



Einführung in XML

Schachteln und Klammern

Dieses Element kann nicht ohne Attribut
daherkommen.



Überblick

- Was ist XML?
- <oXygen/>
- Grundregeln
- Wege nach XML
- XML-Bereiche für digitale Edition
- (Warum überhaupt?)



XML?

- eXtensible Markup Language
- W3C-Standard (aktuell 5. Ausgabe XML 1.0 vom 26.11.2008)
- Auszeichnungssprache zur strukturierten Beschreibung von Inhalten
- Menschen- und maschinenlesbar



Was ist XML?

- ... eine Dokumenten- oder **Inhaltsbeschreibungssprache**
- ... **trennt Struktur und Darstellung** voneinander, Inhalt und Präsentation.
- ... XML-Dokumente werden nach einem **Dokumentenmodell** entwickelt.
 - Dokumentenmodell als XML-Schema



Was ist XML?

Suche nach ***.xml**

- XML ist ein weit verbreiteter Standard.
- XML wird von einer breiten Softwarepalette unterstützt.
- XML hat eine große internationale Nutzer- und Entwicklergemeinschaft
- XML ist die Grundlage vieler Anwendungsstandards („XML is everywhere“).
- XML ist in zunehmendem Maße die Grundlage von Online-Ressourcen.
- XML umfasst eine ganze Familie von begleitenden Standards.



Wie komm' *ich* zu XML?

- Germanist
- Verlagswesen, Korrektorat, Grafikdesign
- Bibliothekswesen: Digitalisierung UB Wien



Wie komm' *ich* zu XML?

- Jobfiles Zeutschel-Scanner für EOD an der UB Wien
- ABBYY FineReader gibt XML-Jobfiles aus zur Weiterverarbeitung
- SVG-Vektorgrafiken
- Geodaten (GML, KML etc.)
- Text: TEI u.a.



Einstellungen für Software

...

```
<FullscanFirst>true</FullscanFirst>
```

```
<SkipCheck>true</SkipCheck>
```

```
<MagnifierSize>200</MagnifierSize>
```

```
<ProcessLogging>true</ProcessLogging>
```

```
<MetaDataOnFragments>true</MetaDataOnFragments>
```

```
<ShowScannerResolution>true</ShowScannerResolution>
```

...



Scalable Vector Graphics

<http://de.wikipedia.org/wiki/SVG#Beispiel>



XML ist ...

- ... einfach.
- ... wie HTML, nur anders:
 - xHTML ist eine Anwendung von XML.
 - XML ist erweiterbar.
 - XML muss wohlgeformt sein.
 - XML kann validiert werden.
 - besonders gut geeignet, um Texte zu strukturieren.



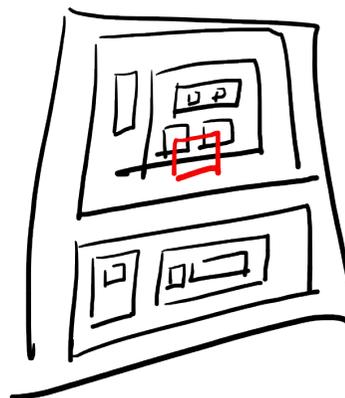
XML Dokumente sind ...

- ... langzeitarchivierbar.
- ... unabhängig von Plattform und Programmen austauschbar.
- Zur Dokumenterstellung reicht theoretisch ein simpler Texteditor aus.
- Moderne Browser können XML-Dokumente anzeigen und gegebenenfalls transformieren (IE).



Nachteile von XML

- sehr „geschwätzig“, d.h. speicherintensiv.
- Overlapping-Problem.



- relativ langsamer Zugriff auf Daten, verglichen mit relationalen Datenbanken.



Alles über ein XML-Dokument

- Ein XML-Dokument kann enthalten:
 - Elemente, eventuell mit Attributen
 - Verarbeitungsanweisungen
 - Kommentare
 - Entitätsreferenzen
- Ein XML-Dokument muss *wohlgeformt* sein und kann *validiert* werden.
- XML Attributwerte müssen in " " stehen.
- XML-Dokumente sind als lineare Zeichenketten kodiert.
- XML-Dokumente beginnen mit einer besonderen Verarbeitungsanweisung.



