



XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

XML Technologien und ihre Bedeutung für LZA und I



XML Technologien und ihre Bedeutung für LZA und I

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Sebastian Rahtz

Information Manager, Oxford University Computing Services

Text Encoding Initiative Board of Directors

Text Encoding Initiative Technical Council

Berlin, October 2007



Outline

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- 1 Markup and digital text
- 2 Validation and schemas
- 3 Using XML: transformation
- 4 Standardized vocabularies
- 5 Conclusions



On the table

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- Markup and digital text
- XML basics
- Schemas
- XSLT
- Vocabularies
- The Text Encoding Initiative
- Interchange



Markup and digital text

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions



A real world object

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions





What shall we record?

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- Where is it?
- What shape is it? what are its decorative features?
- What do we know about Molly Cotton?
- Text of the inscription
- When did we make this record?

Some of this is standard metadata, but we probably want to go deeper.



Before XML

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- In the beginning there was *procedural* markup RED INK ON;
print balance; RED INK OFF
- which being generalised became *descriptive* markup
<balance type='overdrawn'>some
numbers</balance>
- also known as **encoding** or **annotation**

Descriptive markup allows for re-use of data



In more detail

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Markup:

- makes explicit the distinctions we want to make when processing a string of bytes
- works better when we record what we think things *are* than what they *look like*
- adds value by supplying multiple annotations
- facilitates re-use of the same material
 - in different formats
 - in different contexts
 - by different users

It works for text **and** data



What is in a text?

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Upon Julia's Clothes

WHEN as in silks my *Julia* goes,
Then, then (me thinks) how sweetly flowes
That liquefaction of her clothes.

Next, when I cast mine eyes and see
That brave Vibration each way free;
O how that glittering taketh me!



What can markup capture?

Compare

```
<head>Upon Julia's Clothes</head>
<lg>
  <l>Whenas in silks my <hi>Julia</hi> goes,</l>
  <l>Then, then (me thinks) how sweetly flowes</l>
  <l>That liquefaction of her clothes.</l>
</lg>
```

and

```
<s n="1" role="head">
  <w type="pp">Upon</w>
  <w type="np">Julia</w>
  <w type="pos">'s </w>
  <w type="nn2">Clothes</w>
</s>
<s n="2" role="line">
  <w type="adv">Whenas</w>
  <w type="pp">in</w>
  <w type="nn2">silks</w> ...
</s>
```



A manuscript

XML

Technologien
und ihre
Bedeutung für
LZA und I

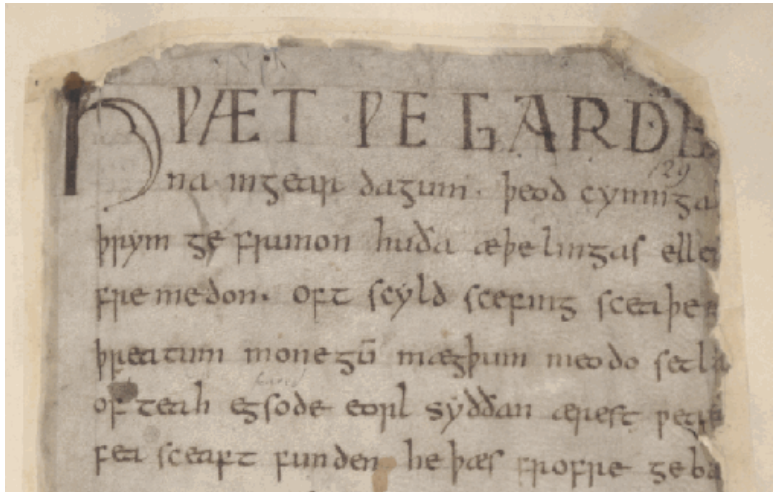
Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions





Choose again what to record. This:

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<hi rend="dropcap">H</hi>&amp;WYN;ÆT WE GARDE <b>na in  
gear-dagum þeod-cyninga <b>þrym gefrunon,  
hu ða æþelingas <b>ellen fremedon. oft scyld scefing sceaþe  
<add>na</add>  
<b>þreatum, moneg<expan>um</expan> mægþum meodo-setl  
<add>a</add>  
<b>of<damage desc="blot"/>teah egsode <sic>eorl</sic> syððan  
ærest wear  
<add>þ</add>  
<b>fea sceaft funden...
```



Or this:

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<lg>
  <l>Hwæt! we Gar-dena in gear-dagum</l>
  <l>beod-cyninga þrym gefrunon,</l>
  <l>hu ða æþelingas ellen fremedon,</l>
</lg>
<lg>
  <l>Oft Scyld Scefing sceaþena þreatum,</l>
  <l>monegum mægþum meodo-setla ofteah;</l>
  <l>egsode Eorle, syððan ærest wearþ</l>
  <l>feasceft funden...</l>
</lg>
```



