



# Einführung in XML

Von Klammern, Schachteln und gefüllten  
Keks





# Überblick

- XML Grundlagen
- Das XML-Dokument und seine Regeln
- XML edieren
- XML – und wie weiter?



# XML Grundlagen

- XML heißt eXtensible Markup Language.
- XML ist ein internationaler W3C-Standard.
  - 16 August 2006: 4. Edition ist der aktuell gültige Standard XML

Knoten prüfen Endtag  
validieren Elternknoten  
Wurzelement Namensraum  
Dokumentenmodell  
Auszeichnung Starttag XML  
wohlgeformt Markup valide  
Attributwert Klammer  
modellieren Schema  
Attribut



# Was ist XML?

„XML definiert eine Syntax, um strukturierte Datenbestände jeder Art mit einfachen, verständlichen Auszeichnungen zu versehen, die zugleich von Anwendungen der unterschiedlichsten Art ausgewertet werden können.“

H. Vonhoegen, Einstieg in XML, Bonn :  
Galileo Press, 2007, S. 31

XML-Dokumente sind also menschen- und maschinenlesbar.





# Was ist XML?

- ... eine Dokumenten- oder Inhalts**beschreibungssprache**
- ... **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 Entwickler-gemeinde
- 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.



# Überblick

- XML Grundlagen
- **Das XML-Dokument und seine Regeln**
- XML edieren
- XML – und wie weiter?



# XML ist ...

- XML ist einfach.
- XML ist 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-Dokument Regeln

- Ein Element hat immer einen Start-Tag und einen End-Tag.
  - Das ist ein **<StartTag>**.
  - Das ist das Ende vom **</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.





# 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**.



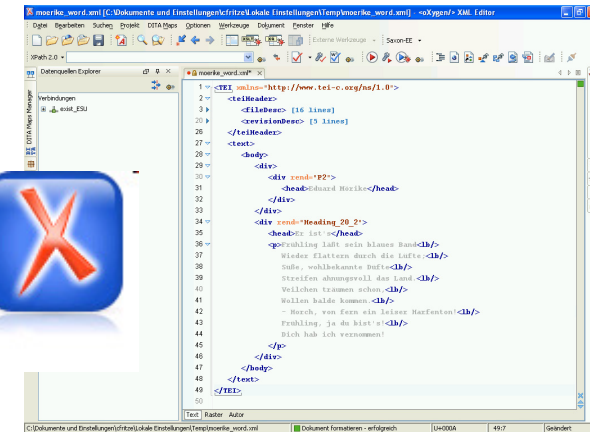
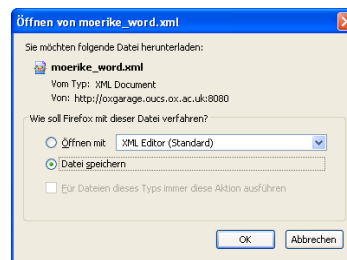
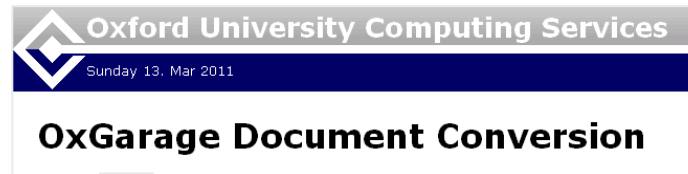
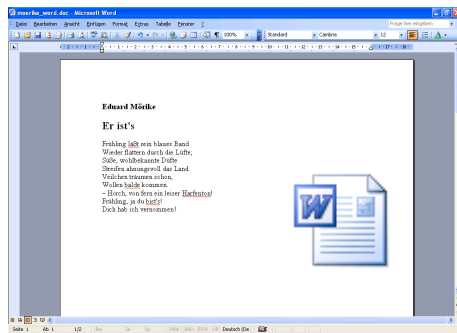


# Überblick

- XML Grundlagen
- Das XML-Dokument und seine Regeln
- **XML edieren**
- XML – und wie weiter?



