Manuscripts in the Digital Age

XML-Based Catalogues and Editions

Katarzyna Anna Kapitan
@KatarzynaAnn
&
N. Kıvılcım Yavuz
@nkivilcimyavuz

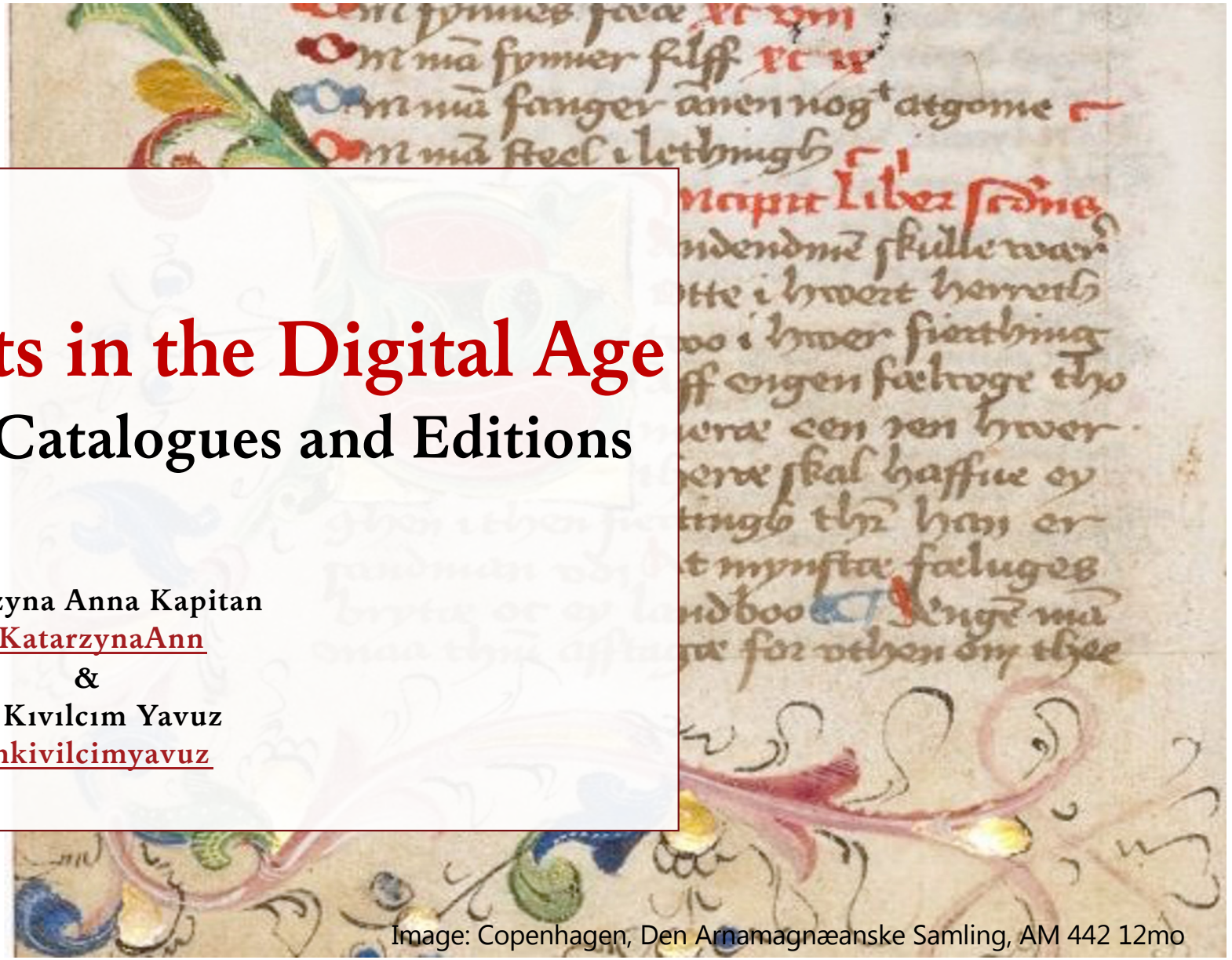