From text to data

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<biblStruct xml:id="Simons1999">
  <analytic>
    <author>
      <forename>Gary</forename>
      <forename>F.</forename>
      <surname>Simons</surname>
    </author>
    <title level="a">Using Architectural Forms to Map
      TEI Data into an
      Object-Oriented Database</title>
  </analytic>
  <monogr>
    <title level="j">Computers and the Humanities</title>
    <imprint>
      <biblScope type="vol">33</biblScope>
      <biblScope type="issue">1-2</biblScope>
      <date>1999</date>
      <biblScope type="pp">85-101</biblScope>
    </imprint>
  </monogr>
</biblStruct>
```



Or mimicking a computer program

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<match pattern="**/*.htm">
  <select type="resource-exists">
    <when test="1/2.xml">
      <redirect-to uri="cocoon:/1/2.xml"/>
    </when>
    <otherwise>
      <read src="1/2.htm" mime-type="text/html"/>
    </otherwise>
  </select>
</match>
```




And back to text again

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<body>
  <div xml:lang="en">
    <ab>DR MOLLY COTTON</ab>
    <ab>O.B.E. F.S.A.</ab>
    <ab>MARY AYLWIN COTTON</ab>
    <ab>NÉE MARSHALL</ab>
    <ab>*1. 8. 1902 +31. 5. 1984</ab>
  </div>
</body>
```



but also data

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<listPerson>
  <person sex="2">
    <persName>
      <forename>Mary</forename>
      <forename>Aylwin</forename>
      <surname>Cotton</surname>
    </persName>
    <birth when="1902-08-01"/>
    <death when="1984-05-31"/>
    <nationality reg="UK"/>
  </person>
</listPerson>
```



Validation and schemas

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

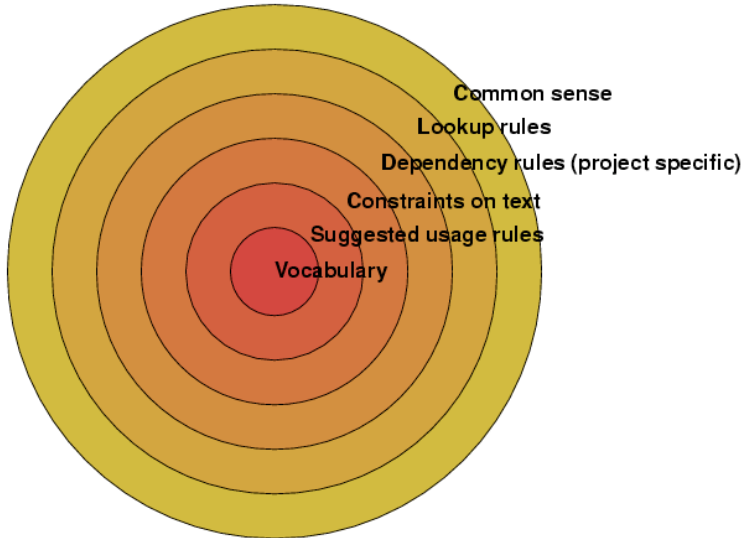
XML has a fixed syntax, and you define the vocabulary rules.
Syntax and vocabulary checking schemas can be written in:

- The W3C schema language
- RELAX NG schema language
- XML DTD language

But there is more to life than checking syntax checking.



What kinds of validation do we need?



XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions



Simple schemas

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

DTD, XSD and RELAX NG schemas specify:

- the name of the root element
- names for all elements used
- names and datatypes and (occasionally) default values for their attributes
- rules about how elements can nest
- and a few other things, depending on the schema language

A schema does *not* specify anything about what elements ‘mean’



ISO Schematron can add data-specific rules

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Before:

```
<element name="identifier">  
  <text/>  
</element>
```

After:

```
<element name="identifier">  
  <text/>  
</element>  
<pattern name="identifier-content">  
  <rule context="x:identifier">  
    <report test="string-length(normalize-space(.))=0">  
Identifier cannot be empty  
    </report>  
  </rule>  
</pattern>
```



Worthy but dull

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<define name="xcri-description">
  <element name="description">
    <ref name="xcri-model-description"/>
    <attribute name="type">
      <text/>
    </attribute>
  </element>
</define>
<pattern name="description-type">
  <rule context="x:presentation/x:description">
    <report
      test="@type and not(@type='Study Hours'
or @type='Contact Pattern' or @type='Contact Hours')"> The
type attribute on the description of a presentation is a fixed
vocabulary;
      this does not match. </report>
    </rule>
  </pattern>
```



Security checks

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<define name="xcri-model-description">
  <!-- ... -->
  <zeroOrMore>
    <choice>
      <ref name="htmlDiv"/>
      <text/>
    </choice>
  </zeroOrMore>
  <pattern name="description-safe">
    <rule context="x:description">
      <report test="!//xhtml:script"> Do NOT allow HTML
        script elements in your description </report>
    </rule>
  </pattern>
</define>
```




Tricksier

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<define name="xcri-postcode">
  <element name="postcode">
    <text/>
    <ns prefix="x" uri="http://xcri.org/profiles/catalog"/>
    <pattern name="postcode-type">
      <rule context="x:postcode">
        <report test="string-length(normalize-
space())=0">Postcode
          element cannot be empty</report>
        <assert
          test="document('postcodes.xml')/postcodes/postcode/@ident-
before(.,' ')"> First half of postcode must match known list
        </assert>
      </rule>
    </pattern>
  </element>
</define>
```



Mixed vocabulary

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<course xmlns="http://xcri.org/profiles/catalog">
  <identifier
xmlns="http://xcri.org/profiles/catalog">007P158AHJ</identifier>
  <title xmlns="http://xcri.org/profiles/catalog">After the Ice
Ages - New Research on Mesolithic Britain</title>
  <description
xmlns="http://xcri.org/profiles/catalog" type="Structure">
    <xhtml:div>
      <xhtml:br/>9.45am Mesolithic Britain: setting
      the scene<xhtml:br/>PROFESSOR NICK BARTON<xhtml:br/>
      ...</xhtml:div>
    </description>
  <subject
xmlns="http://purl.org/dc/elements/1.1/">Archaeology</subject>
</course>
```



ISO NVDL can specify validation of mixed namespaces in stages

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<namespace ns="http://xcri.org/profiles/catalog">
  <validate schema="oxcap.rng">
    <mode>
      <namespace ns="http://purl.org/dc/elements/1.1/">
        <validate schema="dc.rng"/>
      </namespace>
      <namespace ns="http://www.w3.org/1999/xhtml">
        <validate schema="xhtml.rng">
          <mode>
            <anyNamespace>
              <allow/>
            </anyNamespace>
          </mode>
        </validate>
      </namespace>
    </mode>
  </validate>
</namespace>
```



Using XML: transformation

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

The XSLT language is

- Expressed in XML; uses namespaces to distinguish output from instructions
- Purely functional
- Reads and writes XML trees
- Designed to generate XSL FO, but now widely used to generate HTML



How is XSLT used? (1)

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- With a command-line program to transform XML (eg to HTML)
 - Downside: no dynamic content, user sees HTML
 - Upside: no server overhead, understood by all clients
- In a web server *servlet*, eg serving up HTML from XML (eg Cocoon, Axkit)
 - Downside: user sees HTML, server overhead
 - Upside: understood by all clients, allows for dynamic changes



How is XSLT used? (2)

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- In a web browser, displaying HTML from XML on the fly
 - Downside: many clients do not understand it
 - Upside: user gets XML displayed as HTML
- Embedded in specialized program
- As part of a chain of production processes, performing arbitrary transformations



XSLT implementations

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

MSXML Built into Microsoft Internet Explorer

Saxon Java-based, standards leader, implements XSLT 2.0
(basic version free)

Xerces Java-based, widely used in servlets (open source)

libxslt C-based, fast and efficient (open source); used from
Perl, PHP, C etc

transformiix C-based, used in Mozilla (open source)



What do you mean, 'transformation'?

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Take this

```
<recipe>
  <title>Pasta for beginners</title>
  <ingredients>
    <item>Pasta</item>
    <item>Grated cheese</item>
  </ingredients>
  <cook>Cook the pasta and mix with the cheese</cook>
</recipe>
```




and make this

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<html>  
  <h1>Pasta for beginners</h1>  
  <p>Ingredients: Pasta Grated cheese</p>  
  <p>Cook the pasta and mix with the cheese</p>  
</html>
```



How do you express that in XSL?

