Single Source

—

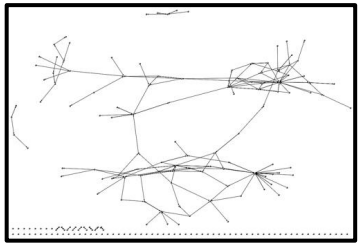
Multiple Outputs

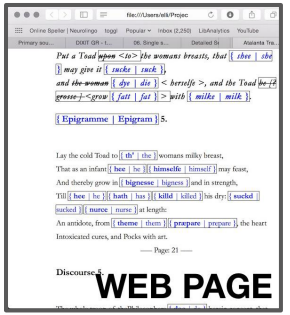
Transforming your TEI Texts

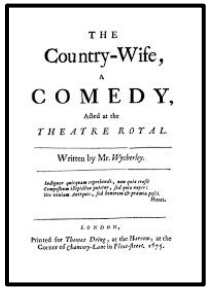
DiXiT workshop, 24-28 April 2017

What To Do With a TEI Text?

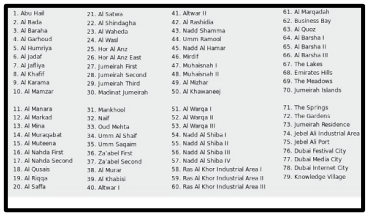
- Proofread it
- Print it out
- Index it
- Analyze it
- Make a corpus






THE
Country-Wife,
A
COMEDY,
Acted at the
THEATRE ROYAL.
Written by Mr. Wycherley.
LONDON,
Printed for Thomas Doring, at the Harrow, at the Corner of Chancery-Lane in Fleet-Street, 1675.

This is accomplished by using a program to *transform (μετασχηματίζει)* the text or to *extract information* from the text



Examples

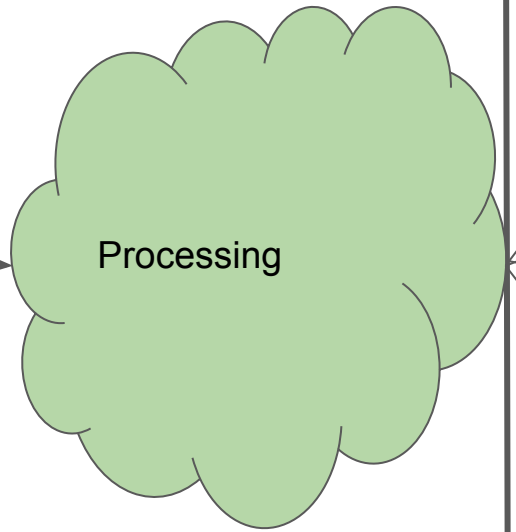
- Shelley Godwin
  http://shelleygodwinarchive.org/sc/oxford/frankenstein/volume/i/#/p1
- Inscriptions of Aphrodisias showing edition, diplomatic transcription, indices:
  http://insaph.kcl.ac.uk/iaph2007/iAph020113.html
- Coptic Scriptorium normalized, diplomatic, analyzed
  http://data.copticscriptorium.org/filter/corpus=Besa.Letters
- Others?

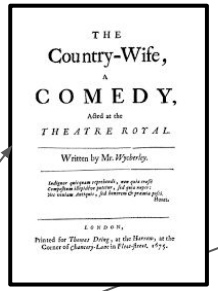
TEI Transformation (Μετασχηματισμός)

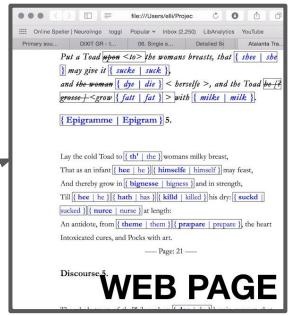
TEI and XML texts are transformed using:

- Specialized XML aware programming languages such as **XSL** and **XQuery**
- More general programming languages that may have **XML-aware libraries**
- Simple substitutions using familiar web languages such as **CSS** and **javascript**