Image: Copenhagen, Den Arnamagnæanske Samling, AM 442 12mo

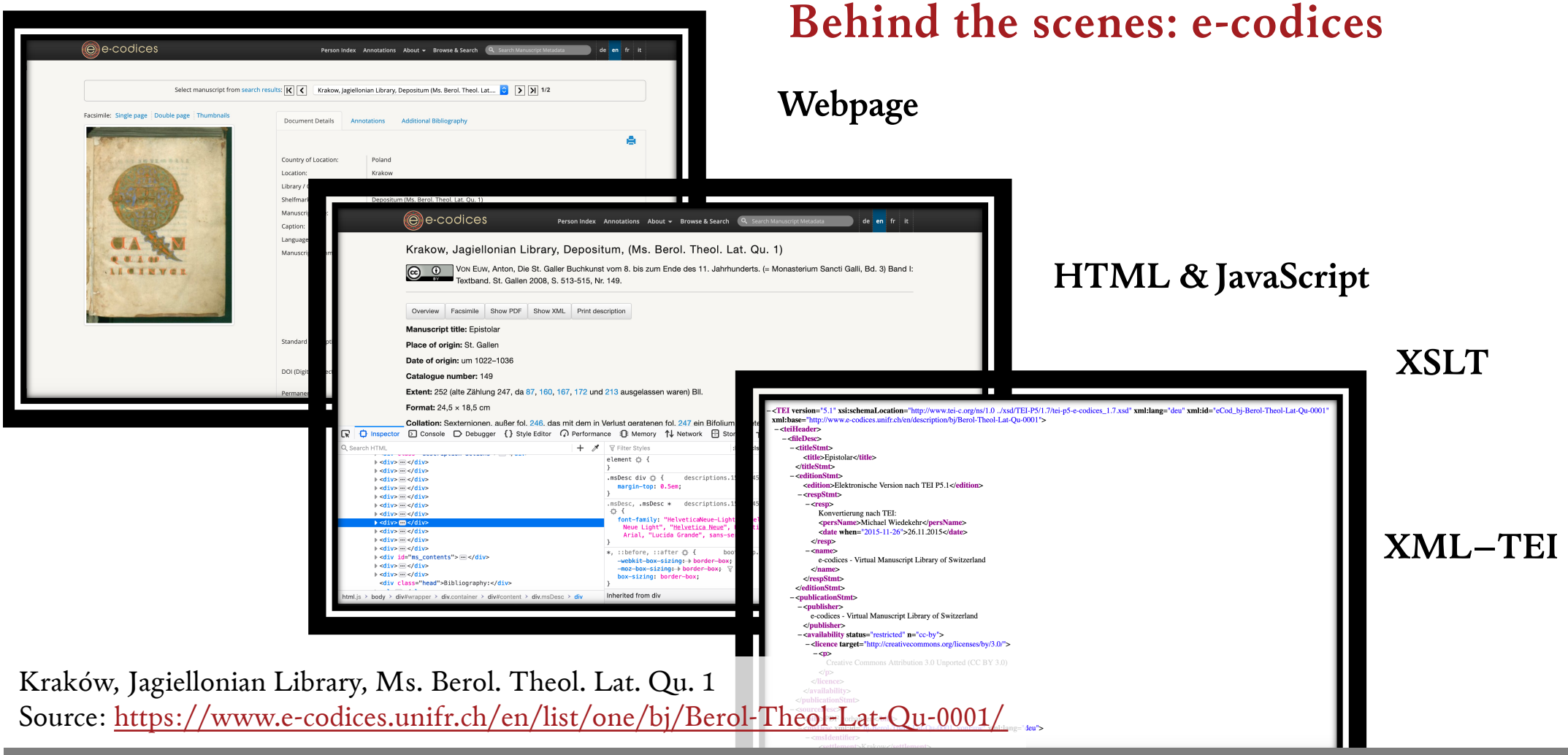
Image: Copenhagen, Den Arnamagnæanske Samling, AM 415 12mo