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:template match="recipe">
    <html>
      <h1>
        <xsl:value-of select="title"/>
      </h1>
      <p>Ingredients:
      <xsl:apply-templates select="ingredients/item"/>
      </p>
      <p>
        <xsl:value-of select="cook"/>
      </p>
    </html>
  </xsl:template>
</xsl:stylesheet>
```



Standardized vocabularies

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- XHTML: web pages
- MathML: maths
- SVG: vector graphics
- RSS/Atom: news feeds
- Docbook/DITA: technical writing
- TEI: digital texts
- MODS/METS: library finding aids

The list goes on and on for each specialism ...



The Text Encoding Initiative

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- defines and names several hundred useful textual distinctions
- provides a set of modules that can be used to define schemas making those distinctions
- provides a customization mechanism for modifying and combining those definitions with new ones using the same conceptual model

The TEI provides a framework for the definition of multiple schemas



TEI Deliverables

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- A set of recommendations for text encoding, covering both generic text structures and some highly specific areas based on (but not limited by) existing practice
- A very large collection (about 530) of element **definitions** with associated **declarations** for various schema languages
- a modular system for creating personalized schemas or DTDs from the foregoing



Legacy of the TEI

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- a way of looking at what ‘text’ really is
- a codification of current scholarly practice
- (crucially) a set of shared assumptions and priorities about the digital agenda:
 - focus on content and function (rather than presentation)
 - identify generic solutions (rather than application-specific ones)



Some things the TEI is famous for

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- Prose, verse and drama
- Editorial markup
- Linguistic markup
- Complex linking
- Describing manuscripts



Some less obvious areas the TEI is in

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- Names, places, persons, organisations and dates
- Digital facsimiles
- Non-hierarchical markup
- Dictionaries
- Figures and tables
- Writing new documents (and these slides ...)
- Describing XML schemas



The Text Encoding Initiative meets historical data

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Publication

Data

Header
Structure

Figures/Tables

Core
Linking

Names/Dates

Feature Structures

Verse
Drama
Analysis
Dictionaries
Transcription
Gaiji
Manuscripts

Historical Documents



Example: some report writing

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<div>
  <head>The Fabrics</head>
  <p> The fabrics identified are described below. Proportions
    of each fabric are listed as percentage figures. Following
    each description is a generalized list of vessels commonly
    found in that fabric at Middleton Stoney.
    <list type="gloss">
      <label>Ra</label>
      <item> Reduced ware; coarse; major inclusions: grog. The
type
                is a 'catch-all' for grog-tempered fabrics, and
different sherds
                may incorporate inclusions of quartz, quartzite, mica,
chalk,
                and red or black iron ore in varying proportions. The
colour
                ranges from grey or black through to pale pinkish or
yellowish
                brown. </item>
      </list>
    </p>
</div>
```



Example: a more complex record from classical archaeology

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<person xml:id="V2-45386">
  <sex value="1"/>
  <persName type="full" xml:lang="el-grc">Θεμιστοκλῆς</persName>
  <persName xml:lang="el-grc-x-lgpn">Qemistoklh3s</persName>
  <persName xml:lang="el-grc-x-perseus">*qemistoklh=s</persName>
  <birth notAfter="-0075" notBefore="-0125">
    <placeName key="Hagnous">Hagnous</placeName>
  </birth>
  <floruit>ii/i BC</floruit>
  <state key="#relationship">
    <label>s. <persName xml:lang="el-grc">Θεόφραστος</persName>
I</persName>
    </label>
  </state>
  <state key="#relationship">
    <label>f. <persName xml:lang="el-grc">Θεόφραστος</persName>
II</persName>
    </label>
  </state>
  <state key="#relationship">
    <label>f. <persName xml:lang="el-grc">Ξενοκλῆς</persName>
II</persName>
```



Hang on (1), why don't you use an SQL database, mapping software, GIS etc?

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- 1 All the data is in a single archival format independent of proprietary software
- 2 All the complexity can be modelled in one place
- 3 If you do have deep recursive text structures, modelling them in an RDB is hard
- 4 Making everything an XML transformation is very powerful



Hang on (2): why aren't you using an ontology in RDF?

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- Luckily, the TEI semantics map neatly and reliably to CIDOC CRM concepts, which means that RDF representation in CIDOC terms is just a transformation away.
- The TEI is a way of instantiating your material for CIDOC CRM material
- The TEI scheme has explicit recording of CIDOC mapping.



A TEI XML geo record

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<place type="college" xml:id="mert">
  <label>Merton College</label>
  <location type="address">
    <address>
      <addrLine>Oxford</addrLine>
      <postCode>OX1 4JD</postCode>
    </address>
  </location>
  <place type="building">
    <label>Lodge</label>
    <location when="2007-01-20T21:26:32.601Z">
      <geo>-1.252216100692749 51.75129113668488</geo>
      <note>recorded by Sebastian Rahtz</note>
    </location>
  </place>
</place>
```



