## **Transforming XML**

Elli Mylonas elli.mylonas@gmail.com





Edvard Munch. The Day After, 1894-95.

#### **Powerful Tools and Platforms**

- Easy and convenient when used as intended
- Can be demanding if you want to customize appearance or behavior.
- Occasionally, it may be difficult to arrange hosting for the tools.
- You may need a simple and perhaps ephemeral solution, for ex. A
  proofreading view, an extracted list of values for quality control or
  a small but global modification to the XML structure or content.

#### **Alternative Scenarios**

- Low budget, low tech solutions
- Small interventions in existing frameworks
- Minimal Computing (DHQ 16, 2022
   <a href="https://dhq.digitalhumanities.org/vol/16/2/index.ht">https://dhq.digitalhumanities.org/vol/16/2/index.ht</a>
   <a href="ml">ml</a>)
- Note: Labor is almost always the same, what changes is how it is distributed

### Do this yourself

- XSLT—An XML aware transformation language that can perform queries (check version - should be using 2.0 or later!)
- XQuery—An XML aware query language that does transformations
- Python with XML library (note XSLT versions may be old)
- Generally, if you want to process XML natively, you need software that can navigate the XML tree - XPath is almost always involved.

Note: Even when using different programming languages, you will be drawing on a similar strategies to perform a transformation.

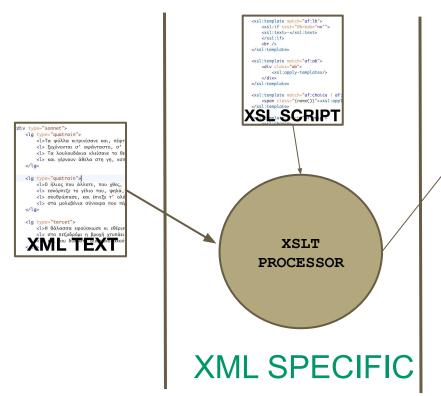
# I Say XSLT, You Say XQuery: Let's Call the Whole Thing Off

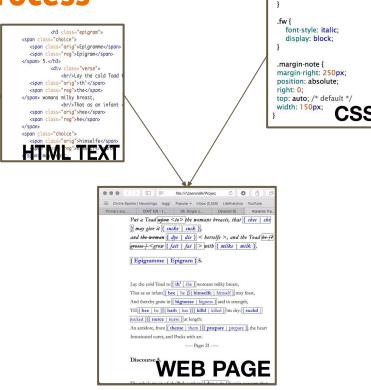
David Birnbaum. Balisage: The Markup Conference 2010

https://www.balisage.net/Proceedings/vol5/author-pkg/Birnb aum01/BalisageVol5-Birnbaum01.html