Mini XML-Dokument

```
<?xml version="1.0" ?>
<root>
  <element attribute="value">
    content
  </element>
<!-- comment -->
</root>
```



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <doc>
3   <rule number="1">Everything is delimited.
4     <elucidation>
5       <subrule>Elements are delimited by <term>start-</term>
6         and <term>end-tags</term>.</subrule>
7       <subrule>Tags are delimited by angle brackets.</subrule>
8       <subrule>Attribute values are delimited by
9         quotation marks.</subrule>
10    </elucidation>
11  </rule>
12  <rule number="2">Everything nests.
13    <elucidation>
14      <subrule>Elements occur within elements.</subrule>
15      <subrule>Attributes are specified within start-tags.</subrule>
16    </elucidation>
17  </rule>
18 </doc>
```



XML Regeln?

- Ihre Expertise ist gefragt!



XML-Dokument Regeln

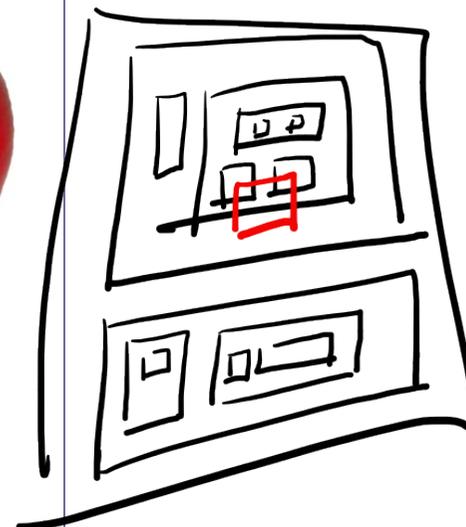
- Ein Element hat immer einen Start-Tag und einen End-Tag.
 - Das ist ein **<StartTag>**.
 - Das Ende markiert **</StartTag>**.
- Die Tagnamen sind case-sensitive.
<tag></tag> und nicht **<tag></Tag>**.
- Leere Elemente – so genannte milestone-Elemente – können abgekürzt notiert werden:
<LeererTag />





XML-Dokument Regeln

- Alle Elemente müssen richtig geschachtelt sein.





Zwischenfrage

- Wer arbeitet mit welcher Textverarbeitung?
 - MS Word
 - Writer (OpenOffice und Derivate)
 - Classical Text Editor
 - Plaintextformate (TUSTEP, LaTeX, XML-Anwendungen, ...)
 - Lotus
 - Word Perfect
 - Andere
- (<http://de.wikipedia.org/wiki/Kategorie:Textverarbeitung>)



Mein erstes XML-Abenteuer

- Layout „[Buchwesen in Wien 1750–1850](#)“
- Aus Access-Datenbank exportiertes .doc
- Rückverwandlung in strukturierten Text (XML)
- DTD zur Kontrolle
- .doc → (halbautomatisch regular expressions) → XML
 - 1) XML → XSL-T → LaTeX → PDF für Print und elektronische Ausgabe
 - 2) XML → halbautomatisch zurück in die Access-Datenbank



Überblick

- Was ist XML?
- `<oxygen/>`
- Grundregeln
- Wege nach XML
- XML-Bereiche für digitale Edition
- (Warum überhaupt?)



XML-Editoren

- Wesentliche Eigenschaften von XML-Editoren
 - Syntaxunterstützung
 - Validierung
 - Schema-Assoziierung
 - anpassbar
 - verschiedene Anzeigemodi
 - Folding-Unterstützung
 - Debugging-Werkzeuge
 - Kontextsensitive Syntaxunterstützung



XML-Editoren

- große Auswahl:
 - XML Spy
 - Stylus Studio
 - XMetaL
 - Syntext Serna
 - Emacs
 - jEdit
 - OpenOffice Writer
 - ...



<oxygen/> XML Editor

- Favorit im geisteswissenschaftlichen Bereich
- Entwickelt von SyncRO Soft Ltd. (Rumänien)
- Marktführer bei XML-Editoren
- Klare Lizenzpolitik, günstige Lizenzen für den akademischen Einsatz (USD 64 + MwSt.)
- www.oxygenxml.com



<oXygen/> XML Editor

- plattformunabhängig
- als Eclipse Plug-in erhältlich
- Subversion-Client for kollaboratives/versioniertes Arbeiten
- TEI-Unterstützung
- Unterstützung aller Schemasprachen
- Syntaxvervollständigung
- Tool-tip-Dokumentation
- XSL-T- und FOP-Unterstützung für Transformationen nach HTML, PDF, ...

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oxygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project exercise3-complete.xml

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>Unpublished</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <div type="verse">
19        <head>PROGRESS.</head>
20        <p>[Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry. --Evening News.] </p>
23        <lg>
24          <l>The little crafts! How soon they die! </l>
25          <l>In cottage doors no shuttle clicks; </l>
26          <l>The hand-loom has been ousted by </l>
27          <l>A large concern with lots more sticks.</l>
28        </lg>
29        <lg>
30          <l>The throb of pistons beats around; </l>
31          <l>Great chimneys rise on Thames's banks; </l>
32          <l>The same phenomena are found </l>
33          <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34        </lg>
35        <lg>
36          <l>No longer now the housewife makes </l>
37          <l>Her rare preserves, for what's the good? </l>

```

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r...
 - lg The little crafts! H...
 - lg The throb of pistor...
 - lg No longer now the...
 - lg 'Tis so with what w...
 - lg 'Tis this that chief...
 - lg That not with tend...