Introduction to HTML

Introduction to XML

Introduction to TEI

Behind the scenes: e-codices



Webpage

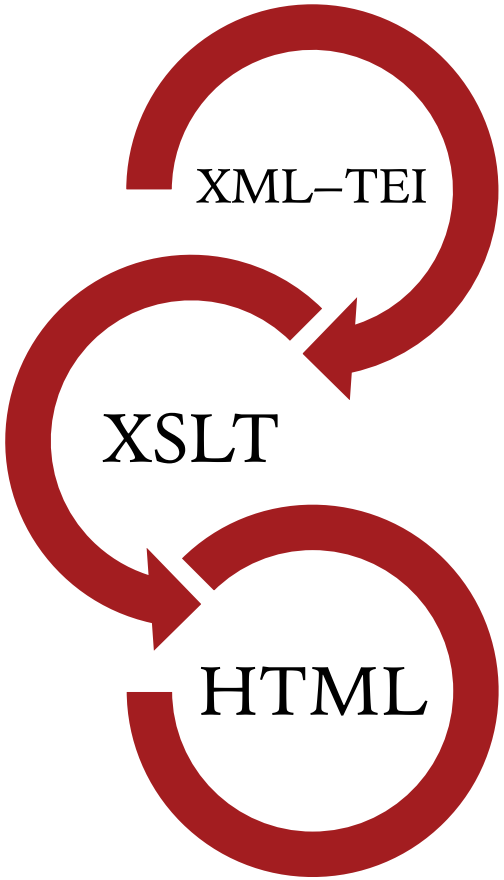
HTML & JavaScript

XSLT

XML–TEI

Kraków, Jagiellonian Library, Ms. Berol. Theol. Lat. Qu. 1
Source: https://www.e-codices.unifr.ch/en/list/one/bj/Berol-Theol-Lat-Qu-0001/

Basic Workflow: From Description to Viewer



Manuscript Description (Raw Data)

Transformation Scenarios

Basic Web Display

Markup Languages

- **What is a markup language?**
  - Computer language
  - Uses tags o define elements within a document
  - Human-readable
- **Examples:**
  - HTML (Hypertext Markup Language)
  - XML (Extensible Markup Language)
  - TeX & LaTeX
  - Scribe, GML (Generalized Markup Language) & SGML (Standard Generalized Markup Language)

HTML - Hypertext Markup Language

- HTML is the standard markup language for web pages.
- Markup:
  - **<tagname> Some text goes here... </tagname>**
  - Start Tag --- Content --- End Tag

HTML - Basic Structure of a Document

```
<!DOCTYPE html>
<html>
   <head>
      <title> This is my title</title>
   </head>
   <body>
      <p>This is a paragraph</p>
   </body>
</html>
```

Basic elements:

<html> – root element of an HTML page
<head> – contains meta information about the document
<body> – contains the visible page content

Other elements:

 - division/section of the page
<table> - table
<img> - image
 – list
<h1> – heading
<p> – paragraph

HTML-based manuscripts description?

```
<!DOCTYPE html>
<html>
    <head>
        <title>My title</title>
    </head>
    <body>
        <div>
            <h2>My shelfmark</h2>
            <p>My description</p>
        </div>
</body>
</html>
```

Exercise 1: HTML-based manuscript description? (1)

- Go to Workshop's Google Folder
- Go to sub-folder Exercises/Week 1/Day2 - Part1
- Create a new folder on your local drive for workshop-related materials
- Download to this new folder two file called **Day_2_AM_30_fol.html** & **Day_2_picture.jpg**
- Double-click on the html file. It should open in your default browser.

