

Publishing Workflow

Tony Graham
Mentea
13 Kelly's Bay Beach
Skerries, Co Dublin
Ireland
info@mentea.net
<http://www.mentea.net>

Version 1.0 – 17 September 2012
Licensed under a Creative Commons Attribution-Noncommercial-Share
Alike 3.0 Unported License



Publishing

Workflow

What is a Workflow?	5
XML and Workflows	5
XML as Source	6
XML as Hub	9
XML as Target	26
One Person's Source is Another's Hub	27
Questions?	28
References	29
Exhibits	31

Learning Objectives

1

- How formats can be better or worse than others
- See ways of doing things

Course Outline

2

- Review workflow concepts
- Advantages and disadvantages of XML
- Tools
- Where to use XML and XSLT
- What makes a good XML format?

This Course is NOT

3

- Guide to available workflow packages
- Prescription for how to organise your own workflow
- Description of XML languages for describing workflows

What is a Workflow?

4

- Wiktionary:
 - A process and/or procedure in which tasks are completed. It may be defined with a flowchart to define actors, actions, results, decisions, and action paths.
- Wikipedia:
 - It is a depiction of a sequence of operations, declared as work of a person, a group of persons,[1] an organization of staff, or one or more simple or complex mechanisms. Workflow may be seen as any abstraction of real work. For control purposes, workflow may be a view on real work under a chosen aspect,[2] thus serving as a virtual representation of actual work. The flow being described may refer to a document or product that is being transferred from one step to another.

What Is A Workflow To You?

5

XML and Workflows

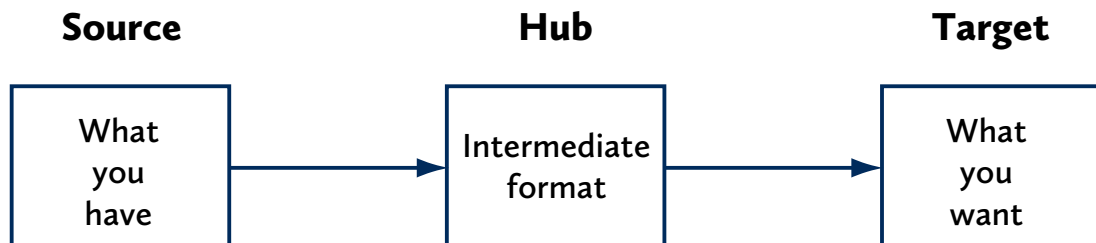
6

- XML is a good use for a workflow
- XML almost always needs processing for:
 - Presentation
 - Alternative output formats
 - Archive format
 - Including changes made during page layout

Workflow Reduced

7

- Workflow is a series of steps
- Every step transforms its source to its target, possibly through a hub format:
 - Source – Original and/or authored format(s)
 - Hub – Intermediate format
 - Target – Result format(s)



XML as Source

8

- Advantages
 - Enforce structure
- Disadvantages
 - May need more than structure
 - Business rules
 - House style
 - Other constraints
 - May need to hide XML from authors

XMLish Source Formats

9

- XML
 - Standard schema
 - Standard schema with own extensions
 - Subset or profile of standard schema
 - Custom schema
- Microsoft Office & Open Office documents
- XHTML(5)

XML as Source: Autodata

10

- Europe's leading provider of technical information to automotive professionals
- Established 1972
- 80 manufacturers
- Over 16,000 models
- 19 languages
- Originally printed books, then CDs, now online
- 2005 – 50% electronic delivery
- 2012 – 95% electronic delivery
- Also supplies XML directly to VARs
- Monthly updates to online and partners



Markup Evolution

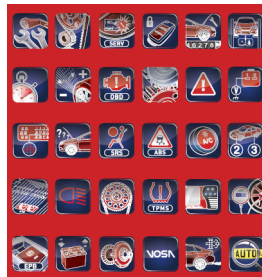
11

- 1990s – SGML-like markup
- Early 2000s – DocBook plus vehicle-specific extensions
- 2013 – Autodata-specific XML vocabulary

Multiple Subject Areas

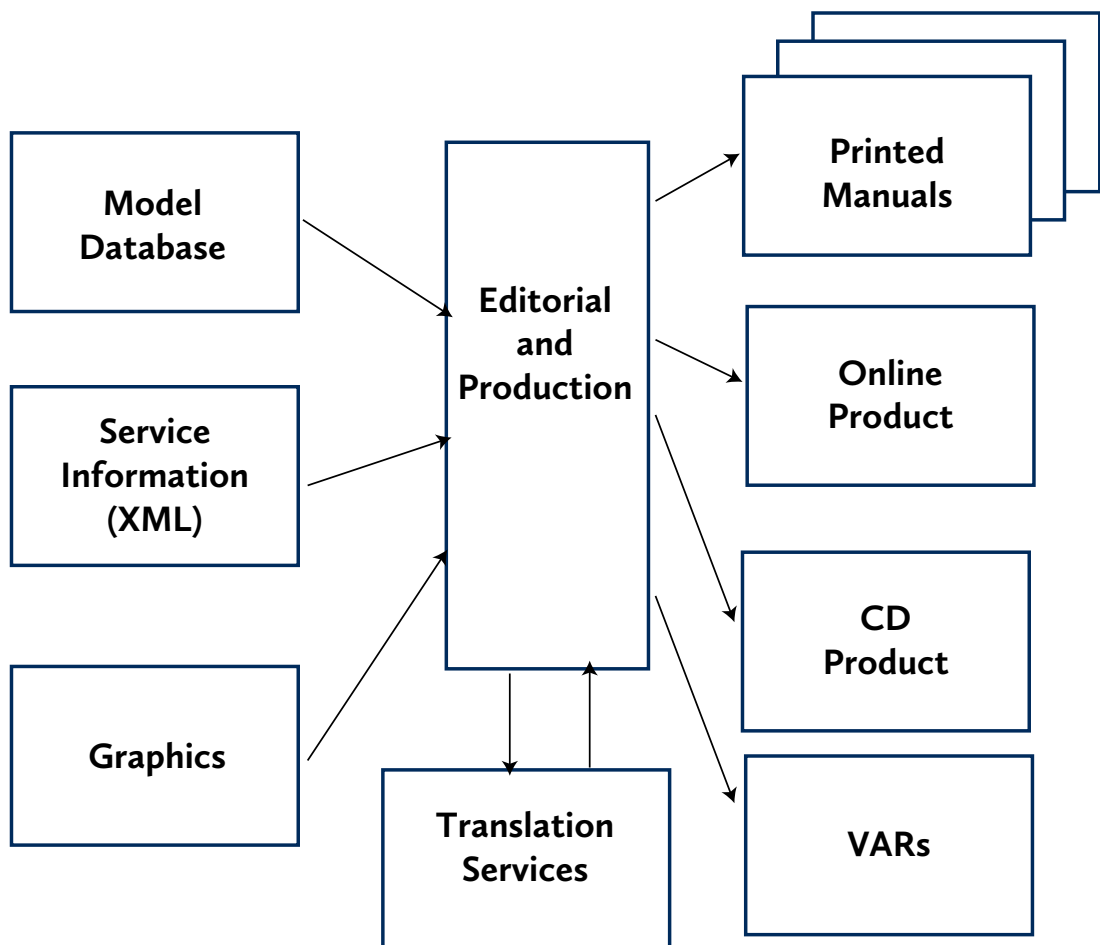
12

- More being added over time
- ABS, engine management systems, diagnostic trouble codes, etc., didn't exist 40 years ago!



