



ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

ENRICHing Electronic Manuscripts with TEI P5 XML

Dr James Cummings

January 2009



Introduction

*ENRICH*ing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

- A brief mention of the TEI and its basic structures
- A discussion of TEI elements for transcription and manuscript description
- A quick mention of the ENRICH project

Overall it is more pedagogic than academic or theoretic



What the TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas, documentation, and texts in various languages and formats
- a large and active open source style user community
- increasing acceptance by major funding bodies

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



What the TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas, documentation, and texts in various languages and formats
- a large and active open source style user community
- increasing acceptance by major funding bodies

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



What the TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas, documentation, and texts in various languages and formats
- a large and active open source style user community
- increasing acceptance by major funding bodies

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



What the TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas, documentation, and texts in various languages and formats
- a large and active open source style user community
- increasing acceptance by major funding bodies

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



What the TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas, documentation, and texts in various languages and formats
- a large and active open source style user community
- increasing acceptance by major funding bodies

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



What the TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas, documentation, and texts in various languages and formats
- a large and active open source style user community
- increasing acceptance by major funding bodies

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



What the TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas, documentation, and texts in various languages and formats
- a large and active open source style user community
- increasing acceptance by major funding bodies

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



Three important things about TEI P5

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

- 1 Being a good digital citizen:
 - Support for multiple schema languages and namespaces
 - Reliance on XML, and hence on Unicode
 - Validation of attributes and datatyping
 - Use of W3C pointers and ISO standards
- 2 Making it flexible:
 - ODD: a single specification language for developers, users, and teachers, integrating schema and documentation;
 - Verifiable conformance
- 3 Old annoyances removed and some new topics added



Modules

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

Module name	Chapter
analysis	Simple Analytic Mechanisms
certainty	Certainty and Responsibility
core	Elements Available in All TEI Documents
corpus	Language Corpora
dictionaries	Dictionaries
drama	Performance Texts
figures	Tables, Formulae, and Graphics
gaiji	Representation of Non-standard Characters and Glyphs
header	The TEI Header
iso-fs	Feature Structures
linking	Linking, Segmentation, and Alignment
msdescription	Manuscript Description
namesdates	Names, Dates, People, and Places
nets	Graphs, Networks, and Trees
spoken	Transcriptions of Speech
tagdocs	Documentation Elements
tei	The TEI Infrastructure
textcrit	Critical Apparatus
textstructure	Default Text Structure
transcr	Representation of Primary Sources
verse	Verse



TEI Macrostructure

As well as the texts, we have detailed metadata about each volume, and images of its pages. These are the three parts of a canonical TEI document:

```
<TEI>
  <teiHeader>
    <!-- required; provides metadata -->
  </teiHeader>
  <facsimile>
    <!-- the text, represented in image form -->
  </facsimile>
  <text>
    <!-- the text, transcribed and marked up -->
  </text>
</TEI>
```

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



TEI Header Structure

The TEI header has four main components:

- `<fileDesc>` (file description) contains a full bibliographic description of an electronic file.
- `<encodingDesc>` (encoding description) documents the relationship between an electronic text and the source or sources from which it was derived.
- `<revisionDesc>` (revision description) summarizes the revision history for a file.
- `<profileDesc>` (text-profile description) provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting. (just about everything not covered in the other header elements)

Only `<fileDesc>` is required; the others are optional.



Below the paragraph...

Inside the `<text>` element people are able to structure things into divisions (`<div>` element) and paragraphs (`<p>`) amongst other options. Inside these we have phrase level elements for mark-up of smaller components. For example:

- boundaries, such as page, column, or line breaks
- highlighting, emphasis and quotation
- editorial changes such as correction, normalization etc.
- names, numbers, dates, addresses...
- links and cross-references
- notes, annotation, indexing
- graphics
- bibliographic citations
- words and other analyses



For Example: Highlighting

By highlighting we mean any combination of typographic features (font, size, hue, etc.) which distinguishes the highlighted text from its surroundings. This may be for many reasons...

- to mark foreign, archaic, technical usages
- for emphasis when spoken
- to show something is not part of the text.. (e.g. cross references, titles, headings)
- or is attributed to some other agency inside or outside the text (e.g. direct speech, quotation)

TEI provides both a generic `<hi>` tag and a large number of specific ones...

