
Transforming XML

Elli Mylonas
elli.mylonas@gmail.com



BærUt!

Sustainable Digital Scholarly Editions

This Presentation: <http://tiny.cc/BaerUT-XSLT>

Too Many Platforms, Too Many Choices!



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
<https://dhq.digitalhumanities.org/vol/16/2/index.html>)
- 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

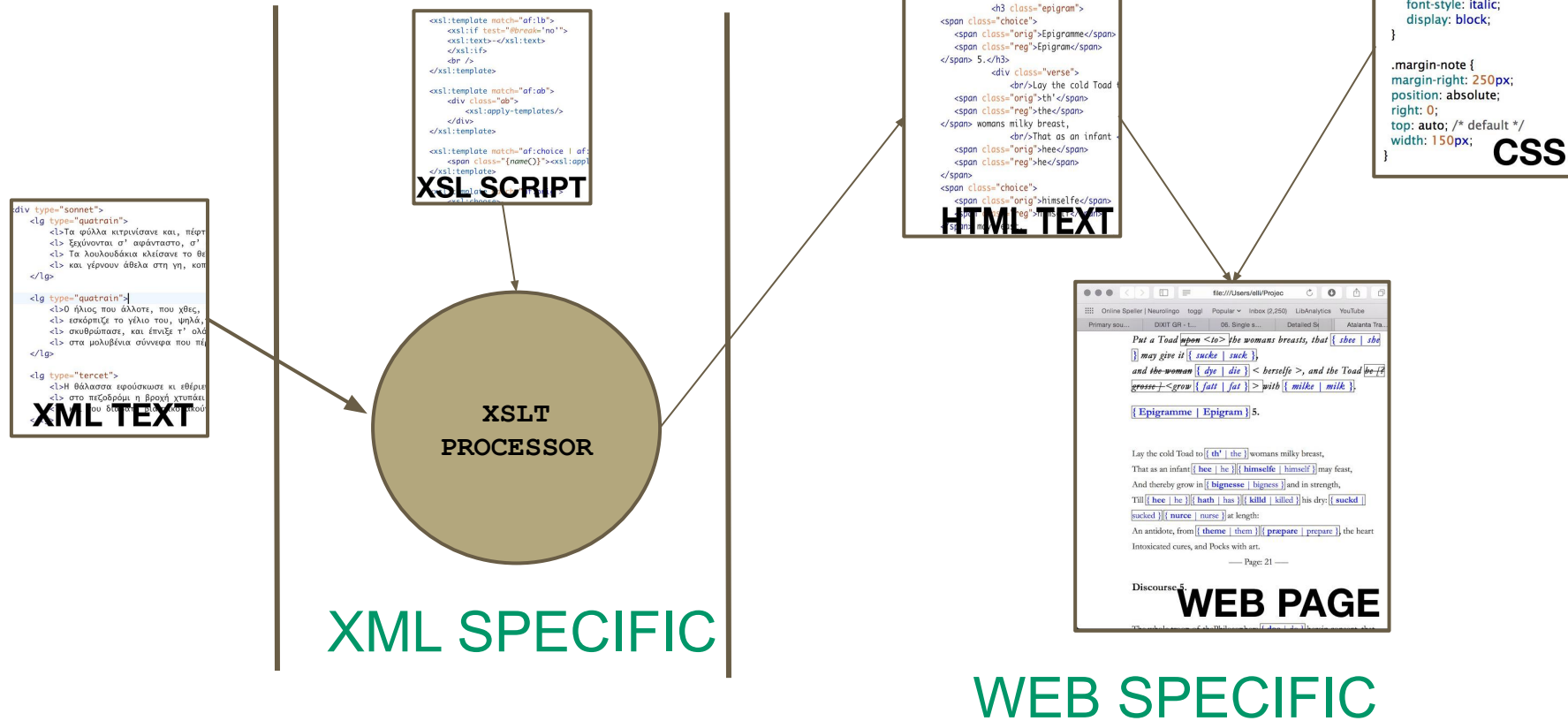
<https://www.balisage.net/Proceedings/vol5/author-pkg/Birnbaum01/BalisageVol5-Birnbaum01.html>