Simplified Workflow

13



XML Origination

14

- Narrative, mostly procedural service information
- Authored by ex-Service Technicians
- Many initially unfamiliar with word processors
- *None* initially familiar with XML
 - Except as “XML hot-fixes” to download to car engine management systems
- XML also for notes associated with technical, quantitative data from database
- Editorial department reviews documents and handles translations
- Consistent terminology maintains house style and minimises translation costs

XML Tools

15

- Origination authors – Quark *XML Author*
- Origination team leads – *XML Author*, Oxygen
- Editorial – Oxygen, Trados translation memory, in-house tools
- Book production – ArborText/3B2
- CD production – Proprietary tool
- Online production – .Net, XSLT 1.0, XSLT 2.0

Current XML Markup

16

- DocBook
- Autodata extensions:
 - Vehicle- and Autodata-specific metadata
 - Service-specific inlines, e.g:
 - Special tool identifiers
 - Diagnostic trouble codes
 - Attributes indicating right-hand or left-hand drive, etc.
 - DocBook `<alt>` allowed in more places to handle RH/LH drive, etc., differences

Limitations of Current Markup

17

- Originally one subject per printed book
- Electronic products allow more correlations between subjects
- DocBook markup still largely visual:
 - Important information in a subject always marked up as same kind of `<table>` or `<segmentedlist>`
- DocBook still very loose:
 - *XML Author* configuration more restrictive than DocBook
 - Some documents converted by-hand from pre-XML markup are valid DocBook but can't be opened in *XML Author*
- Want more service-specific inline markup
- Building products that pull in data from multiple subjects

AutodataML – The Next Generation

18

Autodata-specific schema (and updated *XML Author*) will support:

- More meaningful markup
- Consistent authoring
- Less boilerplate text, e.g., no wiring colour key table
- Markup for voltages, resistances, etc.
- Round-tripping back to (more consistent) DocBook until processes and VARs switch over to new schema

XML as Hub

19

- Advantages
 - Technology neutral
 - Standards-based tools (often free) for transforming into and out of XML
- Disadvantages
 - Someone needs to look at the pointy brackets

Converting Non-XML Sources to XML

20

- Export as XML
- Export as CSV
 - Programmatically using format-specific library
 - General-purpose programming language
 - Language with good regular expression support, e.g., Perl
 - XSLT

Non-XML To XML Using XSLT

21

- `unparsed-text()`
- Saxon collection resolver

```
file:///a/b/c/d?unparsed=yes;select=*.txt
```

- Tokenize line-based format into `<line>` elements

```
<xsl:variable
  name="lines"
  select="tokenize(unparsed-text(concat('file://', $doc-
converted),
                                $encoding),
                                '(&#xD;&#xA;|&#xD;|&#xA;)') [normalize-
space() ne '']"
  as="xs:string"/>
```

Impediments to Moving to XML

22

- Legacy data may not fit a rigid schema
 - Create two-tier XML
- Cost of conversion
 - Will only increase the longer you wait!
- Investment in current systems
 - “Back-convert” to current format until replaced?
- Existing channels, partner agreements, etc.
- Retraining costs

What Make's a Good XML Format?

23

- Describes your data
- Doesn't get in the way
- Conforms to (widely implemented) standards

Describes Your Data

24

- Markup structure matches document structure
- Low-level markup matches your needs:
 - Parts of speech?
 - Part numbers?
 - Cross-references?
 - Bibliographies, and with sufficient detail?

Doesn't Get In The Way

25

- Do you need to provide six levels of markup for what, to you, is a simple structure?
- Are you having to (over)use `role`, `remap`, `class` or other attributes because there aren't elements for what you want?
- Are you adding markup for what things look like because you don't have elements describing what they really are?

Conforms To Standards

26

- Conformance to standards can:
 - Increase portability of data
 - Avoid vendor lock-in
 - Allow more software choices
 - Make it easier for new people to understand your data
- Picking a standard can be easy or hard
 - Sometimes it's obvious
 - This year's obvious may make obsolete what was previously obvious
 - Sometimes an emerging standard may not be widely implemented after all

Now You Have XML, What Can You Do With It?

27

- Edit
- Translate text and leave structure intact
- Store in and serve from a XML Database
- Transform into:
 - Other XML
 - HTML, XHTML, or HTML5
 - XSL-FO for formatting
 - EPUB 2.0 or EPUB 3.0
- Use other, proprietary tools for other outputs
- Exchange with customers, partners, or service providers
- New uses you haven't thought of yet

Delivery Strategies

28

- Static
- On-the-fly
- On-the-fly plus cache
 - Personalised

Tools

29

- Defining your requirements and finding the tool to fit the requirements is harder than just finding an XML-enabled tool
- Tool categories include:
 - Authoring
 - XML processing
 - Presentation
 - CMS
 - Database

Authoring

30

- All, some, or no visible markup?
- Word processor or other editor plug-in or standalone?
- XML document editor or also XSLT, XQuery, WSDL, etc., editor?
- Integration with backend database or CMS or standalone?
- License model and cost?

XML Processing

31

- W3C standards rule!
 - XSLT
 - XQuery
 - XProc
- Is there a language/tool that fits your environment best
 - Programming language
 - Operating system
- What level of standard conformance do you require?
- Will you need a vendor's non-standard extensions?
- Do you need schema-awareness?
- License model and cost?
- Performance requirements?

Presentation

32

- How will documents be displayed to users?
 - (X)HTML(5) in browser?
 - Print
 - EPUB?
- Standards compliance?
- Throughput or other performance requirement?
- Ability to touch-up formatted output?
- Multiple languages/scripts?
- License model and cost?

Content Management System (CMS)

33

- Very broadly applied term
- What do you want managed?
 - XML documents?
 - Word processor and other office files?
 - Graphics and other digital assets?
 - Metadata about files but not files themselves?
- How much history is kept?
- Volume, response time, and other performance constraints?
- License model and cost?

Database

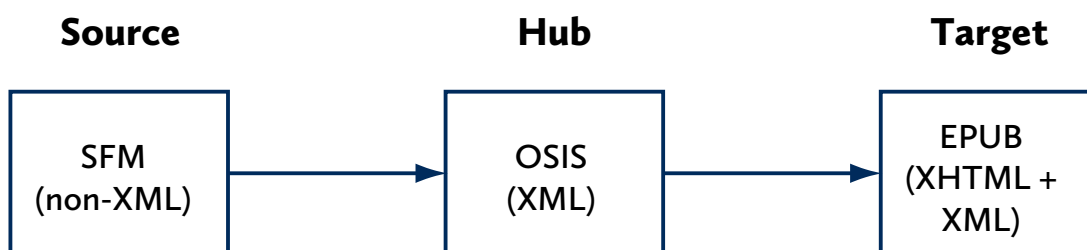
34

- Do you need a dedicated database or will the filesystem do?
- Database for documents or just metadata
- XML? Relational? NoSQL? or what?
- Interfaces to existing systems/databases?
- Integration with editing, CMS, or other functions?
- Volume, response time, and other performance constraints?
- License model and cost?