The Transformation Process





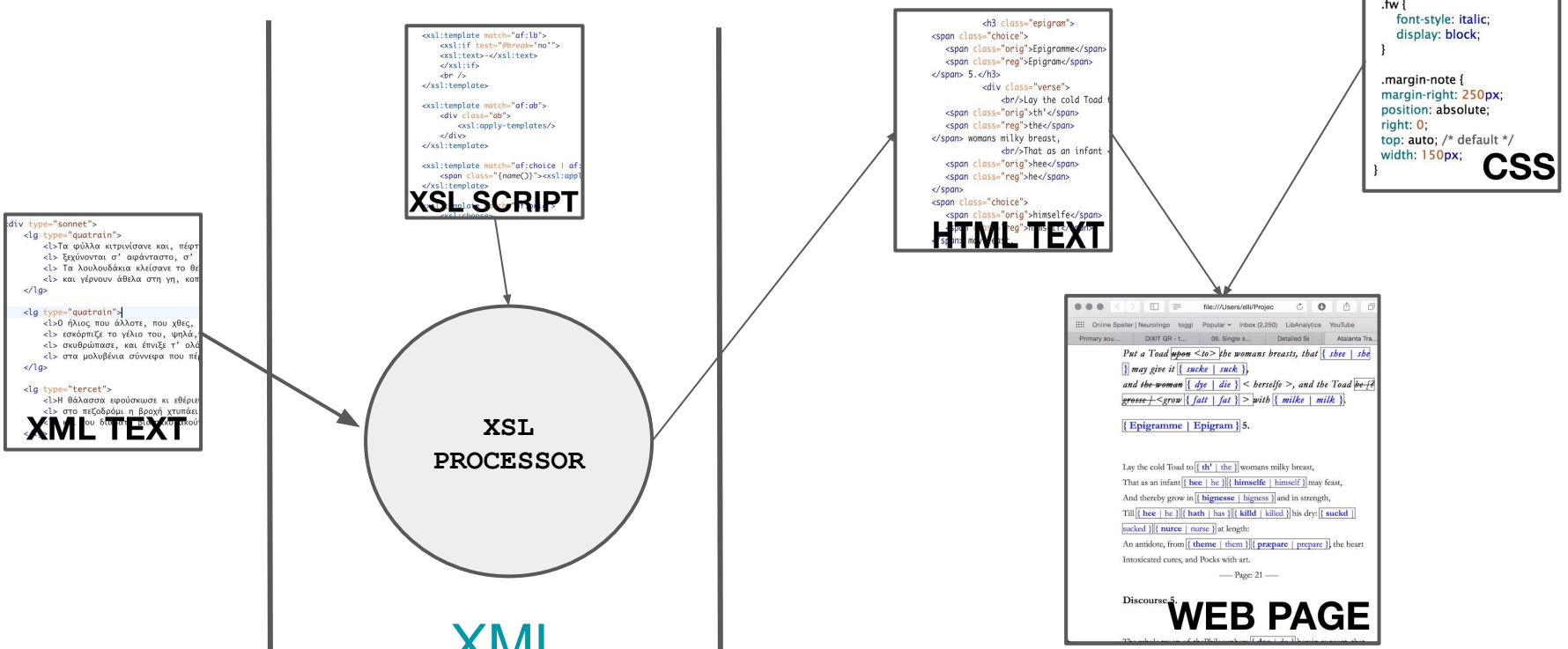


XML INPUT

PROCESSING

COMMON FORMATS

XSLT Example Transformation Process



Terms

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

XQuery: A transformation and query 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