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml U+003C 1:1

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oxygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project exercise3-complete.xml

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <title>Demonstration exercise for Punch</title>
6     </fileDesc>
7     <publication>
8       <p>Unpublished</p>
9     </publication>
10    <sourceDesc>
11      <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
12    </sourceDesc>
13  </teiHeader>
14  <text>
15    <body>
16      <div type="verse">
17        <head> PROGRESS.</head>
18        <p>["Giving evidence recently before a Select Committee of the House of Commons,
19          Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
20          the small hand laundry."--Evening News.]</p>
21        <lg>
22          <l>The little crafts! How soon they die!</l>
23          <l>In cottage doors no shuttle clicks;</l>
24          <l>The hand-loom has been ousted by</l>
25          <l>A large concern with lots more sticks.</l>
26        </lg>
27        <lg>
28          <l>The throb of pistons beats around;</l>
29          <l>Great chimneys rise on Thames's banks;</l>
30          <l>The same phenomena are found</l>
31          <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
32        </lg>
33        <lg>
34          <l>No longer now the housewife makes</l>
35          <l>Her rare preserves, for what's the good?</l>
36        </lg>
37      </div>
38    </body>
39  </text>
40 </TEI>

```

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r
 - lg The little crafts! Hk
 - lg The throb of pisto
 - lg No longer now the
 - lg 'Tis so with what w
 - lg 'Tis this that chiefy
 - lg That not with tend

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml U+003C 1:1

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oxygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project

sample.xpr
css
debugger
fo
import
jsp
nvd1
relaxng
schematron
svg
wsdl
xquery
dita
docbook
ooxml
tei
xhtml
personal.schema.css

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r..."]
 - lg The little crafts! H...
 - lg The throb of pisto...
 - lg No longer now the...
 - lg 'Tis so with what w...
 - lg 'Tis this that chief...
 - lg That not with tend...

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStm>
6         <title>Demonstration exercise for Punch</title>
7       </titleStm>
8       <publicationStm>
9         <p>Unpublished</p>
10      </publicationStm>
11     <sourceDesc>
12       <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13     </sourceDesc>
14   </fileDesc>
15 </teiHeader>
16 <text>
17   <body>
18     <div type="verse">
19       <head> PROGRESS.</head>
20       <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23     </div>
24     <lg>
25       <l>The little crafts! How soon they die!</l>
26     </lg>
27     <lg>
28       <l>In cottage doors no shuttle clicks;</l>
29     </lg>
30     <lg>
31       <l>The hand-loom has been ousted by</l>
32     </lg>
33     <lg>
34       <l>A large concern with lots more sticks.</l>
35     </lg>
36     <lg>
37       <l>The throb of pistons beats around;</l>
38     </lg>
39     <lg>
40       <l>Great chimneys rise on Thames's banks;</l>
41     </lg>
42     <lg>
43       <l>The same phenomena are found</l>
44     </lg>
45     <lg>
46       <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
47     </lg>
48     <lg>
49       <l>No longer now the housewife makes</l>
50     </lg>
51     <lg>
52       <l>'Tis so with what w</l>
53     </lg>
54     <lg>
55       <l>'Tis this that chief</l>
56     </lg>
57     <lg>
58       <l>That not with tend</l>
59     </lg>
60   </body>
61 </text>
62 </TEI>