# Wege von der klassischen Textverarbeitung nach XML



<http://www.oucs.ox.ac.uk/oxgarage/>



# Von Textverarbeitung zu XML

Projekt  Gutenberg-DE

Textsuche bei Gutenberg-DE:

[Startseite](#) [Genres](#) [Neue Texte](#) [Alle Autoren](#) [Alle Werke](#) [Lesetips](#) [Shop](#) [Information](#) [Impressum](#)

Autoren A-Z: [A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#) | [Alle](#)

**Eduard Mörike: Eduard M - Er ist's**

[Quellenangabe](#)

Navigation:

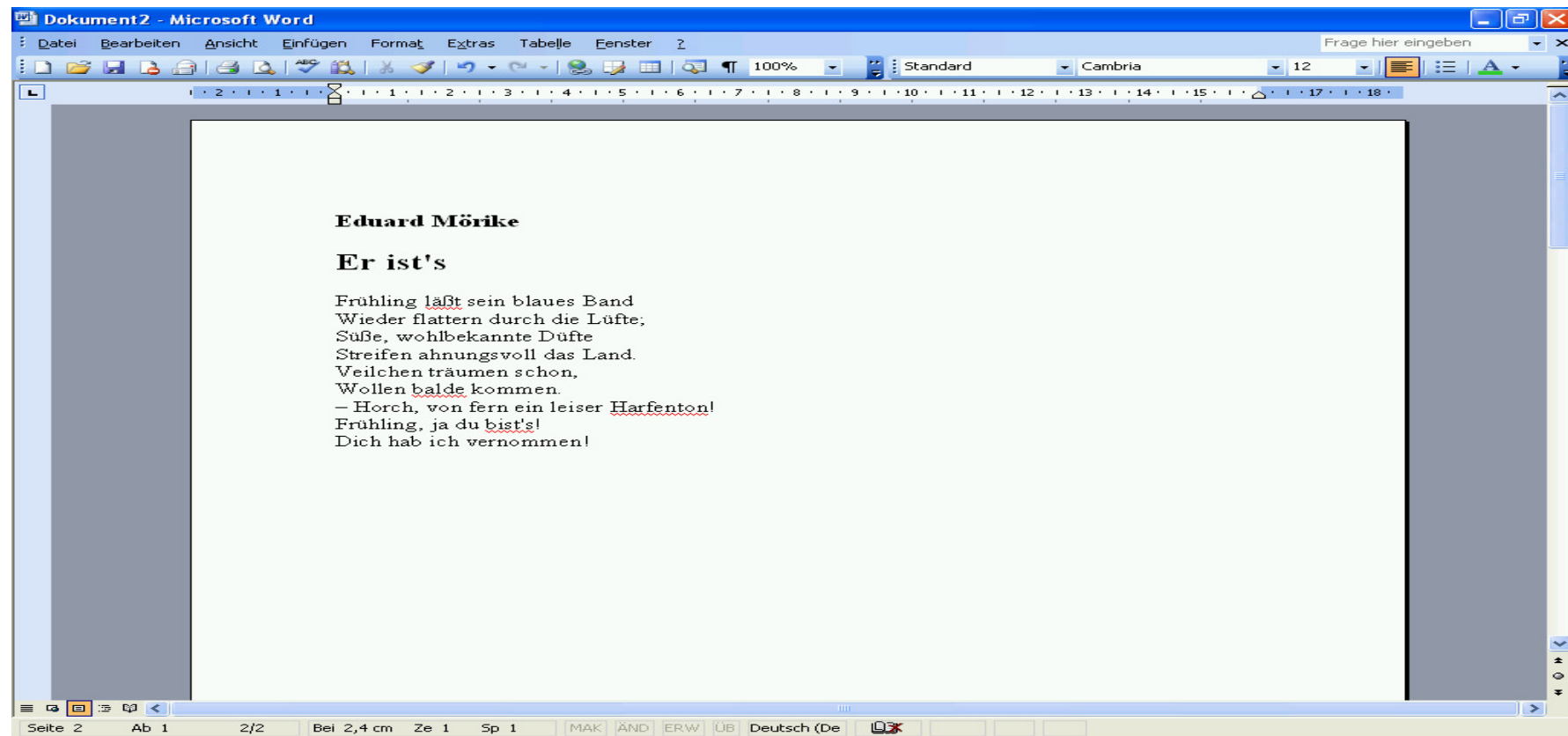
Eduard Mörike

## Er ist's

Frühling läßt sein blaues Band  
Wieder flattern durch die Lüfte;  
Süße, wohlbekannte Düfte  
Streifen ahnungsvoll das Land.  
Veilchen träumen schon,  
Wollen balde kommen.  
– Horch, von fern ein leiser Harfenton!  
Frühling, ja du bist's!  
Dich hab ich vernommen!