XML As Hub: French Bible EPUBs

35

- Creating EPUBs of Bibles for specialist publisher in France
- Input: SFM (non-XML) from Bible Society
 - Troff-like markup used for printed Bibles
- Hub: OSIS, TEI-influenced XML standard for representing scriptures
- Target: EPUB 2.0 – Zip archive containing XHTML and specific XML
- Implemented:
 - Pipeline of XSLT transforms to convert different SFM flavours to OSIS
 - XSLTs to convert OSIS into XHTML and ancillary files
 - Adapted stylesheets to process OSIS received from elsewhere



In the beginning, there was ... SFM

36

```
\c 1
\ms AU COMMENCEMENT
\mr 1--11
\s Dieu crée l'univers et l'humanité
\p
\v 1 Au commencement Dieu créa le ciel et la terre*fa*.
\fm a
\fr 1.1
\fiAu commencement...:|x traduction la plus fréquente de
ce verset. Elle est imitée de l'ancienne version grecque, à laquelle
se réfère très probablement l'évangile de Jean (1.1). Mais le
texte hébreu serait mieux rendu par |iQuand Dieu
commença de créer le ciel et la terre... Dieu dit|x.
\p
\v 2 La terre était sans forme et vide, et l'obscurité
couvrait l'océan primitif. Le souffle de Dieu se déplaçait à
la surface de l'eau*fb*.
\fm b
\fr 1.2
\fiLe v.$2 constitue en hébreu une sorte de parenthèse,
le v.$1 ayant sa suite au v.$3. -- |isans forme et vide:|x
l'expression hébraïque correspondante a donné en français
|itohu-bohu|x. Le jeu de mots pourrait être rendu par |iun
désert en désordre|x. -- |ile souffle de Dieu:|x autre
traduction |iun vent terrible|x.
```

OSIS

37

```
<div type="book">
<title short="GENÈSE" type="main">Genèse</title>
<chapter osisID="Gen.1" sID="Gen.1" n="1"/>
<div type="majorSection"><title>AU COMMENCEMENT</title>
<title type="scope"><reference>1-11</reference></title>
<title type="x-section">Dieu crée l'univers et l'humanité</title>
<p><verse osisID="Gen.1.1" sID="Gen.1.1" n="1"/>Au
commencement Dieu créa le ciel et la terre<note n="a"
type="crossReference"> <reference type="source"
osisRef="Gen.1.1">1.1</reference> <hi type="italic">Au
commencement... :</hi> traduction la plus fréquente de
ce verset. Elle est imitée de l'ancienne version
grecque, à laquelle se réfère très probablement
l'évangile de Jean (1.1). Mais le texte hébreu serait
mieux rendu par <hi type="italic">Quand Dieu commença de
créer le ciel et la terre... Dieu
dit</hi>.</note><verse eID="Gen.1.1"/></p>
<p><verse osisID="Gen.1.2" sID="Gen.1.2" n="2"/>La terre
était sans forme et vide, et l'obscurité couvrait
l'océan primitif. Le souffle de Dieu se déplaçait à la
surface de l'eau<note n="b"
type="crossReference"><reference type="source"
osisRef="Gen.1.2">1.2</reference>Le v. 2 constitue en
hébreu une sorte de parenthèse, le v. 1 ayant sa suite
au v. 3. - <hi type="italic">sans forme et vide :</hi>
l'expression hébraïque correspondante a donné en
français <hi type="italic">tohu-bohu</hi>. Le jeu de
mots pourrait être rendu par <hi type="italic">un désert
en désordre</hi>. - <hi type="italic">Le souffle de
Dieu :</hi> autre traduction <hi type="italic">un vent
terrible</hi>.</note>. <verse eID="Gen.1.2"/>
```

EPUB

38

- XHTML

```
<body class="chapter" id="top"><div class="book">
<h2 class="majorSection nolinegap">AU COMMENCEMENT</h2>
<h2 class="scope wholelinegap"><span class="reference">1-11</span></h2>
<h2 class="x-section halflinelinegap">Dieu crée l'univers et l'humanité</h2>
<p class="p"><a class="verse" id="vGen.1.1"/><span class="displayReference">
<span class="chapterBookName"><a class="chapter-to-index-link"
href="Gen.xml#cGen.1">GEN</a></span><span class="firstChapterNumber">1</span>
</span>Au commencement Dieu créa le ciel et la terre
```

- NCX

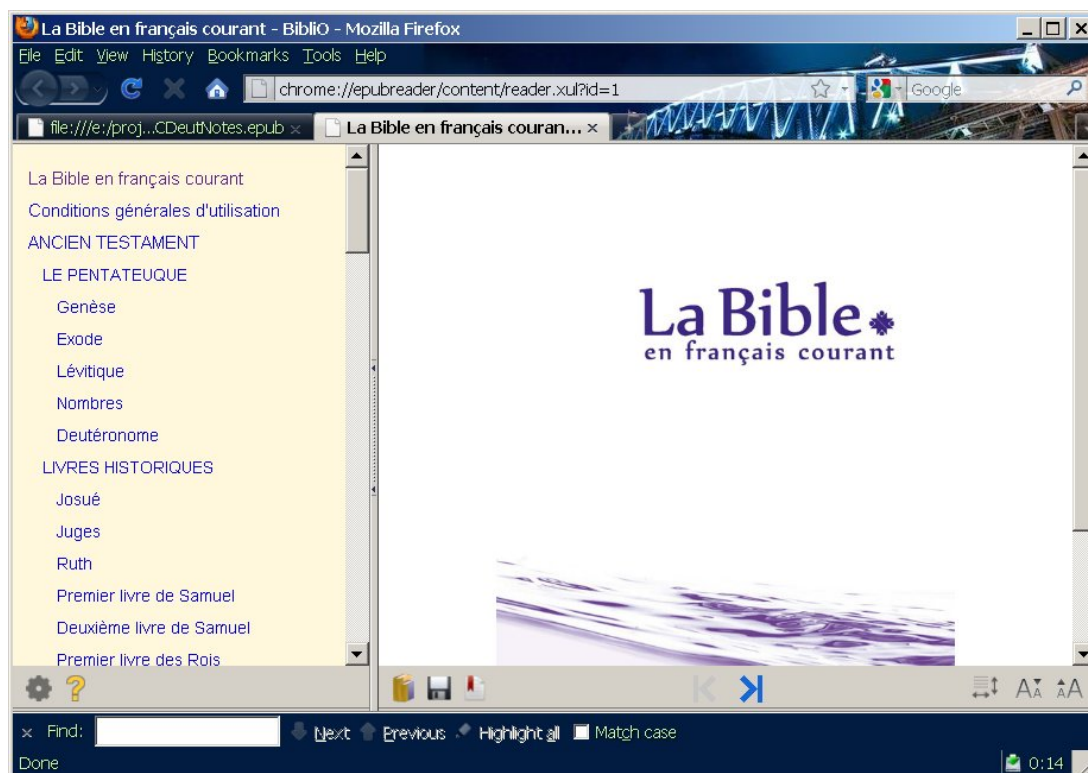
```
<navPoint class="category" id="navpoint-3" playOrder="3">
<navLabel><text>ANCIEN TESTAMENT</text></navLabel>
<content src="Gen.xml"/><navPoint class="category" id="navpoint-3" playOrder="3">
<navLabel><text>LE PENTATEUQUE</text></navLabel>
<content src="Gen.xml"/><navPoint class="book" id="navpoint-3" playOrder="3">
<navLabel><text>Genèse</text></navLabel>
<content src="Gen.xml"/></navPoint>
```

