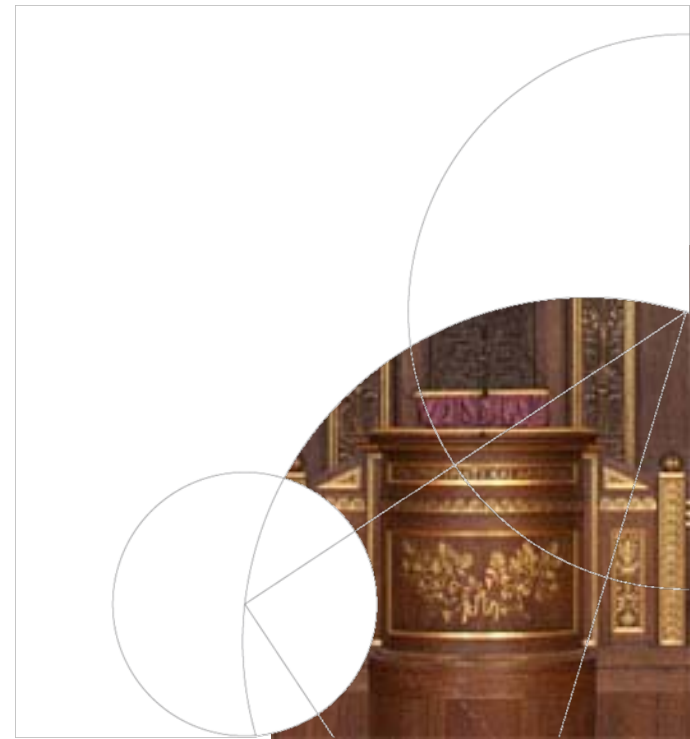


The Text Encoding Initiative and manuscript studies

M. J. Driscoll
Arnamagnæan Institute
Copenhagen

COMSt Conference
Hamburg, 1-3 December 2009



The Text Encoding Initiative



The **TEI** is an international and interdisciplinary standards project established in 1987 to develop, maintain and promulgate hardware- and software-independent methods for encoding humanities data in electronic form (initially SGML, now XML).

The current version of the TEI Guidelines, **TEI P5**, released in November 2007, contains much of potential interest to manuscript scholars.

It provides mechanisms for encoding in XML:

- A detailed description of the text-bearing object
 - A multi-layered transcription of the text(s)
 - Information about digital surrogates (images)
 - Information about real-world entities associated with the manuscript, i.e. the people, places and organisations involved in its production, dissemination and reception
- and linking all these together seamlessly.



Manuscript description

The **TEI manuscript description module** defines a special purpose element which can be used to provide detailed descriptive information about handwritten primary sources. Although originally developed to meet the needs of cataloguers and scholars working with medieval manuscripts in the European tradition, the scheme is general enough that it can also be applied to other traditions and materials, and is potentially useful for any kind of text-bearing object.

The TEI `<msDesc>` element is intended for several different kinds of applications:

- Electronic catalogue record or finding aid
- Metadata component within a digital edition
- Tool for quantitative codicology



Elements within <msDesc>

- <msIdentifier>**: groups information uniquely identifying the manuscript, such as holding institution and shelfmark.
- <msContents>**: provides an itemised list of the intellectual content of the manuscript, with transcriptions of rubrics, incipits, explicits etc., as well as primary bibliographic references.
- <physDesc>**: groups information concerning all physical aspects of the manuscript, its material, size, format, script, decoration, binding, marginalia etc.
- <history>**: provides information on the history of the manuscript, its origin, provenance and acquisition by its current holding institution
- <additional>**: groups other information about the manuscript, in particular administrative information relating to its availability, custodial history, surrogates etc.



