

# Schriftzeichen und Markup Sprachen

# Schriftzeichen und Markup Sprachen - Ziele

- wesentliche Konzepte und Standards aus dem Bereich der Schriftzeichen und Markup Sprachen erklären können
- das Konzept der Metasprache XML erklären können und exemplarisch Sprachen aus der Familie nennen und einordnen können

# Schriftzeichen

- Schriftzeichen (Engl.: "character") sind die kleinsten semantischen Einheiten eines Textes, zum Beispiel Buchstaben, Ziffern, Satzzeichen, Diakritika.
- In einem Zeichensatz (Engl.: "Character Encoding Scheme") werden Schriftzeichen einem bestimmten, jeweils eindeutigen Binärwert zugeordnet.
- Für die Darstellung der Textzeichen eines Zeichensatzes ("Character Encoding Scheme") benötigt man eine Schrift ("Typeface", "Font"), also einen Vorrat von graphischen Zeichen. Eine Schrift ist die Zuordnung von Schriftzeichen zu graphischen Symbolen (glyphes).

# Zeichensätze

- American Standard Code for Information Interchange (ASCII)
  - 8-Bit (früher 7 Bit) Zeichensatz, er an den Zeichen der amerikanischen Sprache orientiert ist
- Extended ASCII
  - untereinander inkompatible Varianten von ASCII, die die Sonderzeichen verschiedener Sprachen berücksichtigen
- ISO Latin-1 (ISO 8859-1)
  - einheitliche 8-Bit Kodierung die viele Textzeichen von nord- und westeuropäischen Sprachen umfasst (8859-2 für osteuropäische Sprachen).
- Unicode (ISO 10646)
  - Eine früher 16, heute bis zu 32 Bit umfassende Kodierung, in der die Schriftzeichen aller bestehenden (und teilweise auch früheren) Kulturen der Welt erfasst werden sollen.
- UCS Transformation Format 16 Bit (UTF-16, RFC 2781)
  - Unicode ist in mehrere Ebenen unterteilt, die Grundlegende wird als Universal Character Set 2" (UCS-2) bezeichnet und kann mittels UTF-16) auf eine variable 1 - 2 Byte Kodierung reduziert werden. Mit UCS-4 werden die höheren Ebenen bezeichnet.
- UCS Transformation Format 8 Bit (UTF-8, RFC 2279)
  - Da viele Anwendungen (noch) auf 1-Byte Kodierung von Zeichen ausgelegt sind wurde mit UTF-8 ein eine variabel 1-4 Byte Kodierung für alle Unicode Zeichen geschaffen, die bei 1Byte Kodierung mit ISO 8859-1 kompatibel ist.

# Zeichensätze: Online Ressourcen

- [www.unicode.org](http://www.unicode.org)
- Dürst, Martin; François Yergeau; Misha Wolf; Asmus Freytag; and Tex Texin; eds. 2001. Character model for the World Wide Web 1.0. W3C working draft 26 January 2001. Cambridge, MA: The World Wide Web Consortium. <http://www.w3.org/TR/charmod/>.
- Implementing Writing Systems: An introduction; [http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsi&item\\_id=IWS-Chapter03](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=IWS-Chapter03)

# Markup Language

- Deutsch: Auszeichnungssprache
- Ein System zur Annotation von Texten, bei dem Annotation und Text syntaktisch unterschieden werden können.
- Geht auf die Zusammenarbeit von Editoren und Druckern zurück.
- IBM GML --> SGML --> XML
- troff --> TeX --> LaTeX

# Typen von Markup Languages

- **Präsentations Markup:** Eingebettete Codes steuern die Präsentation (z.B. Fett, Zeilenabstand ...)
- **Prozedurale Markup:** Anweisungen für ein sequenzielles Verarbeitungsprogramm (z.B. Formatierer) werden in den Text eingebettet.
- **Deskriptiver Markup:** Eigenschaften von Teilen eines Dokumentes werden spezifiziert, um die Spezifika der Verarbeitung von der Struktur und deren Bedeutung zu entkoppeln.