```

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml U+003C 1:1

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oXygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project exercise3-complete.xml

```

1 <?pxygen RNGSchema="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/teilight.rng" type="xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3 <teiHeader>
4 <fileDesc>
5 <titleStm>
6 <title>Demonstration exercise for Punch</title>
7 </titleStm>
8 <publicationStm>
9 <p>Unpublished</p>
10 </publicationStm>
11 <sourceDesc>
12 <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13 </sourceDesc>
14 </fileDesc>
15 </teiHeader>
16 <text>
17 <body>
18 <div type="verse">
19 <head> PROGRESS.</head>
20 <p>["Giving evidence recently before a Select Committee of the House of Commons,
21 Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22 the small hand laundry."--Evening News.]</p>
23 <lg>
24 <l>The little crafts! How soon they die! </l>
25 <l>In cottage doors no shuttle clicks; </l>
26 <l>The hand-loom has been ousted by </l>
27 <l>A large concern with lots more sticks.</l>
28 </lg>
29 <lg>
30 <l>The throb of pistons beats around; </l>
31 <l>Great chimneys rise on Thames's banks; </l>
32 <l>The same phenomena are found </l>
33 <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34 </lg>
35 <lg>
36 <l>No longer now the housewife makes </l>
37 <l>Her rare preserves, for what's the good? </l>

```

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r
 - lg The little crafts! Hk
 - lg The throb of pistor
 - lg No longer now the
 - lg 'Tis so with what w
 - lg 'Tis this that chiefh
 - lg That not with tend

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml U+003C 1:1

The screenshot displays the Oxygen XML Editor interface. The main window shows the XML document `exercise3-complete.xml` with the following content:

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStm>
6         <title>Demonstration exercise for Punch</title>
7       </titleStm>
8       <publicationStm>
9         <p>Unpublished</p>
10      </publicationStm>
11     <sourceDesc>
12       <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13     </sourceDesc>
14   </fileDesc>
15 </teiHeader>
16 <text>
17   <body>
18     <div type="verse">
19       <head> PROGRESS.</head>
20       <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23       <lg>
24         <l>The little crafts! How soon they die! </l>
25         <l>In cottage doors no shuttle clicks; </l>
26         <l>The hand-loom has been ousted by </l>
27         <l>A large concern with lots more sticks.</l>
28       </lg>
29       <lg>
30         <l>The throb of pistons beats around; </l>
31         <l>Great chimneys rise on Thames's banks; </l>
32         <l>The same phenomena are found </l>
33         <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34       </lg>
35       <lg>
36         <l>No longer now the housewife makes </l>
37         <l>Her rare preserves, for what's the good? </l>

```

Annotations in the image include:

- A red circle around the `<div type="verse">` element in the tree view.
- A red circle around the `<div type="verse">` element in the main editor window.
- A blue circle around the `<div type="verse">` element in the main editor window.
- A green circle around the scroll bar in the main editor window.

The status bar at the bottom shows the file path: `D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml`, the ID: `U+003C`, and the line number: `1:1`.

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oXygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project exercise3-complete.xml

```

1 <?pxygen RNGSchema="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/teilight.rng" type="xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3 <teiHeader>
4 <fileDesc>
5 <titleStm>
6 <title>Demonstration exercise for Punch</title>
7 </titleStm>
8 <publicationStm>
9 <p>Unpublished</p>
10 </publicationStm>
11 <sourceDesc>
12 <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13 </sourceDesc>
14 </fileDesc>
15 </teiHeader>
16 <text>
17 <body>
18 <div type="verse">
19 <head> PROGRESS.</head>
20 <p>["Giving evidence recently before a Select Committee of the House of Commons,
21 Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22 the small hand laundry."--Evening News.]</p>
23 </p>
24 </div>
25 <div type="verse">
26 <p>The little crafts! How soon they die! </p>
27 <p>In cottage doors no shuttle clicks; </p>
28 <p>The hand-loom has been ousted by </p>
29 <p>A large concern with lots more sticks.</p>
30 </div>
31 <div type="verse">
32 <p>The throb of pistons beats around; </p>
33 <p>Great chimneys rise on Thames's banks; </p>
34 <p>The same phenomena are found </p>
35 <p>In Sheffield. (Yorks) and Oldham (Lancs).</p>
36 </div>
37 <div type="verse">
38 <p>No longer now the housewife makes </p>
39 <p>Her rare preserves, for what's the good? </p>
40 </div>
41 </body>
42 </text>
43 </TEI>

```

Outline

- TEI http://www.tei-c.org/ns/1.0
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r...
 - lg The little crafts! Hk...
 - lg The throb of pistor...
 - lg No longer now the...
 - lg 'Tis so with what w...
 - lg 'Tis this that chief...
 - lg That not with tend...

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml | U+003C | 1:1

The screenshot shows the Oxygen XML Editor interface. The main window displays an XML document with the following structure:

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>Unpublished</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <div type="verse">
19        <head> PROGRESS.</head>
20        <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23        <lg>
24          <l><|>The little crafts! How soon they die! </l>
25          <l><|>In cottage doors no shuttle clicks; </l>
26          <l><|>The hand-loom has been ousted by </l>
27          <l><|>A large concern with lots more sticks.</l>
28        </lg>
29        <lg>
30          <l><|> The throb of pistons beats around; </l>
31          <l><|>Great chimneys rise on Thames's banks; </l>
32          <l><|>The same phenomena are found </l>
33          <l><|>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34        </lg>
35        <lg>
36          <l><|>No longer now the housewife makes </l>
37          <l><|>Her rare preserves, for what's the good? </l>