# Von Textverarbeitung zu XML







## OxGarage Document Conversion

Please select the type of the document you want to convert.

Convert from: [?](#)



**Presentation documents**



**Spreadsheet documents**



**Text documents**

- Compiled TEI ODD Document
- DocBook Document
- Microsoft Word Document (.doc)
- Microsoft Word Document (.docx)
- ODD Document
- Open Office Text Document (.odt)
- OpenOffice 1.0 Text Document (.sxw)
- Plain Text (.txt)
- Rich Text Format (.rtf)
- TEI P4 XML Document
- TEI P5 XML Document
- TEI Tite XML Document
- WordPerfect Document (.wpd)
- xHTML Document

moerike\_word.xml [C:\Dokumente und Einstellungen\cfritze\Lokale Einstellungen\Temp\moerike\_word.xml] - <oXygen/> XML Editor

Datei Bearbeiten Suchen Projekt DITA Maps Optionen Werkzeuge Dokument Fenster Hilfe

XPath 2.0

Datenquellen Explorer

moerike\_word.xml\*

Verbindungen

exist\_ESU

DITA Maps Manager

Gliederung

Transformationszenarien

```
1 <TEI xmlns="http://www.tei-c.org/ns/1.0">
2   <teiHeader>
3     <fileDesc> [16 lines]
20    <revisionDesc> [5 lines]
26  </teiHeader>
27  <text>
28    <body>
29      <div>
30        <div rend="P2">
31          <head>Eduard Mörike</head>
32        </div>
33      </div>
34      <div rend="Heading_20_2">
35        <head>Er ist's</head>
36        <p>Frühling läßt sein blaues Band<b>
37          Wieder flattern durch die Lüfte,<b>
38          Süße, wohlbekannte Düfte<b>
39          Streifen ahnungsvoll das Land.<b>
40          Veilchen träumen schon,<b>
41          Wollen bald kommen.<b>
42          - Horch, von fern ein leiser Harfenton!<b>
43          Frühling, ja du bist's!<b>
44          Dich hab ich vernommen!
45        </p>
46      </div>
47    </body>
48  </text>
49 </TEI>
50
```

Text Raster Autor

C:\Dokumente und Einstellungen\cfritze\Lokale Einstellungen\Temp\moerike\_word.xml Dokument formatieren - erfolgreich U+000A 49:7 Geändert



# Übung 1

- Such das Frühlingsgedicht von Mörike und speicher es in Deiner Lieblingstextverarbeitung.
- Konvertiere es mit oxGarage nach XML TEI P5.
- Öffne das neue Dokument im oXygen Editor.



# 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
  - Open Office writer
  - ...





# oXygen XML Editor



- Favorit im geisteswissenschaftlichen Bereich
- entwickelt von SynchRo Soft Ltd. (Rumänien)
- Marktführer bei XML Editoren
- klare Lizenzpolitik, günstige Lizenzen für den akademischen Einsatz (USD 64 + MwSt.)
- [www.oxygenxml.com](http://www.oxygenxml.com)



# oXygen XML Editor

- platformunabhängig
- als Eclipse plug-in erhältlich
- subversion client for kollaboratives Arbeiten
- TEI Unterstützung
- Unterstützung aller Schemasprachen
- Syntaxvervollständigung
- Tool-tip Dokumentation
- XSLT 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

- sample.xpr
  - css
  - debugger
  - fo
  - import
  - jsp
  - nvd
  - relaxng
  - schematron
  - svg
  - wSDL
  - xquery
  - dita
  - docbook
  - ooxml
  - tei
  - xhtml