- OPF

```
<spine toc="ncx"><itemref idref="titlepage" linear="yes"/>
<itemref idref="copyright" linear="yes"/>
<itemref idref="toc" linear="yes"/>
<itemref idref="bk-Gen" linear="yes"/>
<itemref idref="intro-Gen1" linear="yes"/>
<itemref idref="ch-Gen-1" linear="yes"/>
```

EPUB

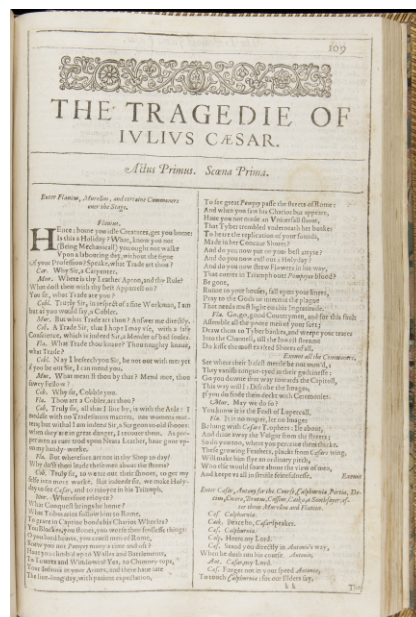
39



Extended Example: Julius Caesar

40

- “The Tragedy of Julius Caesar” by William Shakespeare
- Written 1599?
- “First Folio” printed 1623



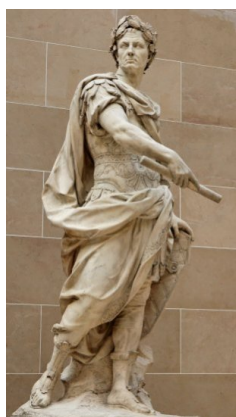
The TEI Does Julius Caesar

41

- <http://ota.ox.ac.uk/id/3014>
- “First Folio” text
- Records page breaks, line breaks, etc.
- Doesn't record initial capitals

Caesar's Anatomy

42



Head and Body

43

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>THE TRAGEDIE OF IVLIVS CAESAR.</title>
        <author>Shakespeare, William, 1564-1616</author>
      </titleStmt>
      ...
    </fileDesc>
    ...
  </teiHeader>
  <text>
    <body>
      <pb n="kk1"/>
      <head>THE TRAGEDIE OF<lb/>IVLIVS CAESAR.</head>
      <milestone unit="compo" n="B"/>
      ...
    </body>
  </text>
</TEI>
```

Inside Caesar's Head

44

- fileDesc – Full bibliographic description
 - titleStmt – Information about the title
 - title
 - author
- publicationStmt – Information concerning publication
 - distributor – Person or other agency responsible
 - idno+ – Any form of identifier
 - availability – Information about the availability
- sourceDesc – Source from which text was derived
 - bibl+ – loosely-structured bibliographic citation
 - biblFull – Fully-structured bibliographic citation
- encodingDesc – Encoding description
- profileDesc – Non-bibliographic aspects of a text
- revisionDesc – Revision history

titleStmt

45

What and who

```
<titleStmt>
  <title>THE TRAGEDIE OF IVLIVS CAESAR.</title>
  <author>Shakespeare, William, 1564-1616</author>
  <editor role="editor">Lee, Sidney, Sir, 1859-1926</editor>
</titleStmt>
```

publicationStmt

46

```

<publicationStmt>
  <distributor>
    <name>University of Oxford Text Archive</name>
    <address>
      <addrLine>Oxford University Computing Services</addrLine>
      <addrLine>13 Banbury Road</addrLine>
      <addrLine>Oxford</addrLine>
      <addrLine>OX2 6NN</addrLine>
    </address>
    <email>ota@oucs.ox.ac.uk</email>
  </distributor>
  <idno type="ota">http://ota.ox.ac.uk/id/3014</idno>
  <idno type="isbn10">1106000137 </idno>
  <idno type="isbn13">9781106000132</idno>
  <availability status="restricted">
    <licence target="http://creativecommons.org/licenses/by-sa/3.0/">
Distributed by the University of Oxford under a Creative Commons
Attribution-ShareAlike 3.0 Unported License
    </licence>
  </availability>
</publicationStmt>

```

sourceDesc

47

```

<sourceDesc>
  <bibl>Revised version of <relatedItem type="older"
target="http://ota.ox.ac.uk/id/0119"/>
</bibl>
  <bibl>The texts were originally prepared by Trevor...</bibl>
  <biblFull>
    <titleStmt>
      <title>THE TRAGEDIE OF IVLIUS CAESAR.</title>
      <author>Shakespeare, William, 1564-1616</author>
      <editor role="editor">Lee, Sidney, Sir, 1859-1926</editor>
    </titleStmt>
    <extent>xxxv, 908 p. : facsims. ; 39 cm.</extent>
    <publicationStmt>
      ...
    <notesStmt>
      <note anchored="true">"One thousand copies of this
facsimile
have been printed"--verso of half t.p.</note>
      ...
    </notesStmt>
  </biblFull>
</sourceDesc>

```

encodingDesc

48

```

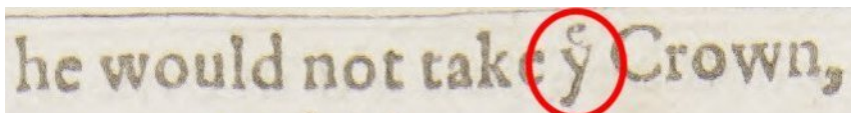
<encodingDesc>
  <charDecl>
    <glyph xml:id="ythe">
      <glyphName>Lower case y with smaller lower case e above</
glyphName>
      <desc>Abbreviation for the in Elizabethan typesetting</
desc>
      <charProp>
        <localName>entity</localName>
        <value>ythe</value>
      </charProp>
    </glyph>
  </charDecl>
  <classDecl>
    <taxonomy xml:id="OTASH">
      <bibl>University of Oxford Text Archive Subject Headings</
bibl>
    </taxonomy>
    <taxonomy xml:id="LCSH">
      <bibl>Library of Congress Subject Headings</bibl>
    </taxonomy>
  </classDecl>
</encodingDesc>

```

TEI Encoding Declaration

49

- First Folio



- TEI encoding declaration

```

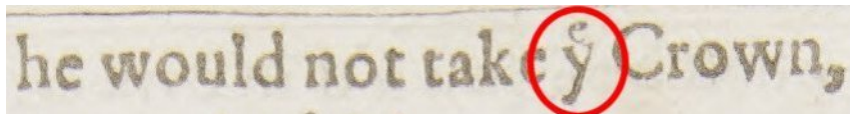
<glyph xml:id="ythe">
  <glyphName>Lower case y with smaller lower case e above</
glyphName>
  <desc>Abbreviation for the in Elizabethan typesetting</desc>
  <charProp>
    <localName>entity</localName>
    <value>ythe</value>
  </charProp>
</glyph>

```

Encoding

50

- First Folio



- TEI markup

```
he would not take <choice>
  <orig>
    <g ref="#ythe"/>
  </orig>
  <reg>the</reg>
</choice> Crown,
```

- Basic formatted result

he would not take **the** Crown,

profileDesc

51