- `<hi>` (highlighted: reason unknown or unimportant)

```
<p>[The rest of this communication is
omitted owing to considerations of
space. -<hi rend="sc">Ed</hi>.] </p>
```

- `<emph>` (emphasized)

```
<said>'E won't bite yer <emph>if you buy 'im</emph>
guv'ner. </said>
```

- `<title>` and `<foreign>`:

```
<p>
<foreign xml:lang="fr">À propos</foreign> of Oxford, it is a
question whether that extremely amusing book
<title>Verdant Green</title> is still much read by freshers.
</p>
```

- `<distinct>` (linguistically marked)

```
But then I remind myself
that the Russian ballet is nothing if not
<distinct>bizarre</distinct>
```




TEI: Transcription

Transcription needs careful encoding, as the aim is to represent all the important features of a primary source without prejudging too much about it... i.e. diplomatic transcript.

Here are some of the kinds of features concerned:

- letter forms
- page layout
- orthography
- word division
- punctuation
- abbreviations
- additions and deletions
- errors and omissions

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



Abbreviation

Abbreviations are highly characteristic of manuscript materials of all kinds. Western MSS traditionally distinguish:

- Suspensions** the first letter or letters of the word are written, generally followed by a point, or other marker: for example **e.g.** for **exempla gratia**
- Contractions** both first and last letters are written, generally with some other mark of abbreviation such as a superscript stroke, or, less commonly, a point or points: e.g. **Mr.** for **Mister**
- Brevigraphs** Special signs or tittels, such as the Tironian nota used for 'et', the letter p with a barred tail commonly used for **per**, the letter c with a circumflex used for **cum** (ĉ) etc.
- Superscripts** Superscript letters (vowels or consonants) are often used to indicate various kinds of contraction: e.g. **w** followed by superscript **ch** for **which**.



Encoding abbreviations (1)

TEI proposes two levels of encoding:

- the whole of an abbreviated word and the whole of its expansion: `<abbr>` and `<expan>`
- abbreviatory signs or characters and the 'invisible' characters they imply: `<am>` and `<ex>`

The Old Icelandic word *hann* ('he') is usually written as a brevigraph in medieval manuscripts, combining the letter **h** with a horizontal stroke representing nasalisation (Unicode character 0305, functionally similar to the modern tilde). It looks like this:





Encoding abbreviations (2)

Depending on editorial policy, we might represent this combination in any one of the following ways:

```
<abbr>h&#x305; </abbr>
```

```
<expan>hann</expan>
```

```
h<am>&#x305; </am>
```

```
h<ex>ann</ex>
```

```
<abbr>h<am>&#x305; </am>  
</abbr>
```

```
<expan>h<ex>ann</ex>  
</expan>
```



Encoding abbreviations (3)

We could also indicate multiple alternatives (at either level) by using the `<choice>` element

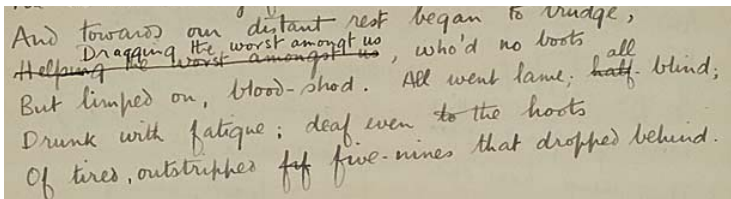
```
h<choice>  
  <am>&#x305; </am>  
  <ex>ann</ex>  
</choice>
```

or

```
<choice>  
  <abbr>h&#x305; </abbr>  
  <expan>hann</expan>  
</choice>
```

And many more possibilities...

- `<add>` (addition) or `` (deletion) are used for evident alterations in the source
- a combined addition and deletion may be marked using `<subst>` (substitution)



And towards our ^{distant} rest began to vudge,
~~Helping the worst amongst us~~ ^{Dragging the worst amongst us}, who'd no boots all
But limped on, blood-shod. All went lame; half-blind;
Drunk with fatigue; deaf even to the hoots
Of tired, outstripped ~~five~~ ^{five}-nines that dropped behind.