```

The Outline view on the left shows the following structure:

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - fileDesc
 - titleStmt
 - title
 - publicationStmt
 - p
 - sourceDesc
 - p
 - text
 - body
 - div "verse"
 - head
 - p
 - lg
 - l
 - l
 - l
 - l
 - lg
 - l
 - l
 - l
 - l
 - lg
 - l
 - l

Annotations in the image include:

- A red circle around the Outline view.
- A blue circle around the 'Save' icon in the toolbar.
- A green circle around the 'Home' icon in the toolbar.

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oXygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project

exercise3-complete.xml

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>Unpublished</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <div type="verse">
19        <head> PROGRESS.</head>
20        <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23      </div>
24      <lg>
25        <l><|>The little crafts! How soon they die! </l>
26        <l><|>In cottage doors no shuttle clicks; </l>
27        <l><|>The hand-loom has been ousted by </l>
28        <l><|>A large concern with lots more sticks.</l>
29      </lg>
30      <lg>
31        <l><|> The throb of pistons beats around; </l>
32        <l><|>Great chimneys rise on Thames's banks; </l>
33        <l><|>The same phenomena are found </l>
34        <l><|>In Sheffield. (Yorks) and Oldham (Lancs).</l>
35      </lg>
36      <lg>
37        <l><|>No longer now the housewife makes </l>
38        <l><|>Her rare preserves, for what's the good? </l>

```

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - fileDesc
 - titleStmt
 - title
 - publicationStmt
 - p
 - sourceDesc
 - p
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r
 - lg The little crafts! H
 - lg The throb of pisto
 - lg No longer now the
 - lg 'Tis so with what w
 - lg 'Tis this that chief
 - lg That not with te

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml | U+003C | 1:1

The screenshot shows the Oxygen XML Editor interface. The main editor window displays the following XML code:

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>Unpublished</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <div type="verse">
19        <head> PROGRESS.</head>
20        <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23      </div>
24      <lg>
25        <l><|>The little crafts! How soon they die! </l>
26        <l><|>In cottage doors no shuttle clicks; </l>
27        <l><|>The hand-loom has been ousted by </l>
28        <l><|>A large concern with lots more sticks.</l>
29      </lg>
30      <lg>
31        <l><|> The throb of pistons beats around; </l>
32        <l><|>Great chimneys rise on Thames's banks; </l>
33        <l><|>The same phenomena are found </l>
34        <l><|>In Sheffield. (Yorks) and Oldham (Lancs).</l>
35      </lg>
36      <lg>
37        <l><|>No longer now the housewife makes </l>
38        <l><|>Her rare preserves, for what's the good? </l>

```

The Outline view at the bottom left shows the following structure:

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - fileDesc
 - titleStmt
 - title
 - publicationStmt
 - p
 - sourceDesc
 - p
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r...
 - lg The little crafts! H...
 - lg The throb of pisto...
 - lg No longer now the...
 - lg 'Tis so with what w...
 - lg 'Tis this that chief...
 - lg That not with te...

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oxygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project

exercise3-complete.xml

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>Unpublished</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <div type="verse">
19        <head> PROGRESS.</head>
20        <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23      </div>
24      <lg>
25        <l><|>The little crafts! How soon they die! </l>
26        <l><|>In cottage doors no shuttle clicks; </l>
27        <l><|>The hand-loom has been ousted by </l>
28        <l><|>A large concern with lots more sticks.</l>
29      </lg>
30      <lg>
31        <l><|> The throb of pistons beats around; </l>
32        <l><|>Great chimneys rise on Thames's banks; </l>
33        <l><|>The same phenomena are found </l>
34        <l><|>In Sheffield. (Yorks) and Oldham (Lancs).</l>
35      </lg>
36      <lg>
37        <l><|>No longer now the housewife makes </l>
38        <l><|>Her rare preserves, for what's the good? </l>

```

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r...
 - lg The little crafts! H...
 - lg The throb of pisto...
 - lg No longer now the...
 - lg 'Tis so with what w...
 - lg 'Tis this that chief...
 - lg That not with te...

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml | U+003C | 1:1

exercise3-complete.xml [D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oXygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project

exercise3-complete.xml

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>Unpublished</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <div type="verse">
19        <head> PROGRESS.</head>
20        <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23        <lg>
24          <l>The little crafts! How soon they die! </l>
25          <l>In cottage doors no shuttle clicks; </l>
26          <l>The hand-loom has been ousted by </l>
27          <l>A large concern with lots more sticks.</l>
28        </lg>
29        <lg>
30          <l>The throb of pistons beats around; </l>
31          <l>Great chimneys rise on Thames's banks; </l>
32          <l>The same phenomena are found </l>
33          <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34        </lg>
35        <lg>
36          <l>No longer now the housewife makes </l>
37          <l>Her rare preserves, for what's the good? </l>