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...

```

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       <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>

```

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 pistor...
        - lg No longer now the...
        - lg 'Tis so with what w...
        - lg 'Tis this that chiefy...
        - 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 <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>

```

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  
nvdI  
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 pistor  
lg No longer now the  
lg 'Tis so with what w  
lg 'Tis this that chiefy  
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       <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       <|>The little crafts! How soon they die! </|>
25       <|>In cottage doors no shuttle clicks; </|>
26       <|>The hand-loom has been ousted by </|>
27       <|>A large concern with lots more sticks.</|>
28     </lg>
29     <lg>
30       <|> The throb of pistons beats around; </|>
31       <|>Great chimneys rise on Thames's banks; </|>
32       <|>The same phenomena are found </|>
33       <|>In Sheffield. (Yorks) and Oldham (Lancs).</|>
34     </lg>
35     <lg>
36       <|>No longer now the housewife makes </|>
37       <|>Her rare preserves, for what's the good? </|>

```

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\esuvorbilder\_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 pistor
        - 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\esuvorbilder\_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

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...

```

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          <|>The little crafts! How soon they die! </|>
25          <|>In cottage doors no shuttle clicks; </|>
26          <|>The hand-loom has been ousted by </|>
27          <|>A large concern with lots more sticks.</|>
28        </lg>
29        <lg>
30          <|> The throb of pistons beats around; </|>
31          <|>Great chimneys rise on Thames's banks; </|>
32          <|>The same phenomena are found </|>
33          <|>In Sheffield. (Yorks) and Oldham (Lancs).</|>
34        </lg>
35        <lg>
36          <|>No longer now the housewife makes </|>
37          <|>Her rare preserves, for what's the good? </|>

```

D:\data\work\esu\esu\vorbilder\_fremd\oxford\_2009\exercise3-complete.xml | U+003C | 1:1

The screenshot shows the Oxygen XML Editor window titled "exercise3-complete.xml [D:\data\work\esu\esu\vorbilder\_fremd\oxford\_2009\exercise3-complete.xml] - <oxygen/> XML Editor". The menu bar includes File, Edit, Find, Project, Perspective, Options, Tools, Document, Window, and Help. The toolbar contains various icons for file operations, XSLT processing, and validation. The Project browser on the left shows a tree structure of files and folders, including "sample.xpr", "css", "debugger", "fo", "import", "jsp", "nvd1", "relaxng", "schematron", "svg", "wsdl", "xquery", "dita", "docbook", "ooxml", "tei", and "xhtml". The Outline view shows the current document structure, with "TEI" selected, and its children "teiHeader" and "text" visible. 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       <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        <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 status bar at the bottom shows the file path "D:\data\work\esu\esu\vorbilder\_fremd\oxford\_2009\exercise3-complete.xml", the encoding "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 <?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       <titleStmnt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmnt>
8       <publicationStmnt>
9         <p>Unpublished</p>
10      </publicationStmnt>
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          <|>The little crafts! How soon they die! </|>
25          <|>In cottage doors no shuttle clicks; </|>
26          <|>The hand-loom has been ousted by </|>
27          <|>A large concern with lots more sticks.</|>
28        </lg>
29        <lg>
30          <|> The throb of pistons beats around; </|>
31          <|>Great chimneys rise on Thames's banks; </|>
32          <|>The same phenomena are found </|>
33          <|>In Sheffield. (Yorks) and Oldham (Lancs).</|>
34        </lg>
35        <lg>
36          <|>No longer now the housewife makes </|>
37          <|>Her rare preserves, for what's the good? </|>

```

Outline

- TEI "http://www.tei-c.org/ns/1.0"
  - teiHeader
    - fileDesc
      - titleStmnt
        - title
      - publicationStmnt
        - p
      - sourceDesc
        - p
  - 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 <?oxygen 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       <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>

```

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 pistor
              - lg No longer now the
                - lg 'Tis so with what w
                  - lg 'Tis this that chiefly
                    - lg That not with tend