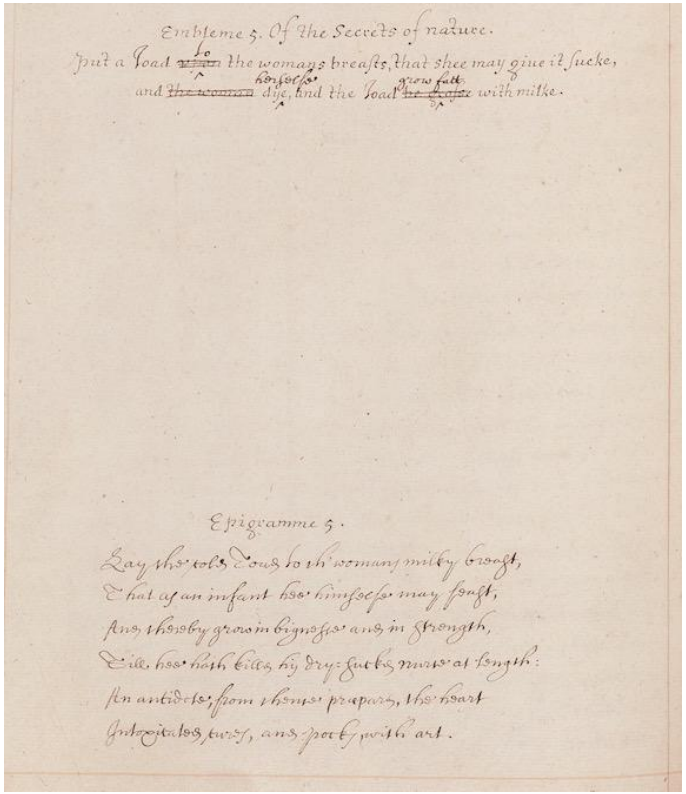
Single texts: Proofreading and Display

- Proofreading native XML is difficult and error-prone.
- It is possible to write scripts and embed them in the editing workflow to enable proofreading.
- It is possible to write scripts to transform XML files for display.

Example: Atalanta proofreading script (easier proofreading of markup)

Example: Τσίρκας letter (different types of display)

Emblem 5, Maier, Michael. *Atalanta Fugiens* (English Translation). Beineke Mellon MS 48.

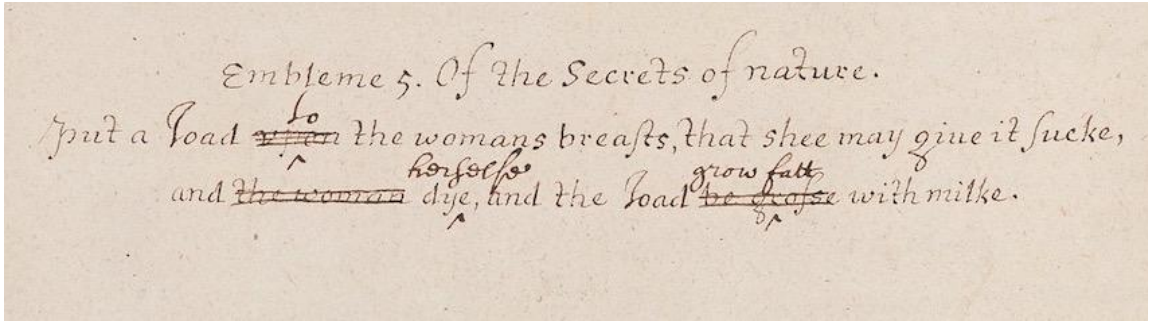


Emblem 5. Of the Secrets of nature.

Put a Toad ~~upon~~ to the womans breasts, that shee may giue it sucke, and ~~the woman~~ herselfe dye, and the Toad ~~be grosse~~ grow fatt with milke.

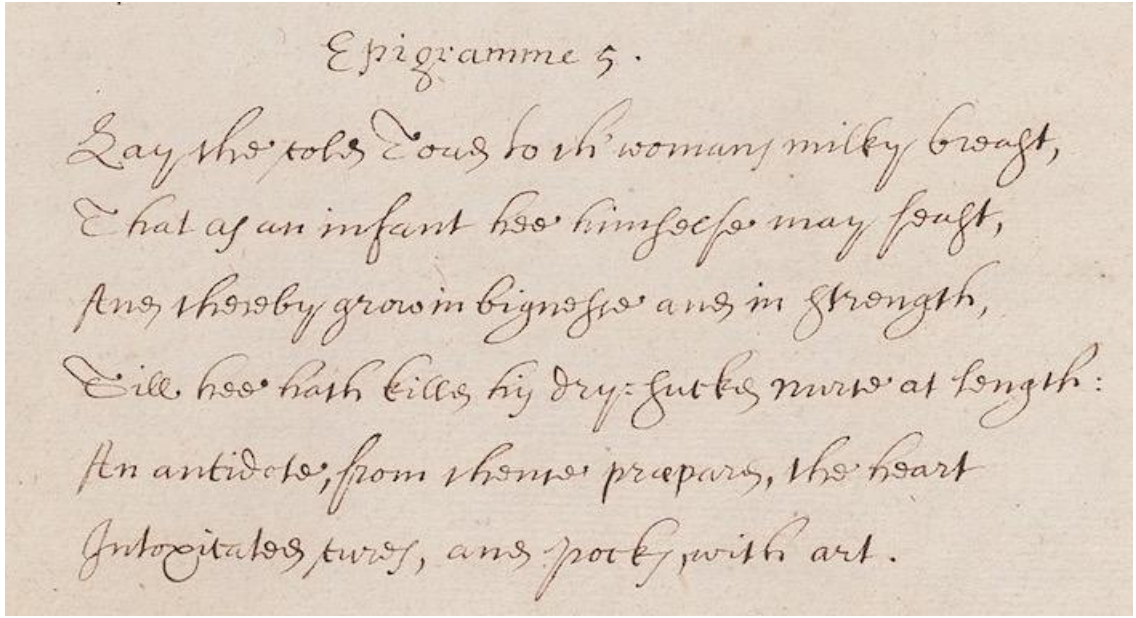
Epigramme 5.