Additions, deletions, and substitutions

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

```
<l>And towards our distant rest began to trudge, </l>
<l>
  <subst>
    <del>Helping the worst amongst us</del>
    <add>Dragging the worst amongst us</add>
  </subst>, who'd no boots
</l>
<l>But limped on, blood-shod. All went lame;
<subst>
  <del>half-</del>
  <add>all</add>
</subst> blind; </l>
<l>Drunk with fatigue ; deaf even to the hoots</l>
<l>Of tired, outstripped <del>fif</del> five-nines that dropped behind. </l>
```




Supplied text

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

Sometimes, a transcript may need to include words not visibly present in the source:

- because the carrier has been damaged or is barely legible
- because of (assumed) scribal error

The `<supplied>` element is provided for use in either situations; the `@reason` attribute is used to distinguish them.

```
...Dragging the worst  
among<supplied reason="omitted">s</supplied>t us...
```



Corrections and emendations

The `<sic>` element can be used to indicate that the reading of the manuscript is erroneous or nonsensical, while `<corr>` (correction) can be used to provide what in the editor's opinion is the correct reading:

```
<sic>giorit</sic>
```

```
<corr>giorir</corr>
```

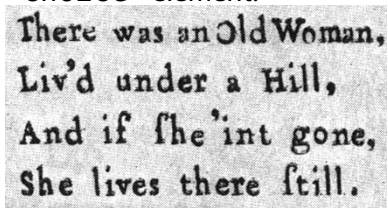
Alternatively, they may be combined within a `<choice>` element, thus allowing the possibility of providing multiple corrections:

```
<choice>  
  <sic>giorit</sic>  
  <corr cert="high">giorir</corr>  
  <corr cert="low">gioret</corr>  
</choice>
```



Normalization

Source texts rarely use modern normalized orthography. For retrieval and other processing reasons, such information may be useful in a transcription. The `<r eg>` (regularized) element is available used to mark a normalized form; the `<or ig>` (original) element to indicate a non-standard spelling. These elements can optionally be grouped as alternatives using the `<choice>` element:



There was an Old Woman,
Liv'd under a Hill,
And if she'nt gone,
She lives there still.

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



Normalization example

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

```
<lg>
  <l>There was an Old Woman, </l>
  <l>
    <choice>
      <orig>Liv'd</orig>
      <reg>Lived</reg>
    </choice> under a hill, </l>
  <l>And if she <orig>'int</orig> gone, </l>
  <l>She lives there still. </l>
</lg>
```



Critical Apparatus: <app>, <rdg>, and <lem>

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

- <app> (apparatus entry) contains one entry in a critical apparatus, with an optional lemma and at least one reading.
- <rdg> (reading) contains a single reading within a textual variation.
- <lem> (lemma) contains the lemma, or base text, of a textual variation.

Example of <app>, <rdg> and <lem>

```

<app>
  <lem wit="#E1">Experience though noon Auctoritee</lem>
  <rdg wit="#La">Experiment thogh noon Auctoritee</rdg>
  <rdg wit="#Ra2">Eryment though none auctorite</rdg>
</app>

```

or

```

<app>
  <lem wit="#E1 #Hg">Experience</lem>
  <rdg type="substantive" wit="#La">Experiment</rdg>
  <rdg type="substantive" wit="#Ra2">Eryment</rdg>
</app>
<app>
  <lem wit="#E1 #Ra2">though</lem>
  <rdg type="orthographic" wit="#Hg">thogh</rdg>
  <rdg type="orthographic" wit="#La">thouh</rdg>
</app>

```



A Simple <app> With No <lem>

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

```
<p type="antiphon"> Populus domini et oves pascuae eius  
<app>  
  <rdg  
    wit="#CA0-B #CA0-V #CA0-R #CA0-D #CA0-F #CA0-S #Ely  
    #Wor #Wcb"> venite adoremus eum</rdg>  
  <rdg wit="#CA0-H #Pet"> venite adoremus deum</rdg>  
  <rdg wit="#CA0-E #Alb2"> venite adoremus dominum</rdg>  
  <rdg wit="#CA0-C #CA0-G #CA0-L #Hyd #Evm"> venite  
    adoremus</rdg>  
</app>  
</p>
```



TEI: Manuscript Description

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

Why are manuscript descriptions special?

- Manuscripts are *unique objects*, sometimes (though not always) of great cultural or political value
- Books, by contrast, often exist in multiple copies, and can be usually described adequately by well-established and formalized bibliographic conventions.
- For manuscripts, there are several traditions, often descriptive or **belle lettriste**, and little consensus.

Similar concerns sometimes apply to other text-bearing objects.