Text Grid Author

D:\data\work\esu\esu\vorbilder\_fremd\oxford\_2009\exercise3-complete.xml U+003C Samstag, 25. Juli 2009

The screenshot shows the Oxygen XML Editor interface. The main text area displays the following XML code:

```

1 <?oxygen 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> [12 lines]
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>

```

The Project browser on the left shows a tree structure with folders like 'css', 'debugger', 'fo', 'import', 'jsp', 'nvd', 'relaxng', 'schematron', 'svg', 'wsdl', 'xquery', 'dita', 'docbook', 'ooxml', 'tei', and 'xhtml'. The Outline pane shows a hierarchical view of the XML document structure, including 'TEI', 'teiHeader', 'text', and 'body'.



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
    - fileDesc
      - titleStmt
        - title
- 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       <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
    - fileDesc
      - titleStm
        - title PROGRESS.
      - publicationStm
        - p ["Giving evidence r
      - sourceDesc
        - p The little crafts! H
        - p The throb of pistor
        - p No longer now the
        - p 'Tis so with what w
        - p 'Tis this that chief
        - p 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 window titled "exercise3-complete.xml". The interface includes a menu bar (File, Edit, Find, Project, Perspective, Options, Tools, Document, Window, Help), a toolbar with icons for file operations and editing, and a project browser on the left. The main editing area 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, Miss C. E. Collet, of the Home Office, said the commercial laundry was killing the small hand laundry."--Evening News.]</p>
21      </div>
22      <lg>
23        <l>The little crafts! How soon they die!</l>
24        <l>In cottage doors no shuttle clicks;</l>
25        <l>The hand-loom has been ousted by</l>
26        <l>A large concern with lots more sticks.</l>
27      </lg>
28      <lg>
29        <l>The throb of pistons beats around;</l>
30        <l>Great chimneys rise on Thames's banks;</l>
31        <l>The same phenomena are found</l>
32        <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
33      </lg>
34      <lg>
35        <l>No longer now the housewife makes</l>
36        <l>Her rare preserves, for what's the good?</l>
37      </lg>
38    </body>
39  </text>
40 </TEI>

```

Annotations in the image include:

- Red circles around the `<fileDesc>` element (line 4), the `<title>` element (line 6), and the `<div type="verse">` element (line 18).
- Green circles around the `<title>` element (line 6), the `<sourceDesc>` element (line 11), and the right-hand side of the editor window.
- A red circle around the `Text` tab in the bottom status bar.



# XML-Dokument Regeln

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

**<name AttrName="Attributwert"> </name>**



# 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



# 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

- Documenttypdefinition (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:

**&lt;** <

**&gt;** >

**&amp;** &



# Namensräume

- Namespaces sind eine der ersten Erweiterung 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"/>
```



# Namensräume

- TEI P5-Namespace: <http://www.tei-c.org/ns/1.0>
- XSLT-Namespace:  
<http://www.w3.org/1999/XSL/Transform>
- XSL-FO-Namespace:  
<http://www.w3.org/1999/XSL/Format>
- XML-Namespace:  
<http://www.w3c.org/XML/1998/namespace>
- SVG-Namespace: <http://www.w3.org/2000/svg>



# Überblick

- XML Grundlagen
- Das XML-Dokument und seine Regeln
- XML edieren
- XML – und wie weiter?



# XML Familie

- XML ... beschreibt strukturierte Daten
- XPath ... erlaubt die Navigation in XML-Daten
- XML Schema ... beschreibt ein striktes Datenmodell
- XSL ... eXtensible Style Language
  - XSLT ... 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/>



```
1 <TEI>
2   <teiHeader>
3     <!-- -->
4   </teiHeader>
5   <text>
6     <!-- -->
7   </text>
8 </TEI>
```



```
5 <text>
6   <body>
7     <head>chapter title</head>
8     <p>text goes here</p>
9   </body>
10 </text>
```



# Übung 2

- Formatiere das Gedicht in Deiner Lieblingstextverarbeitung mit Formatvorlagen und Hervorhebungen
- Konvertiere es wieder mit oxGarage nach XML TEI P5.
- Öffne das Dokument im oXygen Editor.
- Finde den Unterschied.