```

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r...
 - lg The little crafts! H...
 - lg The throb of pisto...
 - lg No longer now the...
 - lg 'Tis so with what w...
 - lg 'Tis this that chief...
 - lg That not with te...

Text Grid Author

D:\data\work\esu\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml | U+003C | 1:1



Übung – Trial 'n' Error

- Ziel: Weitere XML-Regeln finden und dokumentieren unter <http://piratepad.net/xmlregeln>
- Zeit: 10 Minuten
- Arbeitsform: zu zweit (rechte/r NachbarIn)
- Material: [wiki/Text_Encoding_Initiative#Praxisbeispiel](#)
- Handwerkszeug:
 - <Oxygen/>
 - Texteditor
 - oxGarage (www.oucs.ox.ac.uk/oxgarage/)



Überblick

- Was ist XML?
- <oXygen/>
- **Grundregeln**
- Wege nach XML
- XML-Bereiche für digitale Edition
- (Warum überhaupt?)



XML-Dokument-Regeln

- Zeichenkette **xml** darf nicht zu Beginn des Elementnamens stehen.
- **< > & ' #** und **"** sind im Elementnamen nicht erlaubt.
- beliebig viele Zeichen.
- Hinweis: sprechende Namen verwenden.
- Elemente können durch **Attribute** näher bestimmt werden.
- Attribute stehen im Start-Tag.

```
<name AttrName="Attributwert">Inhalt</name>  
<leerelement AttrName="Wert" />
```



XML-Dokument-Regeln

- Elementnamen müssen mit einem **Buchstaben**, **Unterstrich** oder **Doppelpunkt** beginnen.
- Elementnamen können **Buchstaben**, **Zahlen**, **Bindestriche**, **Punkte** oder **Unterstriche** enthalten sowie **Umlaute** und **Akzente**.



XML-Dokument-Regeln

- Weitere Regeln können in der „Grammatik“ eines XML-Schemas angegeben sein
 - Wertbeschränkungen
 - Aufeinanderfolge von Elementen
 - Muss- und Kannelemente bzw. -attribute
 - ...



XML-Dokument-Regeln

- Bei der Dokumentmodellierung muss immer abgewogen werden, welche Informationen elementwürdig, welche attributwürdig sind.
- Mehrere Attribute sind pro Element erlaubt.
- Gleich benannte Attribute in verschiedenen Elementen erlaubt, keine Wiederholung desselben Attributs in einem Element.



XML Check

- Die Prüfung erledigt ein XML-Parser.
- Ein XML Dokument **muss wohlgeformt** sein und **kann valide** sein.
- XML-Dokument prüfen auf Wohlgeformtheit
 - Ein XML-Dokument ist **wohlgeformt**, wenn es den Regeln des XML-Standards entspricht.
- XML-Dokument prüfen auf Gültigkeit/Validität
 - Ein XML-Dokument ist **valide**, wenn es wohlgeformt ist **und** der Grammatik des XML-Schemas entspricht.



XML Prolog

- XML-Deklaration

```
<xml version="1.0" encoding="utf-8">
```

- Verarbeitungsanweisungen (optional)

```
<?xsl-stylesheet type="text/xsl"  
  href="mein.xsl"?>
```

- Einbindung des Dokumentmodells (optional)

- DTD, XML Schema, RelaxNG, Schematron



XML Prolog

- Dokumenttypdefinition (DTD) einbinden

```
<!DOCTYPE tei SYSTEM "tei.dtd">
```

- XML Schema einbinden

```
<TEI
```

```
  xmlns:xsi=
```

```
  "http://www.w3.org/2001/XMLSchema-instance"
```

```
  xsi:noNamespaceSchemaLocation=
```

```
  "pfad/zu/meinemSchema.xsd">
```



Entitäten

- Entitäten stehen für irgendetwas anderes.
- XML allows to establish relations to the entities using references.
- The XML processor replaces the reference by the entity.
- already reserved entity references:

< <

> >

& &



Namensräume

- Namespaces sind eine der ersten Erweiterungen des XML-Standards und wurden bereits im Februar 1998 veröffentlicht.
- Mit der Angabe des Namensraums wird der Kontext angegeben, in dem ein bestimmter Name gilt.
- Namensräume werden über eine URI-Referenz angegeben.
- Namespaces für Elemente und Attribute.