The transcription of primary sources

The **TEI transcription module** provides mechanisms for encoding not only 'the words on the page', but also such features as:

- Page layout – as well as text structure
- Orthography, capitalisation, word division and punctuation – original as well as normalised
- Abbreviations – and their expansions
- Additions and deletions
- Scribal errors and omissions – and editorial emendations
- Variant letter forms



The Gaiji module

In most cases, **Unicode** already covers most of the characters most scholars need. There are, however, many characters and uncommon glyphs which have yet to make it into Unicode. Moreover, one may wish to record palaeographic variants of a single character in order to facilitate scribal identification or for statistical purposes.

Using **the 'Gaiji' module** one can encode characters or glyphs which do not appear in Unicode by defining them in a series of <charDesc> elements in the TEI header and then referring to them using using the <g> element in the body of the text.

Within <charDesc> one then uses either the <char> element to define a character which is not available in Unicode, or <glyph> to annotate an existing character.



Example: Defining a new glyph

A new glyph variant can be defined and assigned to a position in the Unicode Private Use Area (PUA), with a standardised form provided as a fallback:

```
<charDecl>
  <glyph xml:id="z103">
    <glyphName>LATIN LETTER Z WITH TWO STROKES</glyphName>
    <mapping type="standardised">z</mapping>
    <mapping type="PUA">U+E304</mapping>
  </glyph>
</charDecl>
```

This can now be referenced in the text using the `<g>` element:

```
<g ref="#z103"/>
```

It is also possible to provide a default value like this:

```
<g ref="#z103">z</g>
```



The facsimile module

The **<facsimile>** element (a sibling of **<text>**) contains a representation of the written source in the form of a set of images rather than as transcribed or encoded text.

The **<surface>** element defines a written surface in terms of a rectangular space.

The **<zone>** element defines a rectangular area contained within a **<surface>**.

The **@facs** attribute, available globally, points directly from any element to the **<facsimile>**, **<surface>** or **<zone>** to which it corresponds.

In this way one can point from anywhere in a manuscript description or text to the corresponding place on the manuscript page (and vice versa).



An example of <facsimile>

```
<facsimile xml:base="http://www.handrit.org/AM/fo1/">
  <surface start="#LSB-1r" ulx="0" uly="0" lrx="200" lry="300">
    <graphic mimeType="jpeg" xml:id="AM02-5000-1r"
      url="AM02-5000-1r.jpg"/>
    <graphic mimeType="jpeg" url="AM02-5000-1r-thumb.jpg"/>
    <zone ulx="20" uly="20" lrx="70" lry="70">
      <desc>Intitial (detail)</desc>
      <graphic mimeType="jpeg" xml:id="AM02-5000-1r-det"
        url="AM02-5000-1r-det.jpg"/>
    </zone>
  </surface>
  <surface ulx="0" uly="0" lrx="200" lry="300">
    <graphic mimeType="jpeg" xml:id="AM02-5000-1v"
      url="http://www.handrit.org/AM/fo1/AM02-5000-1v.jpg"/>
    <graphic mimeType="jpeg" url="http://www.handrit.org/AM02-5000-
      1v-thumb.jpg"/>
  </surface>
  <!-- -->
</facsimile>
```



Persons, places and organisations

New mechanisms available in P5 make it possible to encode data on people, places and organisations associated with a single manuscript, manuscript collection or corpus of texts; the use of such 'authority files' prevents repetition of information and minimises the risk of error.

It also allows us to map the relationships between the manuscripts and the people who produced and consumed them, to show in a dynamic way how the 'manuscript matrix' worked.



Biographical and prosopographical data

The **<person>** element can be used to encode data pertaining to people, whether actors in history or those living today.