Exercise 1: HTML-based manuscript description? (2)

- Now open the same file with a plain text editor such as Notepad, Sublime Text or Oxygen XML Editor.
- You should see an html file that starts like this:

```
<!DOCTYPE html>
  <html>
        <head>
              <meta charset="utf-8"/>
              <title>Copenhagen, Den Arnamagnæanske Samling, AM 30
fol.</title>
        </head>
<body>
  <h1>Copenhagen, Den Arnamagnæanske Samling, AM 30 fol.</h1>
```

Exercise 1: HTML-based manuscript description? (3)

- Make a copy of your HTML file: Day_2_AM_30_fol.html
- Open the file in an editor
- Edit the file by separating the summary of the contents from the physical description, by creating a new paragraph.
- Create a separate header for the summary, within the contents section.
- Save the changes and display them in the browser

Clear separation of content and presentation

```
<div>
    <h2>Short Description</h2>
    <p>AM 30 fol. is a paper manuscript in folio format comprising 56 leaves gathered into five quires of six conjoint leaves each, with the exception of the last quire, with consists of four conjoint leaves only. The whole codex is made of one type of relatively thick laid paper with a "bull's head" watermark.
    </p>
</div>
```

Clear separation of content and presentation

HTML + CSS (Cascading Style Sheets)

- HTML

```
<h1> My Text </h1>
```

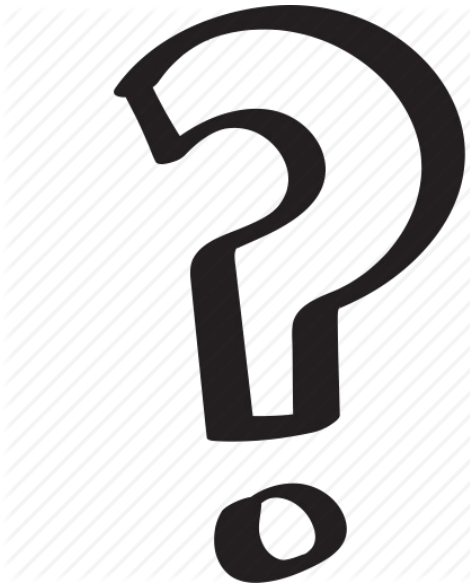
- CSS

```
h1 {
        font-variant: small-caps;
        color: #9D0E0E;
        text-decoration: underline;
        text-align: center;
}
```

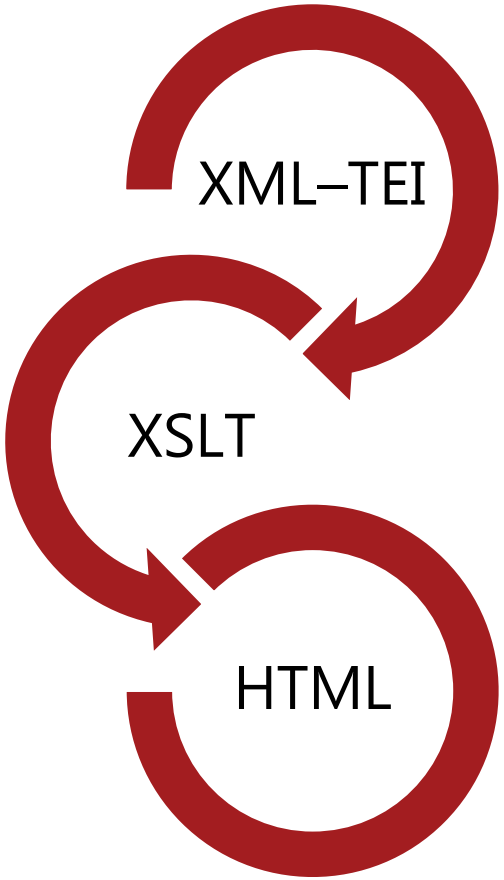
Exercise 2: Associate CSS with your HTML

- Go to Workshop's Google Folder
- Go to sub-folder Exercises/Week 1/Day 2 - Part 1
- Download the file called Day_2_style.css
- Associate your CSS file with your HTML file
  - By adding within <head> the element <link> with the following attributes:
    <link rel="stylesheet" type="text/css" href= "Day_2_style.css"/>
- Make sure you are using " Straight double quotes (U+0022) in your code.
  - On MAC IOS - Unicode Hex Input: alt + shift + 0022
  - On Windows hold alt + 0022 or 0022 – alt – x
    - (On Windows you should use the numeric keypad to type the numbers, not the keyboard)
- Open your HTML document in a browser

Questions



Basic Workflow: From Description to Viewer



Manuscript Description (Raw Data)