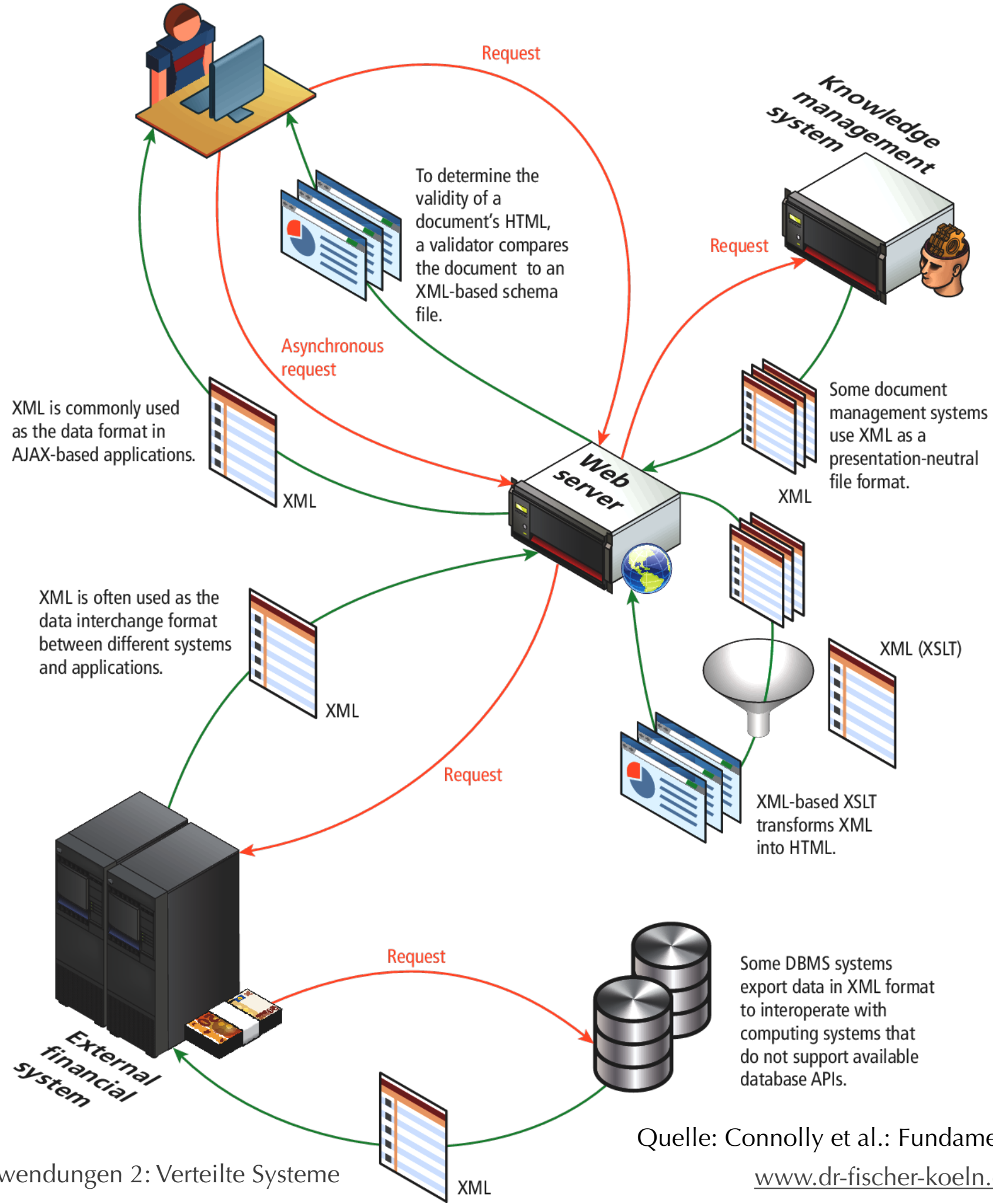
# EXtensible Markup Language

Metasprache

Werkzeuge sind universell

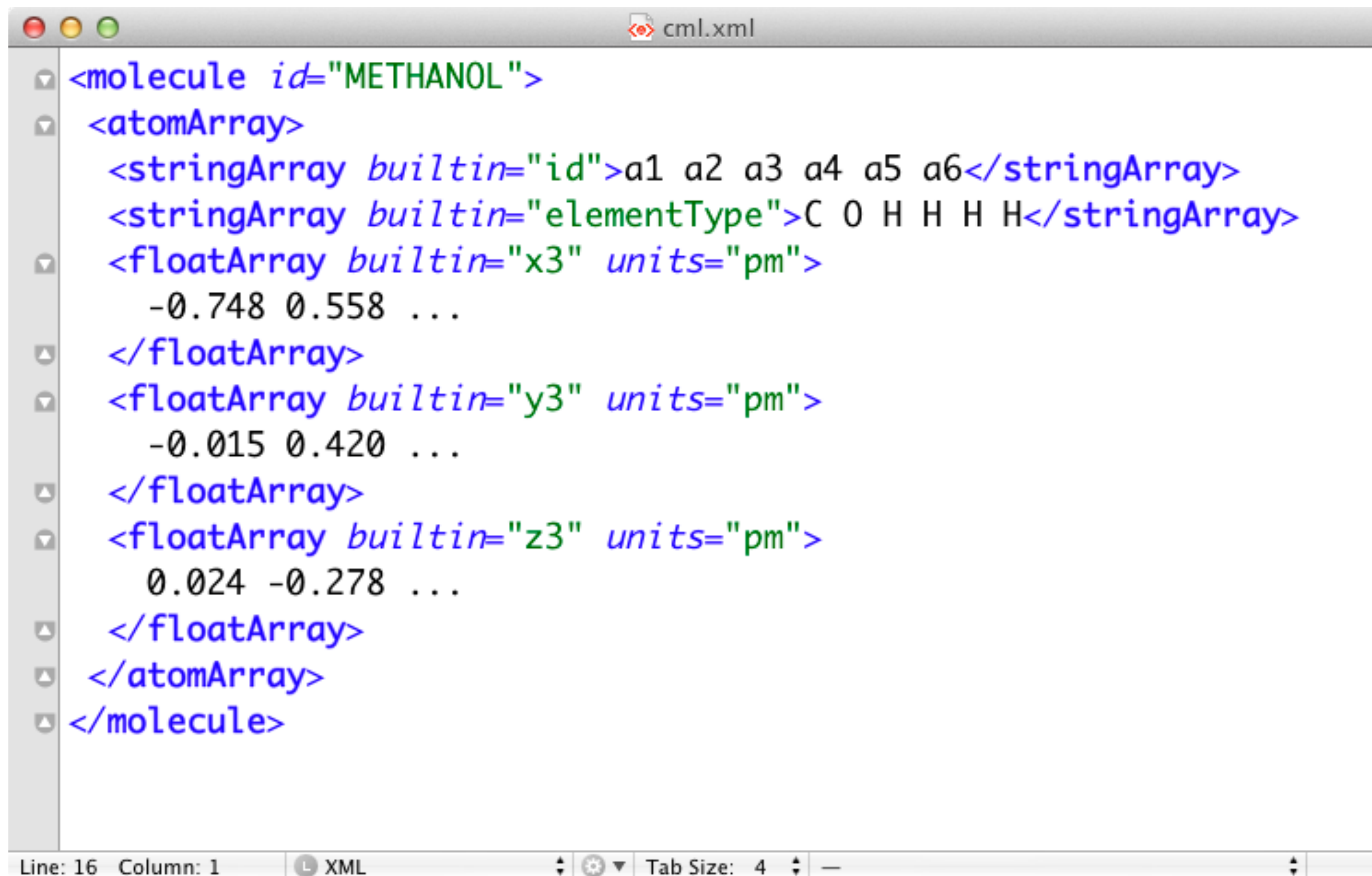
W3C





Quelle: Connolly et al.: Fundamentals of Web Development

# Chemical Markup Language (CML)

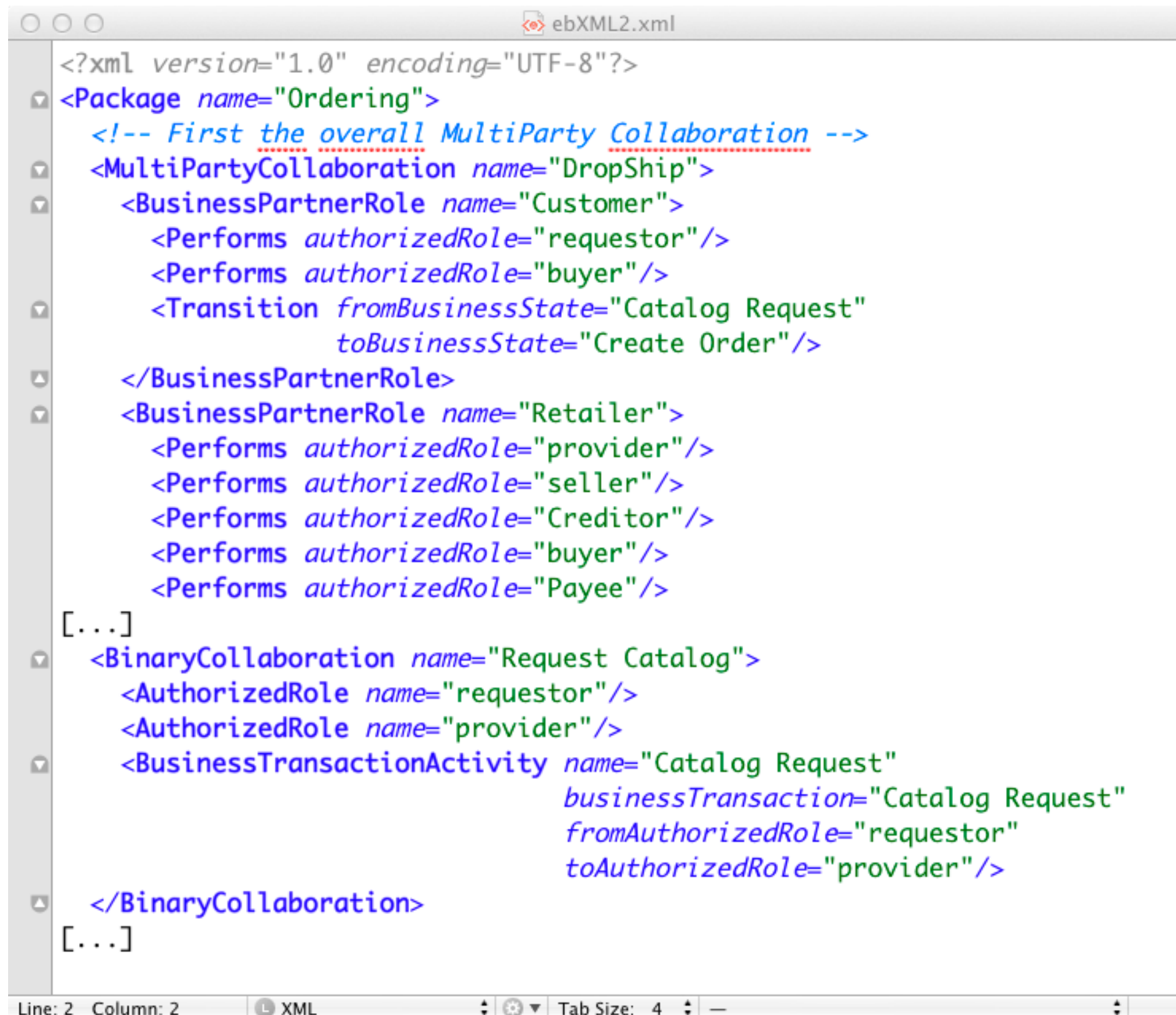


The image shows a screenshot of a text editor window titled 'cml.xml'. The editor displays an XML document for a methanol molecule. The XML structure is as follows:

```
<molecule id="METHANOL">
  <atomArray>
    <stringArray builtin="id">a1 a2 a3 a4 a5 a6</stringArray>
    <stringArray builtin="elementType">C O H H H H</stringArray>
    <floatArray builtin="x3" units="pm">
      -0.748 0.558 ...
    </floatArray>
    <floatArray builtin="y3" units="pm">
      -0.015 0.420 ...
    </floatArray>
    <floatArray builtin="z3" units="pm">
      0.024 -0.278 ...
    </floatArray>
  </atomArray>
</molecule>
```

The editor interface includes a sidebar on the left with expand/collapse icons for each XML element. The status bar at the bottom indicates 'Line: 16 Column: 1', 'XML' encoding, and 'Tab Size: 4'.