Such data include:

- physical characteristics such as sex and eye colour
- cultural characteristics such as socio-economic status and religion
- information on occupation and education, and the events in people's lives such as birth, marriage or appointment to office
- the various names by which a person may be known



An example of <person>

```
<person sex="1" role="scribe" xml:id="BryJon001">
  <persName>
    <forename sort="1">Brynjólfur</forename>
    <surname type="patronymic" sort="2">Jónsson</surname>
  </persName>
  <birth notBefore="1600" notAfter="1649"/>
  <death when="1664"/>
  <residence>
    <settlement type="farm" key="EfsÖxn01">Efstaland</settlement>
  </residence>
  <occupation ref="#bondi"/>
  <socecStatus scheme="#socecStatus" code="#socec4"/>
  <education ref="#edu1">Learnt to write from his father</education>
</person>
```



Specifying relationships

The **<relation>** element can be used to describe any kind of relationship between people or between people and objects.

```
<relation name="employer" active="#ThoSku001"  
passive="#BryJon001"/>
```

```
<relation name="scribe" active="#BryJon001" passive="#AM04-  
0285 #AM08-0109a-II #Papp04-17"/>
```



The <place> element

Information about places can be encoded using the **<place>** element, which, like <person>, contains a number of sub-elements:

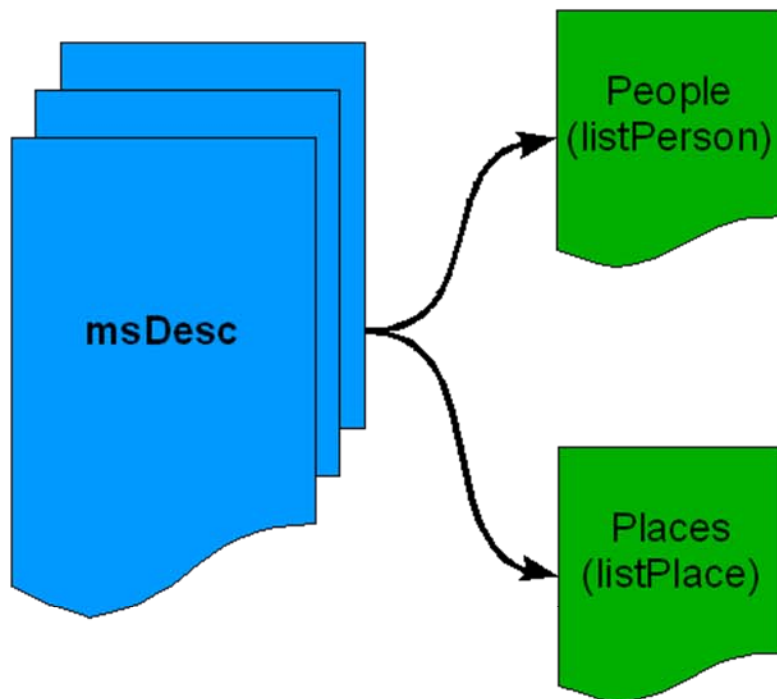
- one or more <placeName> elements
- one or more <location> elements, giving the physical location of the place, for example as a set of geographical co-ordinates
- one or more generic <trait> elements, or the specialised elements <population>, <climate> and <terrain>
- one or more generic <state> and <event> elements
- bibliographical references



An example of the <place> element

```
<place xml:id="VilVil01">  
  <placeName>  
    <settlement type="farm">Villingaholt</settlement>  
    <region type="parish" key="Villin01"/>  
  </placeName>  
  <location>  
    <geo>63.883997 -20.750909</geo>  
  </location>  
  <trait type="propSize" extent="16hh"/>  
  <trait type="propType" ref="#Ben"/>  
</place>
```





```
<person xml:id="ArnMag001">
  <persName xml:lang="is">
    <forename>Árni</forename>
    <surname>Magnússon</surname>
  </persName>
  ...
```

```
<place xml:id="Borga01">
  <placeName xml:lang="is">
    <region type="county">
      Borgarfjarðarsýsla
    </region>
    <region type="geog" key="#Sunnle01"/>
  ...
```