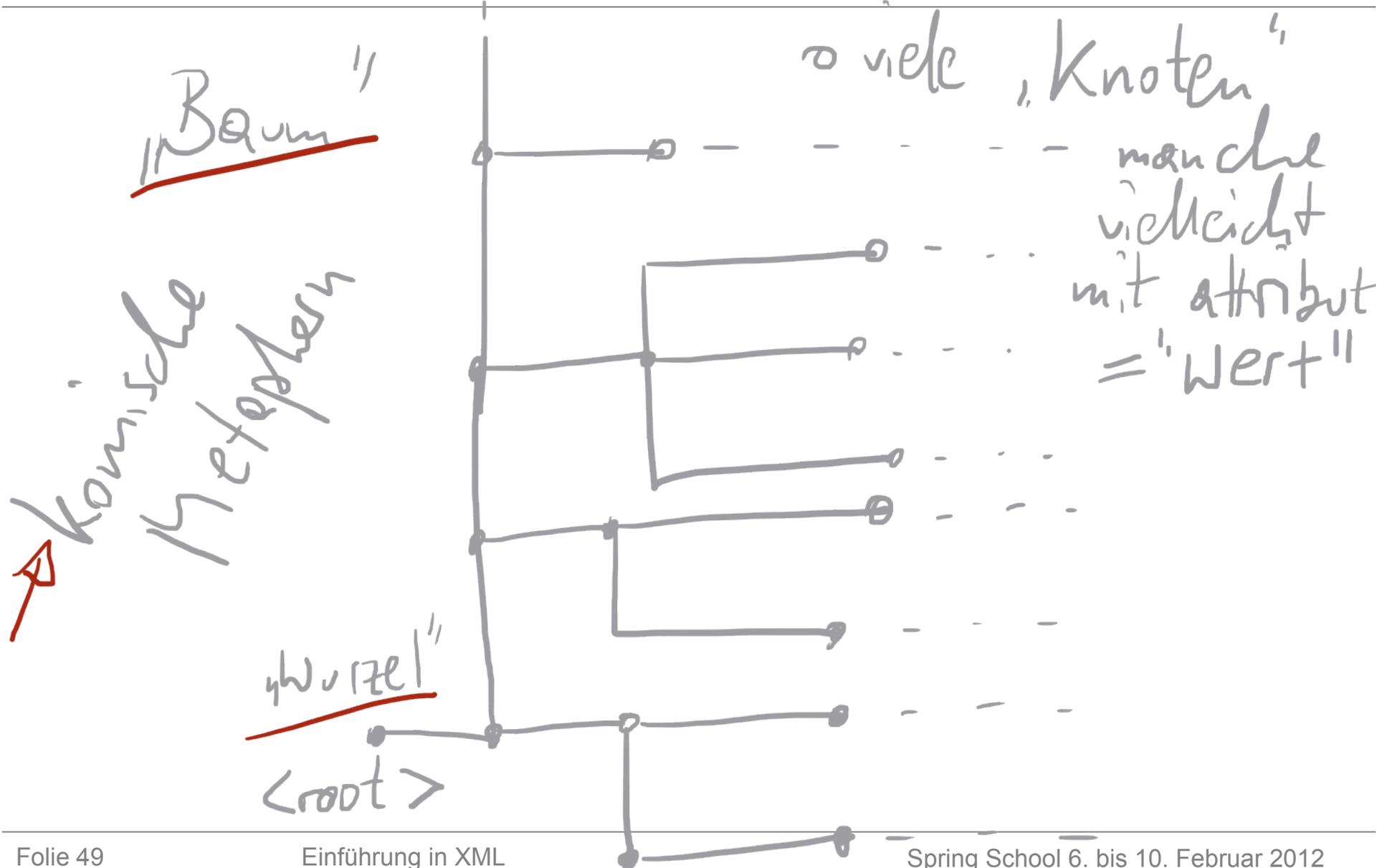


Namensräume

- Zeichenfolge frei wählbar, außer **xml**.
- Beispiel für Deklaration der Namensräume:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0"
xml:lang="en">
```
- Beispiel für Verwendung von Namensräumen :