# e-business XML (ebXML)

A screenshot of an XML editor window titled 'ebXML2.xml'. The editor displays an ebXML document with the following structure: a root element with XML declaration, a <Package name='Ordering'> element, a comment <!-- First the overall MultiParty Collaboration -->, a <MultiPartyCollaboration name='DropShip'> element containing two <BusinessPartnerRole> elements (one for 'Customer' with roles 'requestor' and 'buyer', and a transition from 'Catalog Request' to 'Create Order'; the other for 'Retailer' with roles 'provider', 'seller', 'creditor', 'buyer', and 'payee'), followed by an ellipsis [...], a <BinaryCollaboration name='Request Catalog'> element containing two <AuthorizedRole> elements ('requestor' and 'provider') and a <BusinessTransactionActivity> element with name 'Catalog Request' and transaction details, followed by another ellipsis [...]. The editor interface includes a sidebar with expand/collapse icons, a status bar at the bottom showing 'Line: 2 Column: 2', 'XML', 'Tab Size: 4', and a scrollbar on the right.

```
<?xml version="1.0" encoding="UTF-8"?>
<Package name="Ordering">
  <!-- First the overall MultiParty Collaboration -->
  <MultiPartyCollaboration name="DropShip">
    <BusinessPartnerRole name="Customer">
      <Performs authorizedRole="requestor"/>
      <Performs authorizedRole="buyer"/>
      <Transition fromBusinessState="Catalog Request"
        toBusinessState="Create Order"/>
    </BusinessPartnerRole>
    <BusinessPartnerRole name="Retailer">
      <Performs authorizedRole="provider"/>
      <Performs authorizedRole="seller"/>
      <Performs authorizedRole="Creditor"/>
      <Performs authorizedRole="buyer"/>
      <Performs authorizedRole="Payee"/>
    </BusinessPartnerRole>
    [...]
  <BinaryCollaboration name="Request Catalog">
    <AuthorizedRole name="requestor"/>
    <AuthorizedRole name="provider"/>
    <BusinessTransactionActivity name="Catalog Request"
      businessTransaction="Catalog Request"
      fromAuthorizedRole="requestor"
      toAuthorizedRole="provider"/>
  </BinaryCollaboration>
  [...]
</Package>
```