#### **The Transformation Process** THE DOOT GR - t... D8. Single s... Detailed Si Avalanta 1 Country-Wife, Put a Toad upon <10> the womans breasts, that { thee | she and the woman { dye | die } < berselfe >, and the Toad be [? COMEDY. grosse | < grow { fatt | fat } > with { milke | milk }. { Epigramme | Epigram } 5. THEATRE ROTAL Written by Mr. Wycherley. Lay the cold Toad to [ th' | the ] womans milky breast, That as an infant { hee | he } { himselfe | himself } may feast, And thereby grow in [ bignesse | bigness ] and in strength, Till { hee | he } { hath | has } { killd | killed } his dry: { suckd | sucked ) { nurce | nurse } at length: An antidote, from [ theme | them ] [ præpare | prepare ], the heart LONDON, Printed for Thomas Dring, at the Harrow, at the Gener of Chancery-Lancin Floar-front, 1575. Intoxicated cures, and Pocks with art. div type="sonnet"> WEB PAGE <ld><lg type="quatrain"> <l>Τα φύλλα κιτρινίσανε και, πέφ? <l> ξεχύνονται σ' αφάνταστο, σ' <1> Τα λουλουδάκια κλείσανε το θε <1> και γέρνουν άθελα στη γη, κο <ld><lg type="quatrain"> **Processing** <1>0 ήλιος που άλλοτε, που χθες <l> εσκόρπιζε το γέλιο του, ψηλά, <l> σκυθρώπασε, και έπνιξε τ' ολο <l> στα μολυβένια σύννεφα που πε </10> 43. Nadd Shamma 44. Umm Ramool 63. Al Quez 24. Al Wasi 64. Al Barsha I <ld><lg type="tercet"> 25. Hor Al Anz 26. Hor Al Anz East 45. Nadd Al Hamar 66. Al Barsha III <1>Η θάλασσα εφούσκωσε κι εθέριε 67. The Lakes 68. Emirates Hills z. Al taffiya 27. Jumeirah First 28. Jumeirah Second 29. Jumeirah Third <1> στο πεζοδρόμι η βροχή χτυπάει ΧΙΜ Ευ διατοι Ελλά το πκού 48. Muhaisnah II 69. The Meadows 70. Jumeirah Island 10. Al Mamza 11. Al Manara 32. Naif 33. Oud Mehta 52. Al Warqa II 53. Al Warqa III 72. The Garden 73. Jumeirah Residence 74. Jobel Ali Industrial J 34. Umm Al Shail 35. Umm Sagaim cute alluring foxy beauteous marvelous enticing gorgeous 75. Jobel Ali Port 16. Al Nahda First 56. Nadd Al Shiba III 7. Dubai Media City 78. Dubai Internet City dazzling ravishing handsome 79. Knowledge Village 19. Al Rippa 59. Ras Al Khor Industrial Area II excellent appealing fair good-looking sightly sexygrandelegants tunning XML INPUT | PROCESSING **COMMON FORMATS**

## **XSLT Example Transformation Process**





.pc {

WEB SPECIFIC

#### **Terms**

**XPath:** basis for all XML processing - language for navigating the XML tree.

**XSLT:** A transformation (programming) language that knows about the XML tree structure, XPath and the document's schema

**HTML**: An XML schema (mostly) that is the markup language used for most web pages.

**CSS**: A styling language used on web pages. It can change the appearance of information on a page.

Javascript: A web programming language



## **Norway's Oldest Wheel**

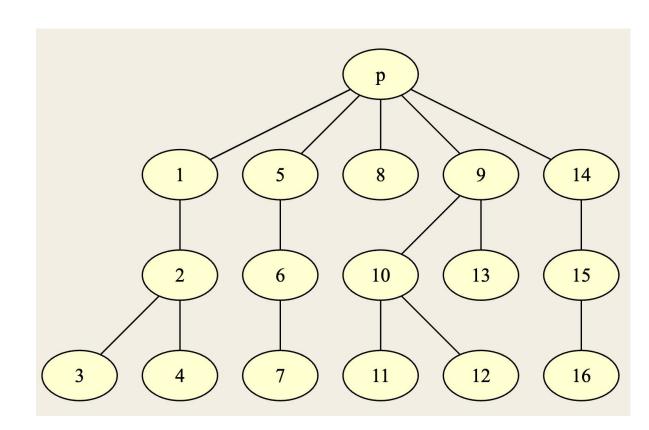
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="letter.rng"?>
<letter type="friendly">
    <salute>Dear Parents/salute>
    <para n="1">I am having a good time a
camp</para>
    <para n="2">Please send cookies</para>
    <closer>hugs and kisses
        <signed>Scooter</signed>
    </closer>
</letter>
```

#### HTML

```
<html>
   <head>
      <title>The Friendly Letter</title>
   </head>
   <body>
      Dear Parents
      [1] I am having a good time a camp
      [2] Please send cookies
      hugs and kisses, Scooter
   </body>
</html>
```

#### **Processing the XML Tree**

The XSLT (or other processor) traverses the XML tree depth first, and acts on each element as it encounters it.



### **XSLT Templates**