Digital edition

- metadata in the header
- transcription in the body, with links to
- images in a `<facsimile>` element

```
<TEI>
  <teiHeader>
    <!-- ... metadata describing the manuscript -->
    <!-- includes a msDesc within the sourceDesc -->
  </teiHeader>
  <facsimile>
    <!-- ... metadata describing the digital images -->
  </facsimile>
  <text>
    <!-- (optional) transcription of the manuscript -->
  </text>
</TEI>
```



msDesc structure

```
<msDesc xml:id="ex2" xml:lang="en">
  <msIdentifier>
    <!-- Repository location, shelfmarks, etc. -->
  </msIdentifier>
  <msContents>
    <!-- Structured description of MS contents -->
  </msContents>
  <physDesc>
    <!-- Physical and codicological description -->
  </physDesc>
  <history>
    <!-- Origin, provenance, acquisition, etc. -->
  </history>
  <additional>
    <!-- Additional bibliographic and curatorial information, and associated
    materials etc. -->
  </additional>
  <msPart>
    <!-- Composite manuscript details -->
  </msPart>
</msDesc>
```

`<msIdentifier>` is the only one that is required.

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



Simple example <msDesc>

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

```
<msDesc xml:id="ex3" xml:lang="en">
  <msIdentifier>
    <settlement>Oxford</settlement>
    <repository>Bodleian Library</repository>
    <idno>MS. Add. A. 61</idno>
    <altIdentifier type="other">
      <idno>28843</idno>
    </altIdentifier>
  </msIdentifier>
  <p>In Latin, on parchment: written in more than one hand of the 13th
cent. in England: 7¼ x 5⅜ in., i + 55 leaves, in double columns: with
a few coloured capitals. </p>
  <p>'Hic incipit Bruitus Anglie,' the De
origine et gestis Regum Angliae of Geoffrey of Monmouth (Galfridus
Monumetensis: beg. 'Cum mecum multa & de multis.' </p>
  <p>On fol. 54v very faint is 'Iste liber est fratris guillelmi de
buria de ... Roberti ordinis fratrum Pred[icatorum],' 14th cent. (?):
'hanauilla' is written at the foot of the page (15th cent.). Bought
from the rev. W. D. Macray on March 17, 1863, for £1 10s. </p>
</msDesc>
```



Structured form of <msDesc> (1)

```
<msDesc xml:id="ex4" xml:lang="en">
  <msIdentifier>
    <settlement>Oxford</settlement>
    <repository>Bodleian Library</repository>
    <idno>MS. Add. A. 61</idno>
    <altIdentifier type="internal">
      <idno>28843</idno>
    </altIdentifier>
  </msIdentifier>
  <msContents>
    <msItem>
      <author xml:lang="en">Geoffrey of Monmouth</author>
      <author xml:lang="la">Galfridus Monumetensis</author>
      <title type="uniform" xml:lang="la">De origine et gestis Regum
Angliae</title>
      <rubric xml:lang="la">Hic incipit Bruitus Anglie</rubric>
      <incipit xml:lang="la">Cum mecum multa & de multis</incipit>
      <textLang mainLang="la">Latin</textLang>
    </msItem>
  </msContents>
<!-- ... -->
</msDesc>
```

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



Structured form of <msDesc> (2)

```
<physDesc>
  <objectDesc form="codex">
    <supportDesc material="perg">
      <support>
        <p>Parchment. </p>
      </support>
      <extent>i + 55 leaves <dimensions scope="all" type="leaf" unit="in">
        <height>7  $\frac{1}{4}$ </height>
        <width>5  $\frac{3}{8}$ </width>
      </dimensions>
    </extent>
  </supportDesc>
  <layoutDesc>
    <layout columns="2">
      <p>In double columns. </p>
    </layout>
  </layoutDesc>
</objectDesc>
<handDesc>
  <p>Written in more than one hand. </p>
</handDesc>
<decoDesc>
  <p>With a few coloured capitals. </p>
</decoDesc>
</physDesc>
```



Structured form of <msDesc> (3)

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

```
<history>
  <origin>
    <p>Written in <origPlace>England</origPlace> in the
    <origDate notAfter="1300" notBefore="1200">13th cent.</origDate>
    </p>
  </origin>
  <provenance>
    <p>On fol. 54v very faint is <quote xml:lang="la">Iste liber est
    fratris guillelmi de buria de
    <gap reason="illegible"/> Roberti ordinis
    fratrum Pred<ex>icatorum</ex>
    </quote>, 14th cent. (?):
    <quote>hanauilla</quote> is written at the foot of
    the page (15th cent.).</p>
  </provenance>
  <acquisition>
    <p>Bought from the rev. <name type="person" key="MCRAYWD">W. D.
    Macray</name> on
    <date when="1863-03-17">March 17,
    1863</date>, for £1 10s.</p>
  </acquisition>
</history>
```