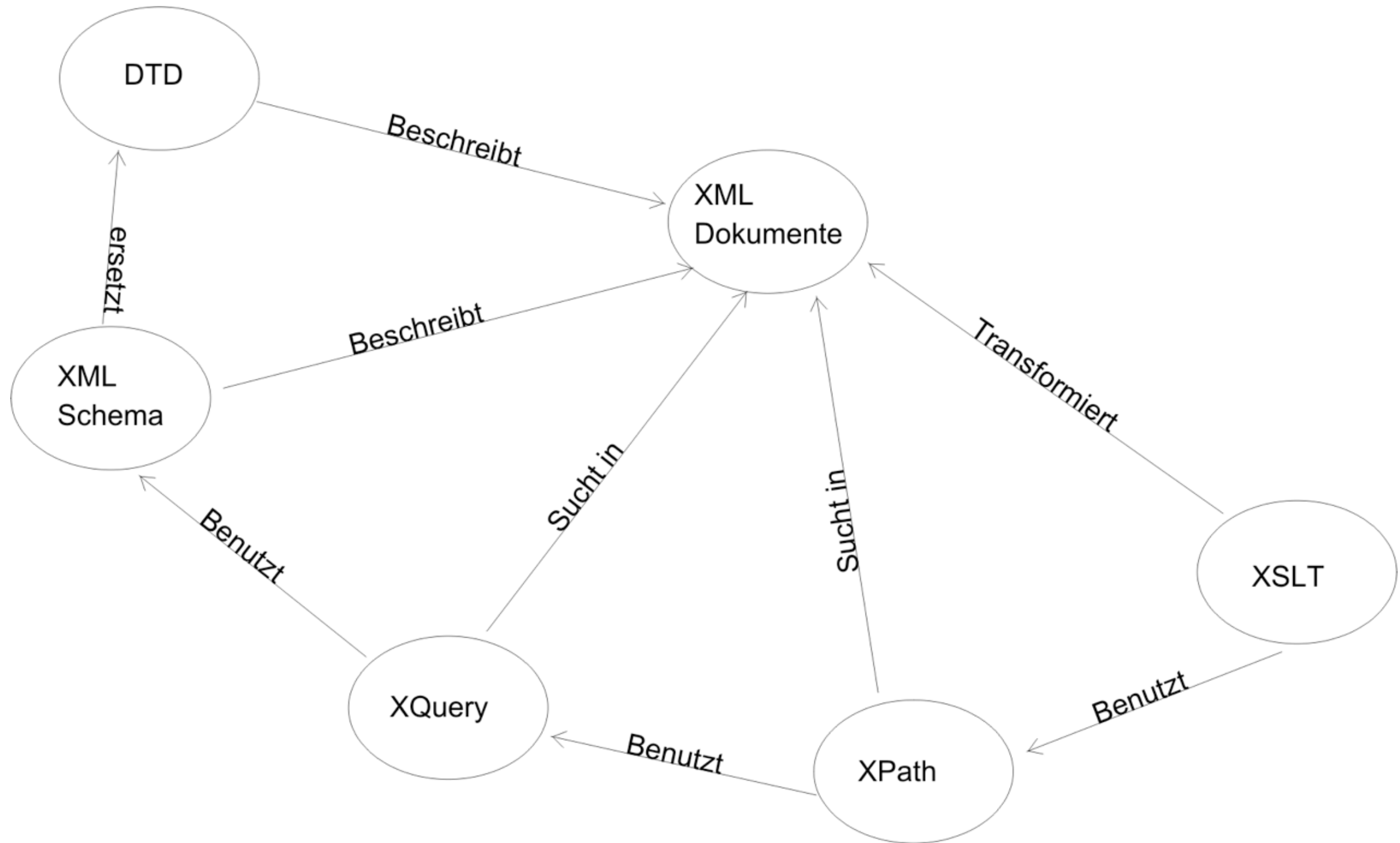
# Theological Markup Language (ThML)

```
ThML.xml
<?xml version="1.0" encoding="UTF-8"?>
<verse>
<l class="t1">O God, a world of empty show,</l>
<l class="t2">Dark wilds of restless, fruitless quest</l>
<l class="t1">Lie round me wheresoe'er I go: </l>
<l class="t3">Within, with Thee, is rest.</l>
</verse><verse>
<l class="t1">And sated with the weary sum</l>
<l class="t2">Of all men think, and hear, and see, </l>
<l class="t1">O more than mother's heart, I come, </l>
<l class="t3">A tired child to Thee. </l>
</verse><verse>
<l class="t1">Sweet childhood of eternal life! </l>
<l class="t2">Whilst troubled days and years go by, </l>
<l class="t1">In stillness hushed from stir and strife, </l>
<l class="t3">Within Thine Arms I lie. </l>
</verse><verse>
<l class="t1">Thine Arms, to whom I turn and cling</l>
<l class="t2">With thirsting soul that longs for Thee; </l>
<l class="t1">As rain that makes the pastures sing, </l>
<l class="t3">Art Thou, my God, to me. </l>
</verse>
<attr><name>G. Ter Steegen</name></attr>
```