Lay the cold Toad to th' womans milky breast,
That as an infant hee himselfe may feast,
And thereby grow in bignesse and in strength,
Till hee hath kill'd his dry-sucke nurse at length:
An antidote, from whence prepar'd, the heart
Intoxicates, turns, and sports with art.



Emblem 5. Of the Secrets of nature.

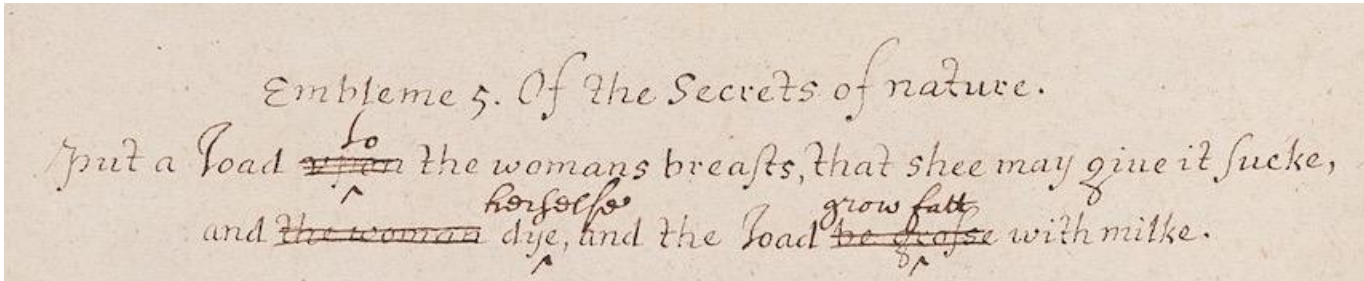
Put a Toad ~~upon~~ to the womans breasts, that shee may giue it sucke, and ~~the woman~~ herselfe dye, and the Toad ~~be grosse~~ grow fatt with milke.



Epigramme 5.

Lay the cold Toad to th' womans milky breast,
That as an infant hee himselfe may feast,
And thereby grow in bignesse and in strength,
Till hee hath kill'd his dry-sucke nurse at length:
An antidote, from whence prepar'd, the heart
Intoxicates, turns, and sports with art.

```
<div type="title" xml:lang="en">
  <head>
    <lb/><choice><orig>Embleme</orig><reg>Emblem</reg></choice> 5. Of the
Secrets of Nature.
    <lb/>Put a Toad <subst><del>upon</del><add>to</add></subst> the womans
breasts, that <choice><orig>shee</orig><reg>she</reg></choice> may give it
<choice><orig>sucke</orig><reg>suck</reg></choice>,
    <lb/>and <del>the woman</del><choice><orig>dye</orig><reg>die</reg></choice>
<add><orig>herselfe</orig></add>, and the Toad <subst><del>be <unclear
reason="illegible"><choice><orig>grosse</orig><reg>gross</reg></choice></unclear><
/del><add>grow <choice><orig>fatt</orig><reg>fat</reg></choice></add></subst>with
<choice><orig>milke</orig><reg>milk</reg></choice>.
  </head>
</div>
```



Αγαπητέ μου Εφαλή 21/6/55

Ἂν ἔχω τὰ «Γραμματα» δὲ ἤσυμα
σὺ μιᾶ συγγαρή κὶ ἔφγεν τοῦ Κυμηρου.