XML INPUT



COMMON FORMATS

XSLT Example Transformation Process



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

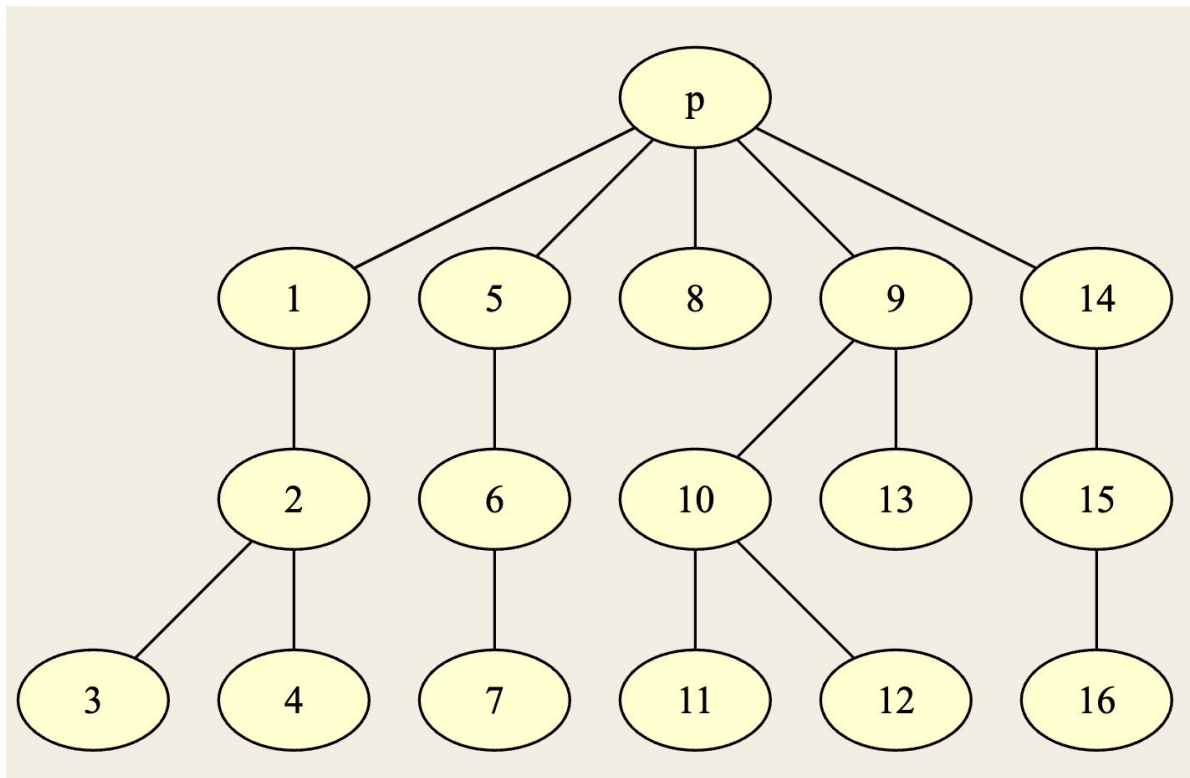
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="letter.rng"?>
<letter type="friendly">
  <salute>Dear Parents</salute>
  <para n="1">I am having a good time a
camp</para>
  <para n="2">Please send cookies</para>
  <closer>hugs and kisses
    <signed>Scooter</signed>
  </closer>
</letter>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

HTML

```
<html>
  <head>
    <title>The Friendly Letter</title>
  </head>
  <body>
    <p class="salute">Dear Parents</p>
    <p>[1] I am having a good time a camp</p>
    <p>[2] Please send cookies</p>
    <p class="closer">hugs and kisses, Scooter</p>
  </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">  
  <p class="salute"><xsl:value-of select="." />,</p>  
</xsl:template>
```

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

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  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 -->
    <p class="salute"><xsl:value-of select="."/>,</p>
  </xsl:template>

  <xsl:template match="para"> <!-- found para, output content of template -->
    <p>[<xsl:value-of select="@n"/>] <xsl:value-of select="."/></p>
  </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.