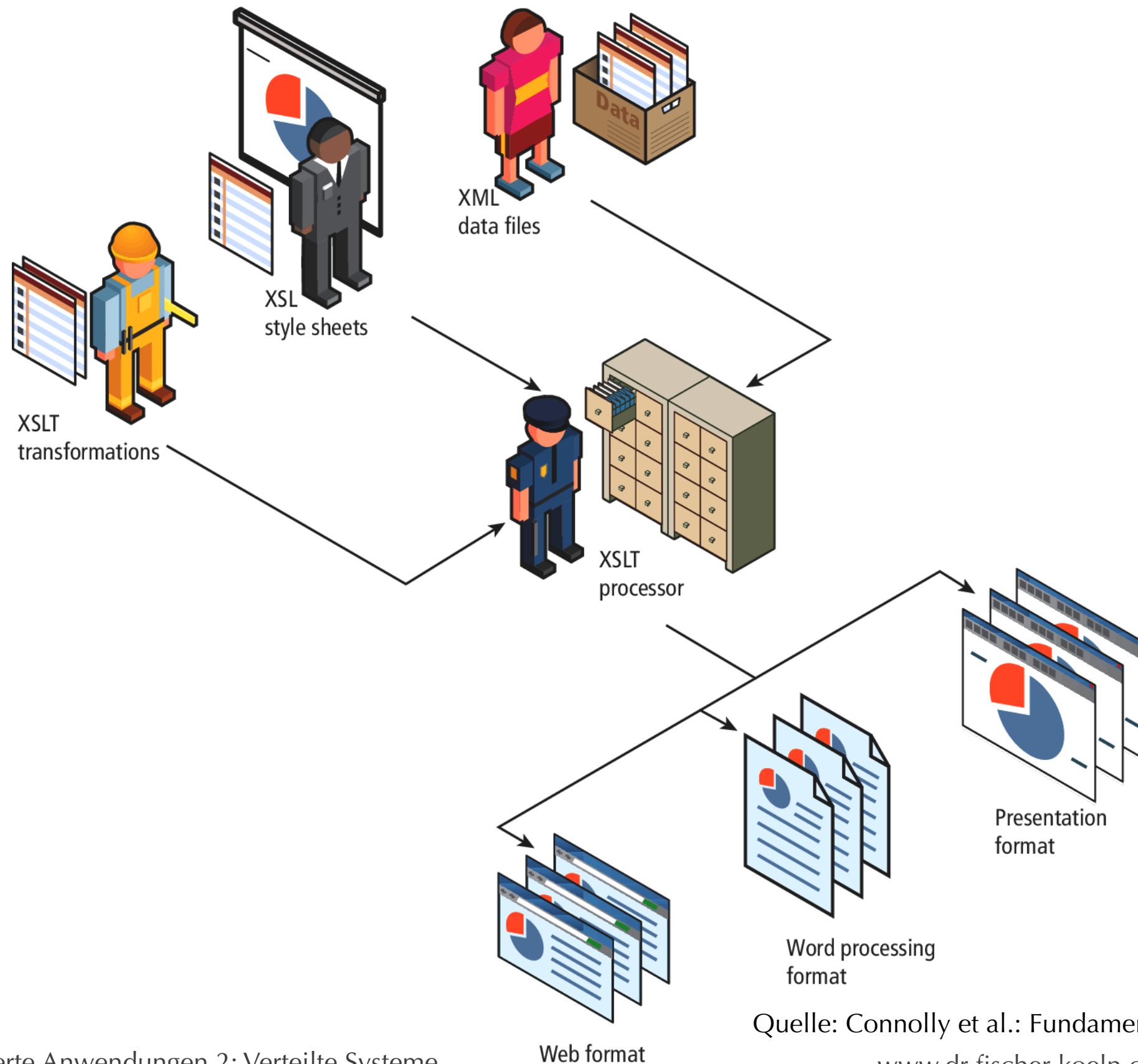
Quelle: <http://www.ccel.org/ThML/ThML1.04.htm>  
(Abruf 16.2.2012)



# Beziehung zwischen XML Spezifikationen



# XML Stylesheet Transformations (XSLT)



```
<?xml version="1.0" encoding="ISO-8859-1"?>
<art>
  <painting id="290">
    <title>Balcony</title>
    <artist>
      <name>Manet</name>
      <nationality>France</nationality>
    </artist>
    <year>1868</year>
    <medium>Oil on canvas</medium>
  </painting>
  <painting id="192">
    <title>The Kiss</title>
    <artist>
      <name>Klimt</name>
      <nationality>Austria</nationality>
    </artist>
    <year>1907</year>
    <medium>Oil and gold on canvas</medium>
  </painting>
  <painting id="139">
    <title>The Oath of the Horatii</title>
    <artist>
      <name>David</name>
      <nationality>France</nationality>
    </artist>
    <year>1784</year>
    <medium>Oil on canvas</medium>
  </painting>
</art>
```