Ὑπάρχαν πληρονομοί;

Ἔγνεται αἰτοῦ σιβιλίσσης γύρω ἀπὸ
τὸ ποσταλιμό, ὅπως ἐδὼ στὸ "Νοτρᾶ";

Γιά τὴν ἔνιδεσι συγγρῆς ἀπὸ τὸ ἔργο
σου: Εἰς μεγάλη μου τοῦ δὲ τὴ σπουδέσει
μιτορᾶ νὰ μεγάλον ἄνεται ἢ ἡ δεσὶ σου
ἐμπόδεξεν;

Εὐχαριστῶ γιά τὰ βιβλία κηρῶ.
Θὰ σοῦ στείλω κι ἐγὼ βιβλία, χρήσιμα
ἴσως.

Ο κ. Ἀλιθέρσης εἶχε γράψει ἱστορία
τῆς λογοτεχνίας. Ὑπάρχει ἀντίτυπο;

Θὰ σοῦ ξαναγράψω περισσότερα
ὁ βερνὴς εἶναι εἰδοποιημένος.

[illegible]

Αγαπητέ μου Εφαλή 21/6/55

Ἂν ἔχω τὰ «Γραμματα» δὲ ἤσυμα
σὺ μιᾶ συγγαρή κὶ ἔφγεν τοῦ Κυμηρου.

Ὑπάρχαν πληρονομοί;

CSS and HTML

TEI

```
<opener>
   <dateline><date when="1955-06-21">21/6/55</date></dateline>
   <salute>Αγαπητέ μου <persName ref="#ST">Στρατή</persName></salute>
</opener>
```

HTML (generated by the XSL transformation)

```
<div class="opener">
    <p class="date">21/6/55</p>
    <p class="salute">Αγαπητέ μου <span class="persName">Στρατή</span> </p>
 </div>
```

CSS and HTML

HTML

```
<div class="opener">
    <p class="date">21/6/55</p>
    <p class="salute">Αγαπητέ μου <span
class="persName">Στρατή</span> </p>
 </div>
```

```
<link rel="stylesheet"
        type="text/css"
        href="tsirkasProof.css"
        media="screen"/>
```

CSS

```
p.date  {
    text-indent: 75%;
    padding-bottom: 10px;
}
```

CSS and HTML

HTML

```
<p>Αν έχω τα Γράμματα θα ετοιμά | σω
μια συλλογή των έργων του <span
class="persName">Σκληρού</span>.
Υπάρχουν κληρονόμοι;</p>
```

```
<link rel="stylesheet"
      type="text/css"
      href="tsirkasProof.css"
      media="screen"/>
```

CSS

```
span.persName {
  text-decoration:
underline;
}
```

Tools for Single Texts

Mostly for display or interaction with a single text

- Boilerplate http://dcl.ils.indiana.edu/teibp/
- Petrarch http://dcl.slis.indiana.edu/petrarchive/

- CETEIcean: http://philomousos.com/CETEIcean/index.html
- Apparatus Example: http://hcayless.github.io/appcrit/
- IIP example: https://brown-university-library.github.io/iip-texts

These are a few examples of tools - there are others.

Working with Single Texts

- Not difficult to implement without technical help (or with the help of a student)
- Can be done using Oxygen or similar tools

- Do not offer much programmatic capability
- Cannot take advantage of large collections
- Requires more technical skills and/or infrastructure

Collections and Corpora

Many web projects are examples of large text corpora. We have come to expect:

- Search over multiple texts
- Search over multiple texts that is structure aware
- Search that relies on controlled vocabularies
- Indexing of metadata (facets)
- Overviews and multiple forms of organization for browsing
- Analysis/visualization of data and metadata

Tools for Collections and Corpora

- eXist - TEI-Publisher
  http://teipublisher.com/exist/apps/tei-publisher/index.html
- Tapas “TAPAS is the TEI Archiving, Publishing, and Access Service hosted by Northeastern University Library's Digital Scholarship Group.”
  http://tapasproject.org
- Customized solutions
  - Drupal
  - Django
  - others

Other TEI Transformations

There are other ways to transform into or out of XML (especially TEI)

- Zotero (bibliographic management software) exports bibliographic entries as TEI fragments
- Oxygen Author mode formatted view that makes it easier to encode texts
- There are people who use their TEI files to create paper publications…

Exercise