*Epigramme***Epigram 6.**

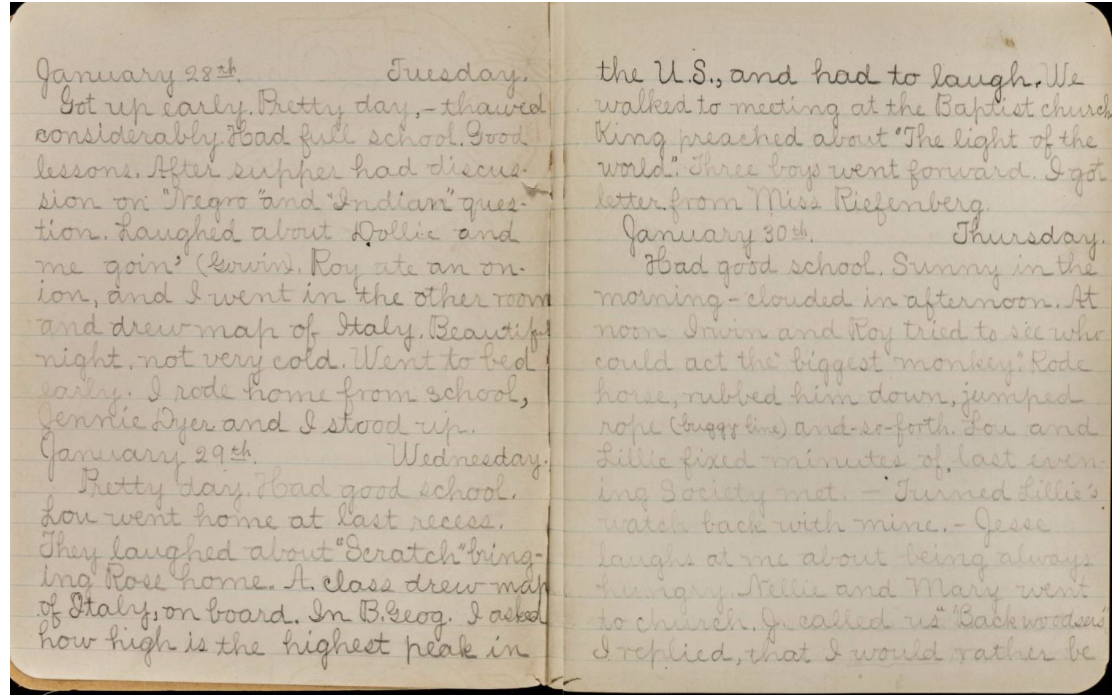
Rusticks**Rustics** their seed to th'**the** fertile earth committ**commit**,
When with their harrows they have made it fitt**fit**:
The Sophi thus their golden seed do**do** sow**sow**
In foliated earth as white as Snow:
This method well observe, and you'**you'll** behold,
As in a glasse**glass**, by 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.



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#creating-and-testing-your-xsl-file>
2. Holmes, Martin. “XSLT” from the DarjahTeach “TEI and XSLT” module.
<https://teach.dariah.eu/mod/page/view.php?id=454> Start after 5:44 where the speaker starts to talk about XSLT.
3. Holmes. “An Introduction to XSLT for Digital Humanists”
<https://web.uvic.ca/~mholmes/dhoxss2013/>
4. W3Schools. “XSLT Tutorial” Best used as a reference for syntax and functions.
https://www.w3schools.com/xml/xml_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