```
<profileDesc>
  <creation>
    <date notAfter="1623"/>
  </creation>
  <langUsage>
    <language ident="eng">English</language>
  </langUsage>
  <textClass>
    <keywords scheme="#LCSH">
      <term type="genre">Plays -- England -- 16th century</term>
      <term type="genre">Plays -- England -- 17th century</term>
      ...
    </keywords>
  </textClass>
</profileDesc>
```

revisionDesc

52

```
<revisionDesc>
  <change when="2010-08-31">Header normalised</change>
</revisionDesc>
```

“Mark Antony, heere take you *Caesars* body”

53

```

<body>
  <pb n="kk1"/>
  <head>THE TRAGEDIE OF<lb/>IVLIIVS CAESAR.</head>
  <milestone unit="compo" n="B"/>
  <lb n="1"/>
  <div>
    <head rend="italic">Actus Primus. Scoena Prima.</head>
    <lb n="2"/>
    <stage rend="italic">Enter Flavius, Murellus, and certaine
Commoners<lb n="3"/>ouer the Stage.</stage>
    <lb n="4"/>
    <sp>
      <speaker rend="italic">Flavius.</speaker>
      <ab>
        <lb n="5" rend="rj"/>Hence: home you idle Creatures,
get you home:<lb n="6"/>Is this a Holiday? What, know you not<lb
n="7"/>
(Being Mechanicall) you ought not walke<lb n="8"/>Vpon a
labouring day, without the signe<lb n="9"/>
Of your Profession? Speake, what Trade art thou?<lb n="10"/>

```

TEI Elements

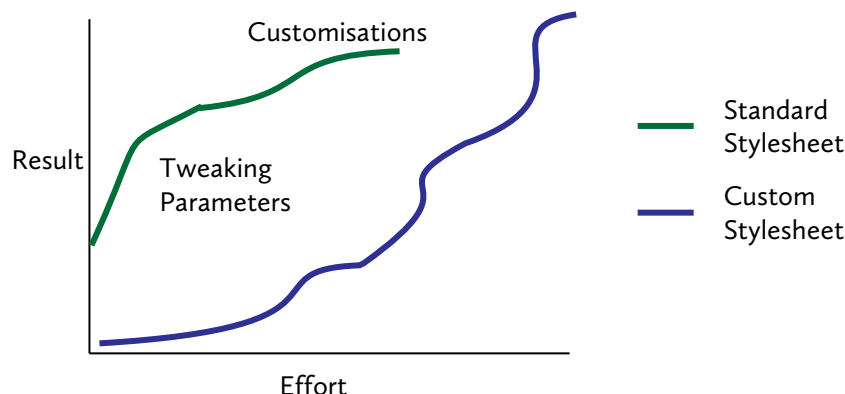
54

- pb – Page break
- head – Heading
- milestone – Marks a boundary point separating any kind of section
- lb – Marks the start of a new (typographic) line
- stage – Any kind of stage direction within a dramatic text or fragment
- sp – An individual speech in a performance text
- speaker – The name of one or more speakers in a dramatic text
- ab – Anonymous container for phrase or inter-level elements

Formatted Output: Standard or Custom Stylesheets?

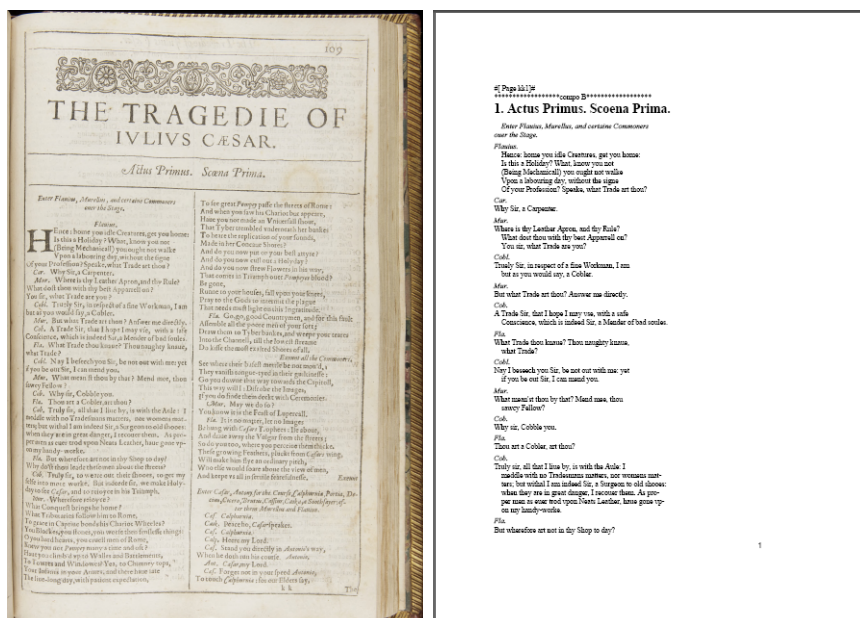
55

- Widely used XML schema often have accompanying standard stylesheets
 - Highly parameterised for tweaking font, page size, etc.
 - Should handle all elements in schema
 - Defined customisation method for overriding default templates and handling extension elements and attributes
- Custom stylesheet:
 - Full control
 - More effort, especially at beginning
 - More likely to have unhandled contexts



Uncustomised TEI Stylesheet

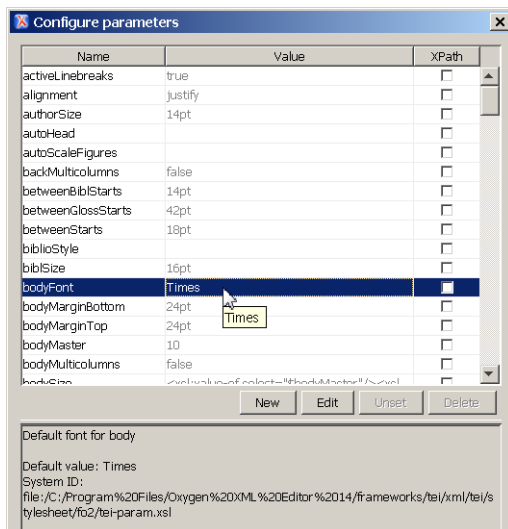
56



TEI Stylesheet Customisations

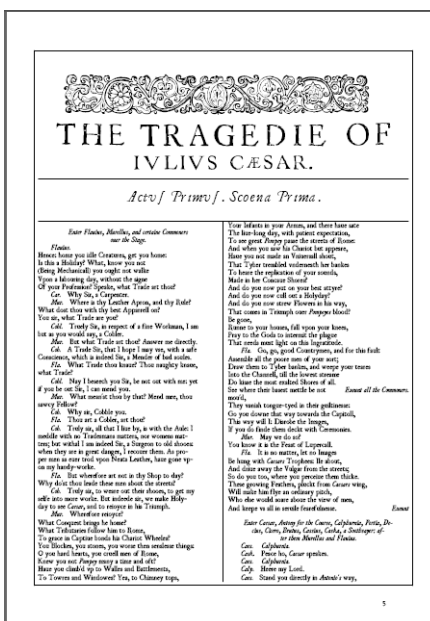
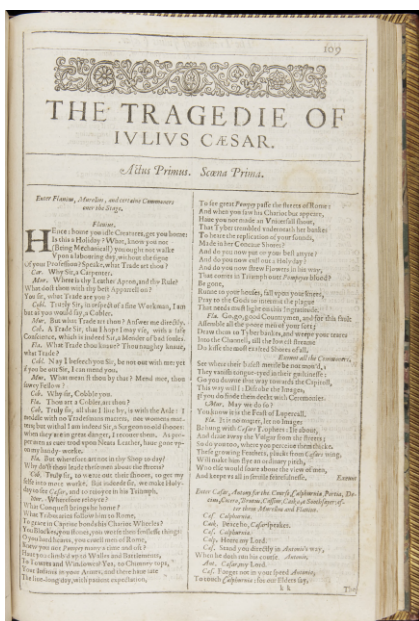
57

Some of the parameters that weren't changed:



Formatted With Custom Stylesheet

58



Julius Caesar as DocBook

59

- “DocBook ... is particularly well suited to books and papers about computer hardware and software (though it is by no means limited to these applications).”

Caesar's Head Into DocBook

60