Displaying TEI data

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions



Web Images News Maps **News!** Products Groups Scholar more »
<http://www.oucs.ox.ac.uk/cgi-bin/oxpoints.cgi?type=college>
Search the map Find businesses

Search Maps

Search Results My Maps **News!**

[Save to My Maps](#)

Oxford University

- ☒ All Souls College
- ☒ All Souls College: The Lodge
- ☒ Balliol College
- ☒ Balliol College: Deila Building
- ☒ Balliol College: Holywell Manor
- ☒ Balliol College: Jowett Walk Buildings
- ☒ Balliol College: Lodge
- ☒ Balliol College: Martin Building
- ☒ Balliol College: Sports Ground
- ☒ Blackfriars
- ☒ Blackfriars: Lodge
- ☒ Brasenose College
- ☒ Brasenose College: Lodge
- ☒ Clarendon Hall
- ☒ Clarendon Hall: Lodge
- ☒ Christ Church
- ☒ Christ Church: Library
- ☒ Christ Church: Lodge
- ☒ Corpus Christi College
- ☒ Corpus Christi College: Lodge



sebastian.rahtz@gmail.com | [Saved Locations](#) | [Help](#) | [My Account](#) | [Sign out](#)

[KML](#) [Print](#) [Email](#) [Link to this page](#)

[Map](#) [Satellite](#) [Hybrid](#)



The map replacement — more information

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Google Maps UK interface showing a map of Oxford University and surrounding areas. The search bar contains the URL: <http://www.oucs.ox.ac.uk/cgi-bin/oxpoints.cgi?type=college>. The search results list various colleges and buildings, including All Souls College, Balliol College, Brasenose College, and Corpus Christi College. A callout box highlights St Edmund Hall: St Edmund Hall, with the URL <http://www.seh.ox.ac.uk/>.

Search Results: My Maps [Save to My Maps](#)

Oxford University

- ☒ All Souls College
 - ☒ All Souls College: The Lodge
- ☒ Balliol College
 - ☒ Balliol College: Detail Building
 - ☒ Balliol College: Holywell Manor
 - ☒ Balliol College: Jowett Walk Buildings
 - ☒ Balliol College: Lodge
 - ☒ Balliol College: Martin Building
 - ☒ Balliol College: Sports Ground
- ☒ Blackfriars
 - ☒ Blackfriars: Lodge
- ☒ Brasenose College
 - ☒ Brasenose College: Lodge
- ☒ Campion Hall
 - ☒ Campion Hall: Lodge
- ☒ Christ Church
 - ☒ Christ Church: Library
 - ☒ Christ Church: Lodge
- ☒ Corpus Christi College
 - ☒ Corpus Christi College: Lodge

St Edmund Hall: St Edmund Hall
<http://www.seh.ox.ac.uk/>
[Search nearby](#) - [Save to My Maps](#)

©2007 Google - Mapdata ©2007 Tele Atlas - Terms of Use



Replacing the text browser?