```
<xsl:template match="msDesc"/>
```





Überblick

- Was ist XML?
- <oXygen/>
- Grundregeln
- **Wege nach XML**
- XML-Bereiche für digitale Edition
- (Warum überhaupt?)



Wege nach XML

- Selberschreiben
- Konvertieren aus anderen Dokumentformaten (z.B. mit oxGarage, oder mit OpenOffice-Exportfiltern oder oder oder)
- Export aus MySQL (z.B. mit phpmyadmin)
- Über Zwischenformate (XML-basiert: z.B. HTML)
- Konverter selberschreiben



Übung

- Heinrich Heines Gedicht „Auf dem Brocken“ in der Lieblingstextverarbeitung (oder einer anderen) mit Formatvorlagen und anderen Formatierungen formatieren
- Mit oxGarage konvertieren nach XML TEI P5
- In <oxygen/> öffnen
- Unterschiede finden zu dem Beispieldokument

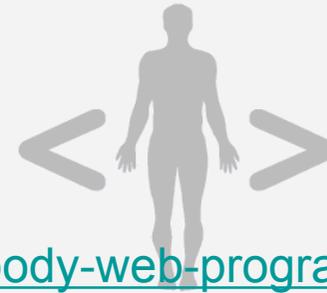


Überblick

- Was ist XML?
- <oXygen/>
- Grundregeln
- Wege nach XML
- **XML-Bereiche für digitale Edition**
- (Warum überhaupt?)



<http://arkarthick.com/2010/12/27/human-body-web-programmed/>



The Human Body in HTML & PHP

```

<human>
  <head>
    <hair /><br />
    <ear align="left" />
    <sight><eye align="left" />
    <eye align="right" /></sight>
    <ear align="right" /><br />
    <nose /><br />
    <form action="alimnet.php"><mouth /></form><br />
    <neck height="8cm" />
  </head>
  <body>
    <tshirt style="background-color: #000;" />
    <arm align="left"><hand /></arm>
    <chestarea><?php if ($sex='female')
{echo '<tit align="left" /><tit align="right" />';}
  else {echo '<nipple align="left" />
    <nipple align="right" />';} ?></chestarea>
    <arm align="right"><hand /></arm>
    <sponsor href="http://www.alvago.com.ar">Alvago Go!</sponsor>
    <br /><tummy><bellybutton /></tummy></tshirt>
    <pants size="short"><underwear>
    <?php include 'private.php'; ?></underwear></pants>
    <leg align="left" />
  <leg align="right" style="tattoo-image: url(img/alvago.gif);" />
    <sneaker align="left" class="nike"><foot /></sneaker>
    <sneaker align="right" class="nike"><foot /></sneaker>
  </body>
</human>

```



XML-Familie

- XML ... beschreibt strukturierte Daten
- XPath ... erlaubt die Navigation in XML-Daten
- XML Schema ... beschreibt ein striktes Datenmodell
- XSL ... eXtensible Style Language
 - XSL-T ... transformiert XML-Dokumente
 - XSL-FO ... beschreibt eine formatierte Ausgabe (z.B. für den Druck)
- XLink ... beschreibt komplexe Links
- XPointer ... beschreibt XML-Zeiger
- XQuery ... ist eine XML-Datenbankabfragesprache
- XForms ... beschreibt Eingabeformulare



XML-Vokabularien

- (X)HTML – Hypertext Markup Language
- **EAD – Encoded Archival Description**
- **TEI – Text Encoding Initiative**
- **CEI – Charters Encoding Initiative**
- MEI – Music Encoding Initiative
- SVG – Scalable Vector Graphics (W3C Standard)
- MathML (W3C Standard)
- CML – Chemical Markup Language
- ...



Informationsquellen

- <http://www.w3c.org/>
- <http://www.xml.org/>
- <http://xml.coverpages.org/>
- <http://www.xmlguru.de/>
- <http://www.xml-web.de/>
- <http://www.zvon.org/>
- TEI by example <http://tbe.kantl.be/TBE/>



Überblick

- Was ist XML?
- <oXygen/>
- Grundregeln
- Wege nach XML
- XML-Bereiche für digitale Edition
- (Warum überhaupt?)



Warum XML?

- Weil's alle tun?
 - Ja
 - Ja, aber
- Standards, Standards, Standards!
- Menschenlesbar
- Strukturadäquat für viele Anwendungen (außer wirklich vieldimensionale oder performancekritische)
- Alternativen sind nicht in Sicht (außer [teure] Insellösungen)
- Warum denn auch nicht?