```

<info>
  <title>THE TRAGEDIE OF IVLIVS CAESAR.</title>
  <author>
    <personname>Shakespeare, William</personname>
    <personblurb><para>1564-1616</para></personblurb>
  </author>
  <publisher>
    <publishername>University of Oxford Text Archive</publishername>
    <address>Oxford University Computing Services
13 Banbury Road
Oxford
OX2 6NN<email>ota@oucs.ox.ac.uk</email></address>
  </publisher>
  <biblioid class="uri">http://ota.ox.ac.uk/id/3014</biblioid>
  <biblioid class="isbn">1106000137</biblioid>
  <biblioid class="isbn">9781106000132</biblioid>
  <legalnotice security="restricted">
    <para>Distributed by the University of Oxford under a Creative
Commons
Attribution-ShareAlike 3.0 Unported License</para></legalnotice>
  <bibliosource>Revised version of http://ota.ox.ac.uk/id/0119</
bibliosource>
  <bibliosource>The texts were originally prepared by Trevor
Howard-Hill for use in his single volume concordances to

```

"DocBook, heere take you Caesars body"

61

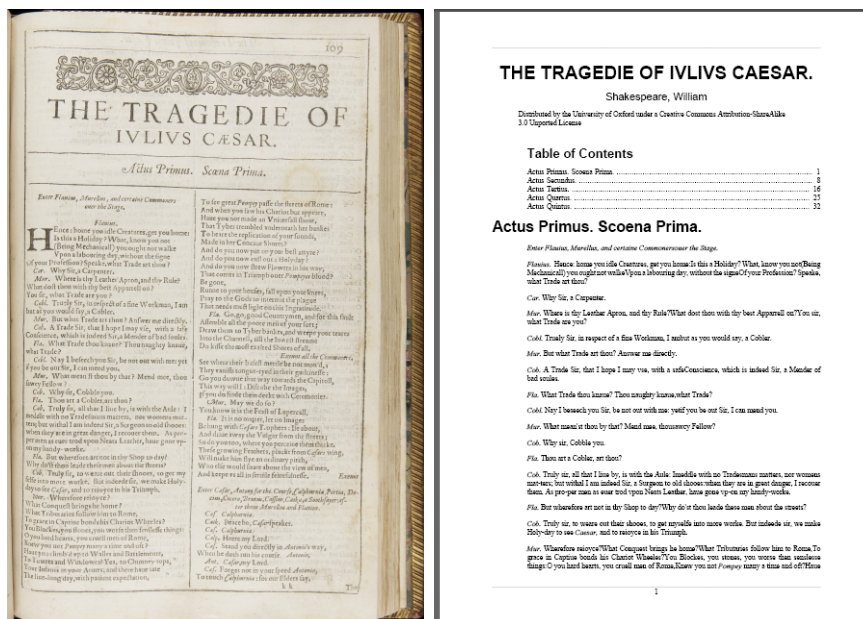
```

<section>
  <title>Actus Primus. Scoena Prima.</title>
  <anchor remap="lb" xml:id="n2"/>
  <para><emphasis role="stage">Enter Flavius, Murellus, and
certaine Commoners
<anchor remap="lb" xml:id="n3"/>ouer the Stage.</emphasis>
  <anchor remap="lb" xml:id="n4"/></para>
  <para role="sp">
    <emphasis role="italic">Flavius.</emphasis>
    <phrase role="ab">
      <anchor remap="lb" xml:id="n5"/>Hence: home you idle
Creatures,
get you home:<anchor remap="lb" xml:id="n6"/>Is this a Holiday?
What,
know you not<anchor remap="lb" xml:id="n7"/>(Being Mechanicall) you
ought not walke<anchor remap="lb" xml:id="n8"/>Vpon a labouring day,
without the signe<anchor remap="lb" xml:id="n9"/>Of your Profession?
Speake, what Trade art thou?<anchor remap="lb" xml:id="n10"/>

```

Formatted DocBook

62



Does It Work?

63

- Metadata elements in `<head>`/`<info>` similar but different
- More bibliographic details in TEI
- No encoding declaration in DocBook
- DocBook doesn't have elements for plays
- Capture original intent with `remap` or `role` attributes
- Stylesheet don't "just work" based on `remap` values
- Formatted output looks like DocBook, not like First Folio

Darwinian Evolution

64

- "The Darwin Information Typing Architecture (DITA) builds content reuse into the authoring process. DITA defines an XML architecture for designing, writing, managing, and publishing many kinds of information in print and on the Web."
- Modular structure means you can put Caesar's ghost before Caesar's death

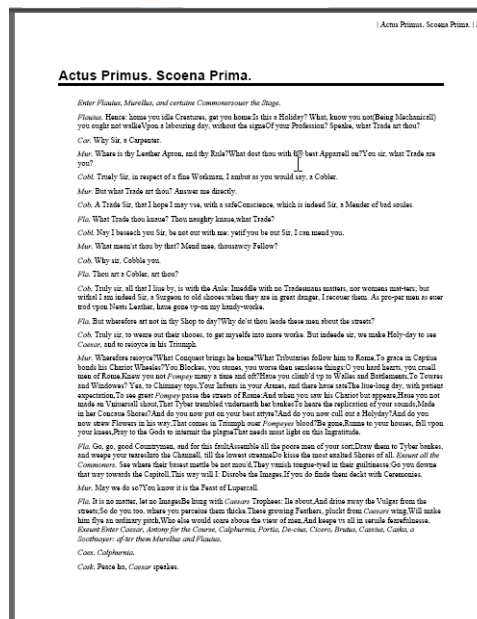
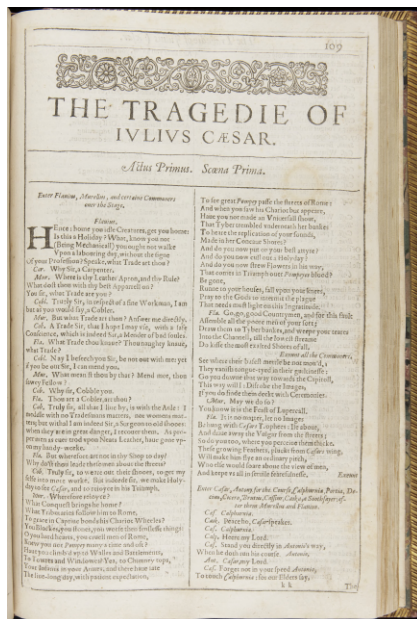
When Caesar Met DITA

65