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

sebastian.rahtz@gmail.com | Saved Locations | Help | My Account | Sign out

Google Maps UK

Web Images News Maps^{News} Products Groups Scholar more...

http://www.oucs.ox.ac.uk/cgi-bin/oxpoints.cgi?type=human

Search the map Find businesses

Search Results My Maps^{News}

Save to My Maps

Oxford University

- ☒ Faculty of Classics
 - ☒ Faculty of Classics: Classics Centre, The Old Boys' School
 - ☒ Faculty of Classics: Ioannou Centre for Classical and Byzantine Studies
- ☒ Faculty of English Language and Literature
 - ☒ Faculty of English Language and Literature: St Cross Building
- ☒ Faculty of History
 - ☒ Faculty of History: The Old Boys' School
- ☒ Faculty of Medieval and Modern Languages
- ☒ Faculty of Music
 - ☒ Faculty of Music: Faculty of Music
- ☒ Faculty of Philosophy
 - ☒ Faculty of Philosophy: Lecture Room
 - ☒ Faculty of Philosophy: Littlegate House
- ☒ Faculty of Philosophy: Philosophy Centre, 10 Merton St
- ☒ Faculty of Philosophy: Ryle Room

☒ Oriental Institute

- ☒ Oriental Institute: Oriental Institute

- ☒ Phonetics Laboratory
- ☒ Phonetics Laboratory: 41 Wellington Square
- ☒ Theology Faculty Library
- ☒ Theology Faculty Library: Theology Faculty Library

Map Satellite Hybrid

KML Print Email Link to this page

Faculty of English Language and Literature: St Cross Building
<http://www.english.ox.ac.uk/> Webteam:
<http://webteam.ox.ac.uk/site/human/english/>
Search nearby - Save to My Maps

©2007 Google - Map data ©2007 Tele Atlas - Terms

http://webteam.ox.ac.uk/site/human/english/



Mapping the world

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

<http://www.oucs.ox.ac.uk/oxpoints/test.html>



Representation of a schema

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

```
<elementSpec module="spoken" ident="pause">
  <classes>
    <memberOf key="model.divPart.spoken"/>
    <memberOf key="att.timed"/>
    <memberOf key="att.typed"/>
  </classes>
  <content>
    <rng:empty/>
  </content>
  <attList>
    <attDef ident="who" usage="opt">
      <gloss>A unique identifier</gloss>
      <desc>supplies the identifier of the
        person or group pausing.
      </desc>
      <datatype>
        <rng:ref name="data.pointer"/>
      </datatype>
    </attDef>
  </attList>
  <desc>a pause either between or within utterances.</desc>
</elementSpec>
```



Roma: building ad-hoc TEI schemas

TEI Roma: generating validators for the TEI Modules

Navigation tabs: [New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#)

List of TEI Modules			
	Module name	A short description	Changes
add	analysis	Simple analytic mechanisms	
add	certainty	Certainty and uncertainty	
add	core	Elements common to all TEI documents	
add	corpus	Header extensions for corpus texts	
add	declarefs	Feature system declarations	
add	dictionaries	Printed dictionaries	
add	drama	Performance texts	
add	figures	Tables, formulae, and figures	
add	gaji	Character and glyph documentation	
add	header	The TEI Header	
add	iso-fs	Feature structures	
add	linking	Linking, segmentation and alignment	
add	msdescription	Manuscript Description	
add	namesdates	Names and dates	
add	nets	Graphs, networks and trees	

List of selected Modules

[remove](#) [core](#)

[remove](#) [tei](#)

[remove](#) [header](#)

[remove](#) [textstructure](#)

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions



Conclusions

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

What XML buys you:

- No need to invent a markup language for textual data
- Plenty of vocabularies to choose from
- Standardized processing



Great XML success stories

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- XSLT: simple language for transforming XML
- XPath: powerful language for addressing XML trees
- XHTML: taming HTML
- XML DOM: standardized API
- RSS: news feeds
- KML: overlays on maps
- AJAX: dynamic web pages reading XML
- TEI: of course...



What XML gets right

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- Mandatory Unicode
- Stream-based extensible markup
- Relative ease of writing a parser
- Namespaces
- Transformations from tree to tree
- Community acceptance of a simple standard



What XML gets wrong

XML
Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

- Legacy of SGML makes for a few awkwardnesses (eg entities)
- Namespaces were a slight afterthought
- Not amenable to old-style line-based tools
- Only models hierarchies



Final lesson

XML

Technologien
und ihre
Bedeutung für
LZA und I

Markup and
digital text

Validation and
schemas

Using XML:
transformation

Standardized
vocabularies

Conclusions

Whatever else you do, make your data available to other machines.

XML is at its best for inter-process communication