Transformation Scenarios

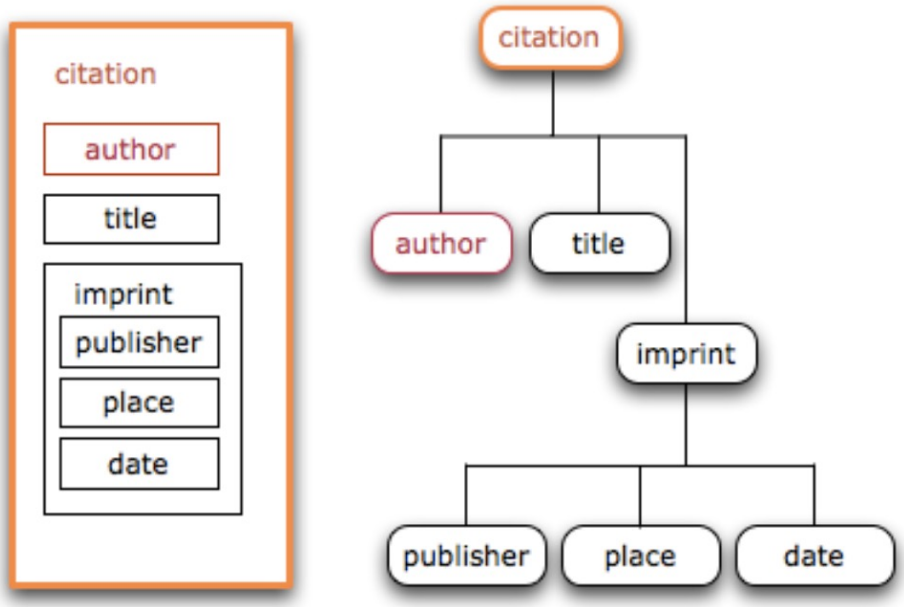
Basic Display (CSS)

XML – Extensible Markup Language

- Storing structured data
- International standard, non-proprietary
- Standard text format (expressed in plain text)– easy to parse and read for computer programs.
- Widely used to export and share structured data.
- Hardware and software independent

XML – Extensible Markup Language: Structure

XML Structure




```
<?xml version="1.0"
encoding="UTF-8"?>
<citation>
  <author>Katherine Hayles</author>
  <title>Writing Machines</title>
  <imprint>
    <publisher>MIT Press</publisher>
    <place>Cambridge, MA</place>
    <date>2002</date>
  </imprint>
</citation>
```

XML Elements & Tags

```
<!DOCTYPE html>
<html>
    <head>
        <title>This is a title</title>
    </head>
    <body>
        <p>This is a paragraph</p>
    </body>
</html>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<myRoot>
        <myContent>
                <content>
                        Here is my content
                </content>
        </myContent>
</myRoot>
```