Intellectual Content

- May simply use paragraphs of text...
- ... or a tree of `<msItem>` elements
- ... optionally preceded by a prose summary

We can describe the content in general terms:

```
<msContents>  
  <p>An extraordinary charivari of heroic deeds and  
    improving tales, including an early version of  
  <title>Guy of Warwick</title> and several hymns.  
  </p>  
</msContents>
```

or we can provide detail about each distinct item:

```
<msContents>  
  <summary>An extraordinary charivari of heroic deeds,  
    improving tales, and hymns</summary>  
  <msItem>  
    <!-- details of Guy of Warwick here -->  
  </msItem>  
  <msItem>  
    <!-- other items here -->  
  </msItem>  
</msContents>
```



The carrier 1

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

The `<objectDesc>` contains just paragraphs, or `<supportDesc>` and `<layoutDesc>`

```
<objectDesc form="codex">
  <supportDesc material="mixed">
    <p>Early modern <material>parchment</material> and
    <material>paper</material>. </p>
  </supportDesc>
  <layoutDesc>
    <layout columns="1" ruledLines="25 32"/>
  </layoutDesc>
</objectDesc>
```




The carrier 2

A more complex substructure with specific elements for `<support t>`, `<extent>`, `<foliation>`, `<collation>`, `<condition>`.

Multiple layouts may also be specified:

```
<layoutDesc>
  <layout ruledLines="25" columns="1">
    <p>
      <locus from="1r-202v"/>
      <locus from="210r-212v"/>
      Between 25 and 32 ruled lines. </p>
    </layout>
  <layout ruledLines="34 50" columns="1">
    <p>
      <locus from="203r-209v"/>Between 34 and 50 ruled lines. </p>
    </layout>
  </layoutDesc>
```

<handDesc> examples

```
<handDesc hands="2">
```

```
<p>The manuscript is written in two contemporary hands, otherwise unknown, but clearly those of practised scribes. Hand I writes ff. 1r-22v and hand II ff. 23 and 24. Some scholars, notably Verner Dahlerup and Hreinn Benediktsson, have argued for a third hand on f. 24, but the evidence for this is insubstantial.</p>
```

```
</handDesc>
```

```
<handDesc hands="3">
```

```
<handNote xml:id="Eirsp-1" scope="minor" script="other">
```

```
<p>The first part of the manuscript, <locus from="1v" to="72v:4">fols 1v-72v:4</locus>, is written in a practised Icelandic Gothic bookhand. This hand is not found elsewhere.</p>
```

```
</handNote>
```

```
<handNote xml:id="Eirsp-2" scope="major" script="other">
```

```
<p>The second part of the manuscript, <locus from="72v:4" to="194v">fols 72v:4-194</locus>, is written in a hand contemporary with the first; it can also be found in a fragment of <title>Knýtlinga saga</title>, <ref>AM 20b II fol.</ref>.</p>
```

```
</handNote>
```

```
<handNote xml:id="Eirsp-3" scope="minor" script="other">
```

```
<p>The third hand has written the majority of the chapter headings. This hand has been identified as the one also found in <ref>AM 221 fol.</ref>.</p>
```

```
</handNote>
```

```
</handDesc>
```



<additions>

The `<additions>` element can be used to list or describe any additions to the manuscript, such as marginalia, scribblings, doodles, etc., which are considered to be of interest or importance.

`<additions>`

`<p>`The text of this manuscript is not interpolated with sentences from Royal decrees promulgated in 1294, 1305 and 1314. In the margins, however, another somewhat later scribe has added the relevant paragraphs of these decrees, see pp. 8, 24, 44, 47 etc. `</p>`

`<p>`As a humorous gesture the scribe in one opening of the manuscript, pp. 36 and 37, has prolonged the lower stems of one letter f and five letters p and has them drizzle down the margin. `</p>`

`</additions>`



<accMat>

<accMat> (accompanying material) contains details of any significant additional material which may be closely associated with the manuscript being described, such as non-contemporaneous documents or fragments bound in with the manuscript at some earlier historical period.

