The default namespace may be declared using a `xmlns`.
- Other namespaces must all use a specially declared prefix.

```xml
<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:math="http://www.mathml.org">
<p>...<math:expr>...</math:expr>...</p>...</TEI>
```

The `xml` namespace is used by the TEI for global attributes `@xml:id` and `@xml:lang`
The Doctype Declaration

You may sometimes find an optional "Document Type" declaration at the start of a document:

```xml
<?xml version="1.0" ?>
<!DOCTYPE greeting SYSTEM "greeting.dtd []">
```

- The DTD is one way of associating the document with its schema (but is not used by W3C or RELAX NG for this purpose)
- The DTD subset is used to provide declarations additional to those in the schema, for example for external files
- The DTD subset may be internal, external, or both

DTDs are now considered old-fashioned – RELAX NG schemas are preferred.
<?xml version="1.0" encoding="utf-8" ?>
<div n="1">
  <head>SCENE I. On a ship at sea: a tempestuous noise of thunder and lightning heard.</head>
  <stage>Enter a Master and a Boatswain</stage>
  <sp>
    <speaker>Master</speaker>
    <ab>Boatswain!</ab>
  </sp>
  <sp>
    <speaker>Boatswain</speaker>
    <ab>Here, master: what cheer?</ab>
  </sp>
  <sp>
    <speaker>Master</speaker>
    <ab>Good, speak to the mariners: fall to't, yarely,</ab>
    <ab>or we run ourselves aground: bestir, bestir.</ab>
  </sp>
  <stage>Exit</stage>
</div>
Example deconstructed: root node

```xml
<?xml version="1.0" encoding="utf-8" ?>
<div n="1">
<!-- .... -->
</div>
</div>
```
Example deconstructed: head

<head>SCENE I. On a ship at sea: a tempestuous noise of thunder and lightning heard.</head>
Example deconstructed: stage direction and speech

<stage>Enter a Master and a Boatswain</stage>
<sp>
  <speaker>Master</speaker>
  <ab>Boatswain!</ab>
</sp>
An XML Tree For The Tempest
What does it mean to be **well-formed**?

1. there is a single root node containing the whole of an XML document
2. each subtree is properly nested within the root node
3. names are always case sensitive
4. start-tags and end-tags are always mandatory (except that a combined start-and-end tag may be used for empty nodes)
5. attribute values are always quoted

Note: You can be **valid** in addition to being well-formed. This means you obey the rules of a specified schema, such as the TEI.
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>`
- `<seg><foo>some</foo> <bar>text</bar></seg>`
- `<seg><foo>some <bar></foo> text</bar></seg>`
- `<seg type="text">some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type="text">some text<seg/>`
- `<seg type="text">some text<gap/></seg>`
- `<seg type="text">some text</Seg>`
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>
- `<seg><foo>some</foo> <bar>text</bar></seg>
- `<seg><foo>some <bar></foo> text</bar></seg>
- `<seg type="text">some text</seg>
- `<seg type='text'>some text</seg>
- `<seg type=text>some text</seg>
- `<seg type="text">some text</seg>`
- `<seg type="text">some text<seg/>
- `<seg type="text">some text<gap/></seg>
- `<seg type="text">some text</Seg>`
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>`
- `<seg><foo>some</foo> <bar>text</bar></seg>`
- `<seg><foo>some <bar>text</bar></seg>`
- `<seg type="text">some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type="text">some text<seg/>`
- `<seg type="text">some text<gap/></seg>`
- `<seg type="text">some text</Seg>`
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>`
- `<seg><foo>some</foo> <bar>text</bar></seg>`
- `<seg><foo>some <bar></foo> text</bar></seg>`
- `<seg type="text">some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type="text">some text<seg/>`
- `<seg type="text">some text<gap/></seg>`
- `<seg type="text">some text</Seg>`
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>`
- `<seg><foo>some</foo> <bar>text</bar></seg>`
- `<seg><foo>some <bar></foo> text</bar></seg>`
- `<seg type="text">some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type="text">some text<seg/>`
- `<seg type="text">some text<gap/></seg>`
- `<seg type="text">some text</Seg>`
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>`
- `<seg><foo>some</foo> <bar>text</bar></seg>`
- `<seg><foo>some <bar></foo> text</bar></seg>`
- `<seg type="text">some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type="text">some text</seg>`
- `<seg type="text">some text<seg/>`
Test your XML knowledge

- Which are correct?
  - `<seg>some text</seg>`
  - `<seg><foo>some</foo> <bar>text</bar></seg>`
  - `<seg><foo>some <bar></foo> text</bar></seg>`
  - `<seg type="text">some text</seg>`
  - `<seg type='text'>some text</seg>`
  - `<seg type=text>some text</seg>`
  - `<seg type="text">some text<seg/>`
  - `<seg type="text">some text<gap/></seg>`
  - `<seg type="text">some text</Seg>`
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>`
- `<seg><foo>some</foo> <bar>text</bar></seg>`
- `<seg><foo>some </bar><foo> text</bar></seg>`
- `<seg type="text">some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type="text">some text<seg>`
- `<seg type="text">some text<gap/></seg>`
- `<seg type="text">some text</Seg>`
Test your XML knowledge

Which are correct?

- `<seg>some text</seg>`
- `<seg><foo>some</foo> <bar>text</bar></seg>`
- `<seg><foo>some <bar></foo> text</bar></seg>`
- `<seg type="text">some text</seg>`
- `<seg type='text'>some text</seg>`
- `<seg type=text>some text</seg>`
- `<seg type="text">some text<seg/>`
- `<seg type="text">some text<gap/></seg>`
- `<seg type="text">some text</Seg>`
Test your XML knowledge

- Which are correct?
  - `<seg>some text</seg>`
  - `<seg><foo>some</foo> <bar>text</bar></seg>`
  - `<seg><foo>some <bar></foo> text</bar></seg>`
  - `<seg type="text">some text</seg>`
  - `<seg type='text'>some text</seg>`
  - `<seg type=text>some text</seg>`
  - `<seg type="text">some text<seg/>`
  - `<seg type="text">some text<gap/></seg>`
  - `<seg type="text">some text</Seg>`
XML is an international standard

- XML requires use of ISO 10646 (also known as Unicode)
  - a 31 bit character repertoire including most human writing systems
  - encoded as UTF8 or UTF16
- other encodings may be specified at the document level
- language may be specified at the element level using \texttt{@xml:lang}

The \texttt{@xml:id} attribute is another W3C-defined attribute.