<p>According to <name type="person" key="ArnMag01">Árni Magnússon</name>, this manuscript received the name <title type="ms">Hulda</title> (i.e. <q>the hidden one</q>) in <name type="place" key="Borga01">Borgarfjörður</name>, where in the <date from="1660" to="1669">1660s</date> it was loaned secretly from farm to farm without the owner's permission or knowledge.</p>



AM 66 fol. (“Hulda”)

Hulda; Iceland, s. XIV 3/4

The sagas of the kings of Norway in the version known as *Hulda* go up to and include *Magnúss saga Erlingssonar*. They cover the period between 1035 and 1117, from the adolescence of [Magnus the Good](#) to the battle of Re. Because of a lost gathering at the beginning of the manuscript, the text begins in the sixteenth chapter of *Magnúss saga góða*. According to [Árni Magnússon](#), this manuscript received the name *Hulda* (i.e. “the hidden one”) in Borgarfjörður, where in the 1660s it was loaned secretly from farm to farm without the owner's permission or knowledge.

CONTENTS

([2r-141v](#)) *Heimskringla - Noregs konunga sögur*

Language of text: Old Norse/Icelandic

Begins: “eigi konung j fylkingu. hann leitar [...] einum.”

Explicit: “ok hann myndi vera því meiri hermaðr en jarl sem hann var yngri”

Final rubric: “ok lykr her at segja | fra þeim Magnusi konungi ok erlingi jarli”

[Fornmanna sögur](#), VI, 28:30-448, VII, 1-326 Rafn, [Antiquités Russes](#), II, 19-76 Konráð

Gíslason, [Fire og fyrretyve Prøver](#), 489-491 (chapter XXXI) and in part, 492-499

(chapter XXXII) Storm, [Sigurd Ranessøns Proses](#), 1-29 Louis-Jensen, [Hulda](#)

(forthcoming)

Document

Contents

Images

Metadata

Country Denmark

Settlement Copenhagen

Repository Den Arnamagnæanske Samling

Shelfmark AM 66 fol.

Alt. ID 77

Type: internal

Subtype: KKKat

XML

Source [AM02-0066-en.xml](#)

Source [AM02-0066-da.xml](#)

handrit.is

[Home](#)[Browse](#)[Search](#)

English

JÓN ERLENDSSON

DETAILS

Name	Jón Erlendsson
Birth	ca. 1605
Death	August 1672
Occupation	Priest
Role	Scribe
Residence	Villingaholt , Iceland [map it] 

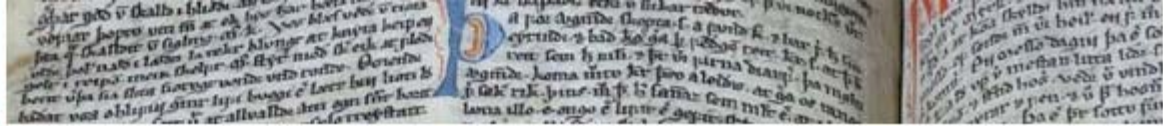
BIBLIOGRAPHY

Íslenzkar æviskrár III, pp. 105.106

ASSOCIATED MANUSCRIPTS

Shelfmark	Title	Origin	Role
AM 1 a fol.	Sögubrot af nokkrum fornkonungum í Dana ok Svía veldi	Iceland, ca. 1650	Scribe
AM 6 fol.	Völsunga saga and Ragnars saga loðbrókar with Krákumál	Iceland, s. XVII	Scribe
AM 43 fol.	Noregs konunga sögur	Iceland, s. XVII	Scribe
AM 49 fol.	On Norwegian Kings	Iceland, s. XVII	Scribe





GEOGRAPHIC REFERENCE

Results

Villingaholt

Map **Satellite** **Hybrid**

Villingaholt
 Villingaholtshreppur
 Árnessýsla
 Vilvil01
 63.883997
 -20.750909

Map

20 km
20 mi

POWERED BY Google

Terms of Use

For more information...

www.tei-c.org

mjd@hum.ku.dk