<accMat> A copy of a tax form from 1947 is included in the envelope with the letter. It is not catalogued separately. **</accMat>**

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



History Example

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

```
<history>
  <origin>
    <p>Written in <origPlace>England</origPlace> in the
    <origDate notAfter="1300" notBefore="1200">13th
      cent. </origDate>
    </p>
  </origin>
  <provenance>
    <p>On fol. 54v very faint is <q>Iste liber
      est fratris guillelmi de buria de <gap reason="illegible"/>
      Roberti ordinis fratrum Pred<expan>icatorum</expan>
    </q>,
      14th cent. (?): <q>hanauilla</q> is written at the
      foot of the page (15th cent.). </p>
  </provenance>
  <acquisition>
    <p>Bought from the rev. <name type="person">W. D. Macray</name> on
    <date when="1863-03-17"> March 17, 1863</date>,
      for 1pound 10s. </p>
  </acquisition>
</history>
```



Administrative information

- record history
- availability
- custodial history
- miscellaneous remarks

```
<adminInfo>
  <recordHist>
    <source>
      <p>Information transcribed from <ref target="IMEV123">IMEV 123</ref>
    </p>
    </source>
  </recordHist>
  <custodialHist>
    <custEvent type="conservation" notBefore="1961-03" notAfter="1963-02">
      <p>Conserved between March 1961 and February 1963 at Birgitte Dalls
        Konserveringsværksted. </p>
    </custEvent>
    <custEvent type="photography" notBefore="1988-05-01" notAfter="1988-05-30">
      <p>Photographed in May 1988 by AMI/FA. </p>
    </custEvent>
    <custEvent type="other" notBefore="1989-11-13" notAfter="1989-11-13">
      <p>Dispatched to Iceland 13 November 1989. </p>
    </custEvent>
  </custodialHist>
</adminInfo>
```



EU-funded eContent+ program to:

- Create seamless access to distributed information about manuscripts and rare old printed books in Europe
- Connect existing digital libraries, bring aboard those who do not have them
- Provide access to these in Manuscriptorium in their own languages and their own virtual interface
- To standardise metadata on a subset of TEI P5 XML



Workpackage Leaders

- National Library of the Czech Republic, Prague
- AiP Beroun, s r.o., Beroun, Czech Republic
- Oxford University Computing Services, Oxford, United Kingdom
- Centro per la comunicazione e l'integrazione dei media, Florence, Italy
- SYSTRAN S.A., Paris, France
- Institute of mathematics and informatics, Vilnius, Lithuania
- Biblioteca Nacional de España, Madrid, Spain

But *many many* more participating institutions and content partners!

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



Workpackages

- WP1: Project Management
- WP2: Preparation for system implementation and content enhancement
- **WP3: *Standardization of shared metadata***
- WP4: User personalization
- WP5: Personalization for contributors
- WP6: Multilingual and sophisticated access
- WP7: Evaluation, testing, and validation
- WP8: Dissemination and exploitation

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



ENRICH: Workpackage 3

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description

- Conversion between MASTER (and other formats) and TEI P5 / ENRICH ODD
- Implementation of the OAI harvester into existing Manuscriptorium platform
- Enhancement of internal use of METS containerization format to use benefits of TEI P5
- Improvement and generalization of use of Unicode treatment throughout interface



Work done so far...

- A comparison between MASTER, Manuscriptorium and TEI P5, with resolutions of inconsistencies for ENRICH
- An ENRICH ODD specification recording the customisation from the TEI, but also providing additional constraints, and detailed internationalised project documentation
- A draft XSLT stylesheet for transforming MASTER to TEI P5 (ENRICH)
- An initial test corpus of 1064 manuscript descriptions from various sources
- Highly detailed documentation and reference information for the ENRICH specification
- Extensive training materials and slides for those teaching local encoders how to use the specification

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description



Conclusions

- I've barely scratched the surface looking at TEI Markup that is beneficial to digital philology!
- Electronic scholarly editions **should**:
 - properly document their source material
 - responsibly indicate scribal and editorial interventions
 - where possible make the underlying encoded text available for scholarly re-use
 - collaborate with existing EU and international initiatives
 - use accepted international standards and recommendations to aid interoperability and preservation
- Using XML, and preferably TEI P5 XML, is a step in the right direction!

(And of course this presentation was written using TEI P5 XML, and will be freely available online.)

ENRICHing
Electronic
Manuscripts
with TEI P5
XML

Dr James
Cummings

Introduction

TEI:
Transcription

TEI: Manuscript
Description