```
<topic id="topic_qgj_w4y_jh">
  <title>Actus Primus. Scoena Prima.</title>
  <body>
    <unknown otherprops="lb" id="n2"/>
    <p><i props="stage">Enter Flavius, Murellus, and certaine
Commoners<unknown
otherprops="lb" id="n3"/>ouer the Stage.</i>
    <unknown otherprops="lb" id="n4"/></p>
    <p props="sp">
      <i props="italic">Flavius.</i>
      <unknown props="ab">
        <unknown otherprops="lb" id="n5"/>Hence: home you idle
Creatures, get you home:<unknown otherprops="lb" id="n6"/>Is
this a Holiday? What, know you not<unknown otherprops="lb"
id="n7"/>(Being Mechanicall) you ought not walke<unknown
otherprops="lb" id="n8"/>Vpon a labouring day, without the
signe<unknown otherprops="lb" id="n9"/>Of your Profession?
Speake, what Trade art thou?<unknown otherprops="lb"
id="n10"/>
```

Formatted DITA

66



XML as Target

67

- Advantages
 - Only if it suits your needs
 - Archive format, e.g., NLM
 - Interchange format
- Disadvantages
 - Not good for display without a processor for XML format

XML As Target: Central Bank of Ireland

68

- Central Bank of Ireland introduced new reporting requirements
- Monthly and quarterly reports required as XML
- Samples provided as spreadsheet and XML
- XML represents selected cells from spreadsheet form
 - Mostly <Form>, <Row>, and <Column>
- *Only* XML to be uploaded to repository
- Validated results returned as spreadsheet
- Economists at Financial Services Firm could do spreadsheets but not XML
- Implemented XSLT transform from exported Excel 2003 XML

Example

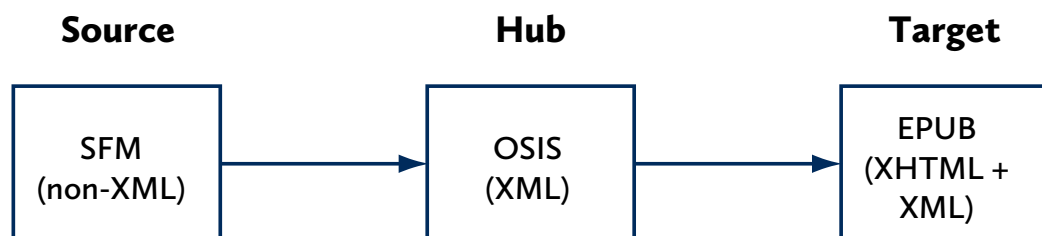
69

```
<ENTITIESForm>
  <Form institutionType="Other">
    <Row name="1">Finance
      <Column number="2">1000</Column>
      <Column number="3">270</Column>
      <Column number="4">375</Column>
      <Column number="5">355</Column>
    </Row>
  </Form>
  <Form institutionType="Investment Firms">
    <Row name="2">
      <Column number="2"></Column>
      <Column number="3"></Column>
      <Column number="4"> </Column>
      <Column number="5"></Column>
    </Row>
  </Form>
</ENTITIESForm>
```

One Person's Source is Another's Hub

70

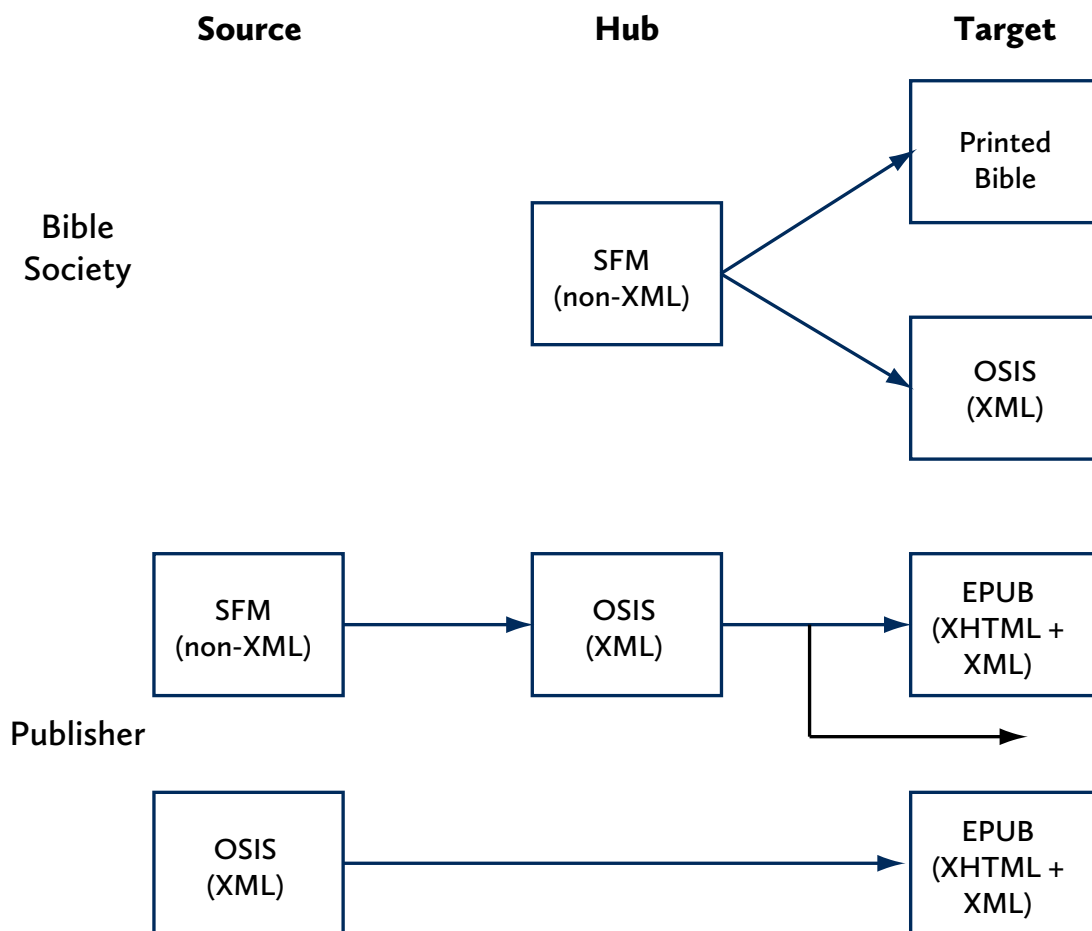
- SFM, OSIS, and EPUB from French Bible EPUBs from slide 35



- Bible Society receives back OSIS to sell or exchange with other groups
 - SFM is hub format since used for both print and OSIS
- Publisher uses OSIS as source, hub, and target formats:
 - Source when received from elsewhere
 - Hub when producing EPUB from SFM
 - Target to be returned to Bible Society

One Person's Source is Another's Hub

71



Questions?

72

References

73

- TEI for “The Tragedy of Julius Caesar”
<http://ota.ox.ac.uk/id/3014>
- “*Julius Caesar* (play)” at Wikipedia
https://en.wikipedia.org/wiki/Julius_Caesar_%28play%29
- First Folio at Wikipedia
https://en.wikipedia.org/wiki/First_Folio
- TEI P5 Guidelines
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index-toc.html>
- Central Bank of Ireland - Reporting Requirements
<http://www.centralbank.ie/regulation/industry-sectors/credit-institutions/pages/reporting.aspx>
- Open Scriptural Information Standard (OSIS)
<http://bibletechnologies.net/>
- DocBook
<http://www.docbook.org/>
- DITA
<http://dita.xml.org/>

Image Credits

74

- slide 42 – Public domain
http://commons.wikimedia.org/wiki/File:Julius_Caesar_Coustou_Louvre_MR1798.jpg
- slide 49 (detail) – Public domain
http://130.132.81.94/dl_crosscollex/brbldl/oneITEM.asp?pid=10000551&iid=1378077&srctype=

Exhibit 1

Inside Caesar's Head