```
<xsl:template match="salute">
 <xsl:value-of select="."/>,
</xsl:template>
<xsl:template match="para">
 <q>
   [<xsl:value-of select="@n"/>] <xsl:value-of select="."/>
 </xsl:template>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"</pre>
   xmlns:xs="http://www.w3.org/2001/XMLSchema"
   exclude-result-prefixes="xs" version="2.0">
   <xsl:output method="html"/>
   <xsl:template match="/">
      <html>
                                  <!-- output the HTML element -->
          <xsl:apply-templates/> <!-- Continue processing -->
      </html>
   </xsl:template>
   <xsl:template match="salute"> <!-- found salute, output content of template -->
       <xsl:value-of select="."/>,
   </xsl:template>
   <xsl:template match="para"> <!-- found para, output content of template -->
       [<xsl:value-of select="@n"/>] <xsl:value-of select="."/>
   </xsl:template>
</xsl:stylesheet>
```

## **XSLT** for Proofreading

#### **Atalanta Fugiens from Furnace and Fugue Project**

#### XML file:

https://github.com/Brown-University-Library/atalanta-texts/blob/master/english/emblem06.xml

#### HTML view:

https://github.com/Brown-University-Library/atalanta-texts/blob/master/english-html/emblem06.html

Quick transformation to help encoder check their work.

#### EpigrammeEpigram 6.

RusticksRustics their seed to th'the fertile earth committcommit,
When with their harrows they have made it fittfit:
The Sophi thus their golden seed doedo sowesow
In foliated earth as white as Snow:
This method well observe, and you'leyou'll behold,
As in a glasseglass, by wheat, your budding gold.

Sample proofreading display

#### **XSLT for HTML Formatting**

Two versions of a student assignment working with the same source text.

By permission of Sarah Collins and Alina Weseloh.

the U.S., and had to laugh. We January 28th. Juesday. Got up early. Pretty day, - thawed walked to meeting at the Baptist church considerably. Had full school Good King preached about "The light of the lessons. After supper had discus. world. Three boys went forward. I go sion on "Tregro and Indian" question, Laughed about Dollie and me goin, (Swin, Roy ate an on. Had good school, Sunny in the ion, and I went in the other room and drew map of Italy, Beautiful night, not very cold. Went to bed early. I rode home from school, ennie Dyer and I stood up. Ganerary 29th Low went home at last recess. They laughed about "Scratch" bringing Rose home. It class drew make of Staly, on board. In B. Seog. I asked how high is the highest heak in I replied, that I would rather be

#### **Learn XSLT**

- 1. "Transforming Data for Reuse and Re-publication with XML and XSL." on the Programming Historian site
  - https://programminghistorian.org/en/lessons/transforming-xml-with-xsl#cr
    eating-and-testing-your-xsl-file
- Holmes, Martin. "XSLT" from the DariahTeach "TEI and XSLT" module. <a href="https://teach.dariah.eu/mod/page/view.php?id=454">https://teach.dariah.eu/mod/page/view.php?id=454</a> Start after 5:44 where the speaker starts to talk about XSLT.
- 3. Holmes. "An Introduction to XSLT for Digital Humanists" <a href="https://web.uvic.ca/~mholmes/dhoxss2013/">https://web.uvic.ca/~mholmes/dhoxss2013/</a>
- 4. W3Schools. "XSLT Tutorial" Best used as a reference for syntax and functions. <a href="https://www.w3schools.com/xml/xml">https://www.w3schools.com/xml/xml</a> xslt.asp

## A Few Things to Come Back To

- It is possible to validate and transform XML using XSLT and XQuery from the command line using the Saxon XML software
- Oxygen has helpful videos explaining how to use the software and also about XML concepts and tools
- Look into the summer schools or conference workshops in specific tools

#### Wednesday



#### Thursday



#### **Friday**