# XSLT Transformation

Listing17.04.xml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<html xsl:version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  • xmlns="http://www.w3.org/1999/xhtml">
  <body>
    <h1>Catalog</h1>
    <ul>
      <xsl:for-each select="/art/painting">
        <li>
          <h2><xsl:value-of select="title"/></h2>
          <p>By: <xsl:value-of select="artist/name"/><br/>
            Year: <xsl:value-of select="year"/>
            [<xsl:value-of select="medium"/>]</p>
        </li>
      </xsl:for-each>
    </ul>
  </body>
</html>
```

Line: 1 Column: 1

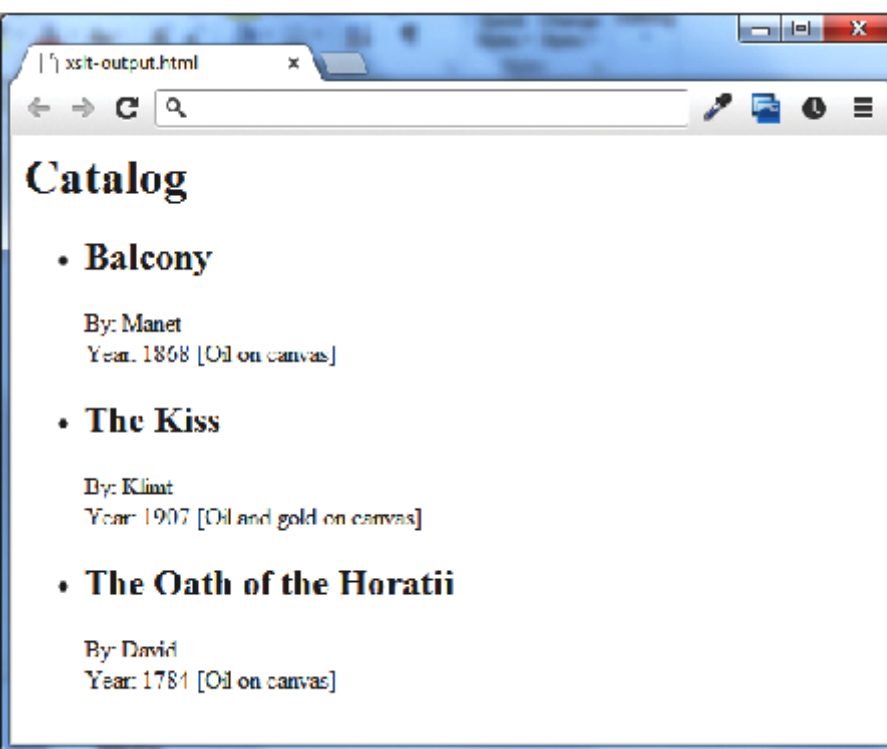
XML

Tab Size: 4

Quelle: Connolly et al.: Fundamentals of Web Development



# Ergebnis der XSLT Transformation



```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml">
<body>
<h1>Catalog</h1>
<ul>
  <li>
    <h2>Balcony</h2>
    <p>By: Manet<br/>
    Year: 1868 [Oil on canvas]</p>
  </li>
  <li>
    <h2>The Kiss</h2>
    <p>By: Klimt<br/>
    Year: 1907 [Oil and gold on canvas]</p>
  </li>
  <li>
    <h2>The Oath of the Horatii</h2>
    <p>By: David<br/>Year: 1784 [Oil on canvas]</p>
  </li>
</ul>
</body>
</html>
```

# Schriftzeichen und Markup Sprachen - Zusammenfassung

- Konzepte und Standards aus dem Bereich der Schriftzeichen und Markup Sprachen
- Konzept der Metasprache XML und exemplarische Sprachen aus der XML Familie