```

<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>THE TRAGEDIE OF IVLIVS CAESAR.</title>
      <author>Shakespeare, William, 1564-1616</author>
    </titleStmt>
    <publicationStmt>
      <distributor>
        <name>University of Oxford Text Archive</name>
        <address>
          <addrLine>Oxford University Computing Services</
addrLine>
          <addrLine>13 Banbury Road</addrLine>
          <addrLine>Oxford</addrLine>
          <addrLine>OX2 6NN</addrLine>
        </address>
        <email>ota@oucs.ox.ac.uk</email>
      </distributor>
      <idno type="ota">http://ota.ox.ac.uk/id/3014</idno>
      <idno type="isbn10">1106000137</idno>
      <idno type="isbn13">9781106000132</idno>
      <availability status="restricted">
        <licence target="http://creativecommons.org/licenses/by-
sa/3.0/">
          Distributed by the University of Oxford under a Creative Commons
          Attribution-ShareAlike 3.0 Unported License
        </licence>
      </availability>
    </publicationStmt>
    <sourceDesc>
      <bibl>Revised version of <relatedItem type="older"
target="http://ota.ox.ac.uk/id/0119"/>
      </bibl>
      <bibl>The texts were originally prepared by Trevor
Howard-Hill for use in his single volume concordances to
Shakespeare (OUP, 1969f). They have since been reformatted
to modern standards and carefully proofread by staff of
Oxford University Press' Shakespeare Department for use in
the new "Old Spelling" Oxford Shakespeare, under the
general editorship of Dr Stanley Wells: <title>The complete
works
/ William Shakespeare</title>; general editors, Stanley
Wells and Gary Taylor ; editors Stanley Wells ... [et al.]
; with introductions by Stanley Wells. -- Oxford :
Clarendon Press, 1986. -- (Oxford Shakespeare). -- ISBN
0-19-812926-2</bibl>
      <biblFull>
        <titleStmt>
          <title>THE TRAGEDIE OF IVLIVS CAESAR.</title>
          <author>Shakespeare, William, 1564-1616</author>
          <editor role="editor">Lee, Sidney, Sir, 1859-1926</

```

```

editor>
    </titleStmt>
    <extent>xxxv, 908 p. : facsim. ; 39 cm.</extent>
    <publicationStmt>
        <publisher>Clarendon Press</publisher>
        <pubPlace>Oxford</pubPlace>
        <date>1902</date>
    </publicationStmt>
    <notesStmt>
        <note anchored="true">"One thousand copies of this
facsimile have been printed"--verso of half t.p.</note>
        <note anchored="true">Facsim. reprint of ed.
published, London : printed by Issac Iaggard and Ed.[ward] Blount,
1623 with original t.p.: Mr. William Shakespeares comedies,
histories, & tragedies</note>
        <note anchored="true">Original colophon reads:
Printed at the charges of W.[illiam] Iaggard, Ed.[ward] Blount, I.
[ohn] Smithweeke [i.e. Smethwick], and W.[illiam] Aspley, 1623</
note>
        <note anchored="true">Contents: The tempest. The two
gentlemen of Verona. The merry wives of Windsor. Measvre, for
measure. The comedie of errors. Much adoe about nothing. Loues
labour's lost. A midsommer nights dreame. The merchant of Venice.
As you like it. The taming of the shrew. All's well, that ends
well. Twelfe night, or what you will. The winters tale. The life
and death of King Iohn. The life and death of King Richard the
second. The first part of Henry the fourth. The second part of
Henry the fourth. The life of Henry the fift. The first part of
Henry the sixt. The second part of Henry the sixt. The third part
of Henry the sixt. The tragedy of Richard the third. The famous
history of the life of King Henry the eight. The tragedie of
Troylus and Cressida. The tragedy of Coriolanvs. The lamentable
tragedy of Titus Andronicus. The tragedie of Romeo and Ivliet. The
life of Tymon of Athens. The tragedie of Ivliivs Caesar. The
tragedie of Macbeth. The tragedie of Hamlet, Prince of Denmarke.
The tragedie of King Lear. The tragedie of Othello, the moore of
Venice. The tragedie of Anthonie, and Cleopatra. The tragedie of
Cymbeline</note>
    </notesStmt>
</biblFull>
</sourceDesc>
</fileDesc>
<encodingDesc>
    <charDecl>
        <glyph xml:id="ythou">
            <glyphName>Lower case y with smaller lower case u
above</glyphName>
            <desc>Abbreviation for thou in Elizabethan
typesetting</desc>
            <charProp>
                <localName>entity</localName>
                <value>ythou</value>
            </charProp>
        </glyph>
        <glyph xml:id="ythe">
            <glyphName>Lower case y with smaller lower case e
above</glyphName>

```

```

        <desc>Abbreviation for the in Elizabethan typesetting</
desc>
        <charProp>
            <localName>entity</localName>
            <value>ythe</value>
        </charProp>
    </glyph>
    <glyph xml:id="ythat">
        <glyphName>Lower case y with smaller lower case t
above</glyphName>
        <desc>Abbreviation for that in Elizabethan typesetting</
desc>
        <charProp>
            <localName>entity</localName>
            <value>ythat</value>
        </charProp>
    </glyph>
    <glyph xml:id="wwhich">
        <glyphName>Lower case w with smaller lower case c
above</glyphName>
        <desc>Abbreviation for which in Elizabethan
typesetting</desc>
        <charProp>
            <localName>entity</localName>
            <value>wwhich</value>
        </charProp>
    </glyph>
</charDecl>
<classDecl>
    <taxonomy xml:id="OTASH">
        <bibl>University of Oxford Text Archive Subject
Headings</bibl>
    </taxonomy>
    <taxonomy xml:id="LCSH">
        <bibl>Library of Congress Subject Headings</bibl>
    </taxonomy>
</classDecl>
</encodingDesc>
<profileDesc>
    <creation>
        <date notAfter="1623"/>
    </creation>
    <langUsage>
        <language ident="eng">English</language>
    </langUsage>
    <textClass>
        <keywords scheme="#LCSH">
            <term type="genre">Plays -- England -- 16th century</
term>
            <term type="genre">Plays -- England -- 17th century</
term>
            <term type="genre">Comedies -- England -- 16th century</
term>
            <term type="genre">Comedies -- England -- 17th century</
term>
            <term type="genre">Tragedies -- England -- 16th
century</term>

```

```
        <term type="genre">Tragedies -- England -- 17th
century</term>
      </keywords>
    </textClass>
  </profileDesc>
  <revisionDesc>
    <change when="2010-08-31">Header normalised</change>
  </revisionDesc>
</teiHeader>
```




summer school