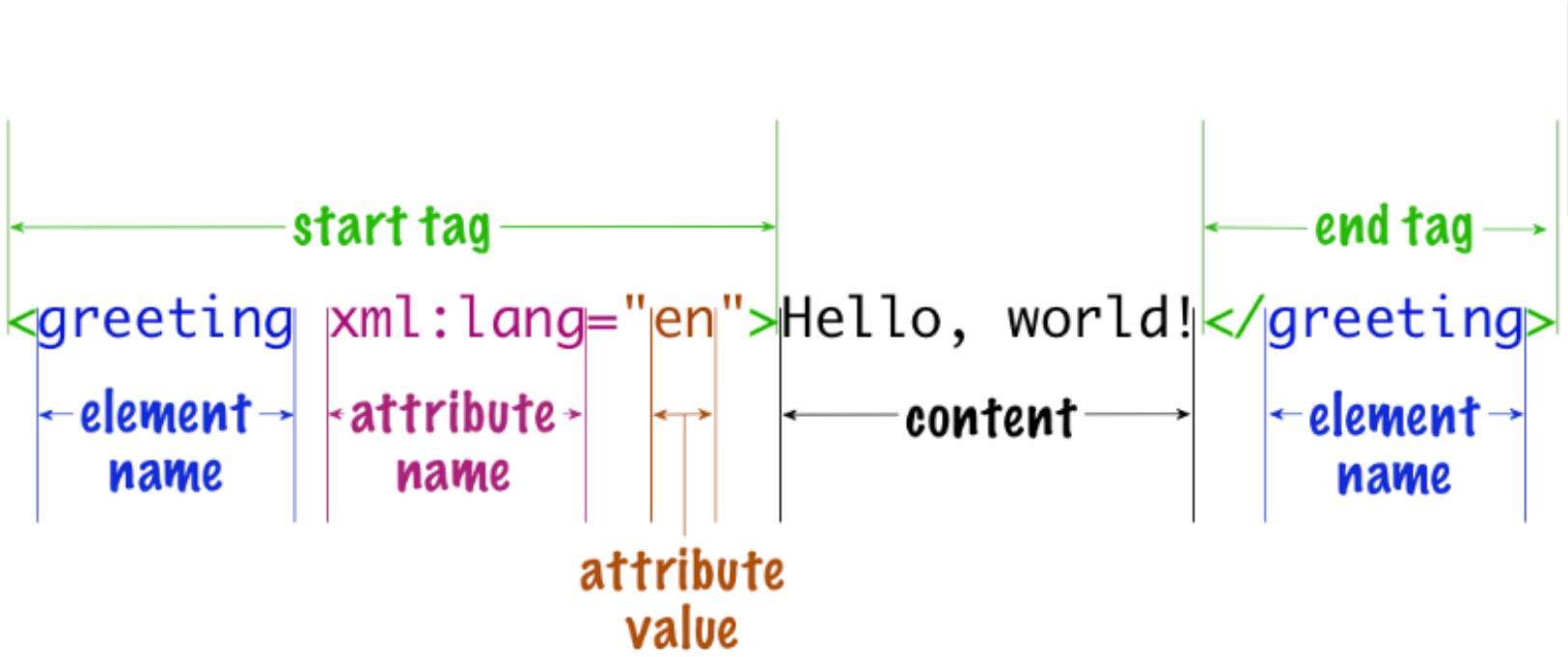
XML Elements & Tags

- Text is divided into elements (the nouns of the encoding — content objects).
- elements have start-tags and end-tags
  - <heading>My heading</heading>
- start-tags have < ... >
  - <heading>
- end-tags have </ ... >
  - </heading>

XML Attributes

- **Attributes** are adjectives of XML, they describe the properties of the elements.
- Any number of **attributes** can be specified on a given start-tag, but only one with a given name.
- Attribute values describe the attributes.

- **<person job="musician" age="55">**Paul Simon**</person>**



Quiz (Group Work)

- Which of the following are well-formed XML?
  1. <name type="person>Pearl S. Buck</name>
  2. <name type="person">Toni Morrison<name>
  3. <name="person">Carl Sagan</name>
  4. <name type="person">Kurt Vonnegut</name>
  5. <name type=person>John Cleese</name>
  6. <name type="person"><forename>Frances</forename>
     <surname>Perkins</surname></name>

Clear Separation of Content and Presentation

```
<?xml version="1.0" encoding="UTF-8"?>
<myRoot>
        <myContent>
                <content>
                        Here is my content
                </content>
        </myContent>
</myRoot>
```

```
<?xml version="1.0" encoding="UTF-8"?>

<workshop name="MiDA”>
        <instructors>
          <name>
            <firstName>Katarzyna</firstName>
            <lastName>Kapitan</lastName>
          </name>
          <name>
            <firstName>N. Kıvılcım</firstName>
            <lastName>Yavuz</lastName>
          </name>
        </instructors>
        <participants>
          <name>
            <firstName>Participant</firstName>
            <lastName>One</lastName>
            <affiliation/>
          </name>
          <name>
            <firstName>Participant</firstName>
            <lastName>Two</lastName>
            <affiliation/>
          </name>
          <name>
            <firstName>Participant</firstName>
            <lastName>Three</lastName>
            <affiliation/>
          </name>
        </participants>
    </workshop>
```

XML Structure

Workshop:

Instructors:

KAK

NKY

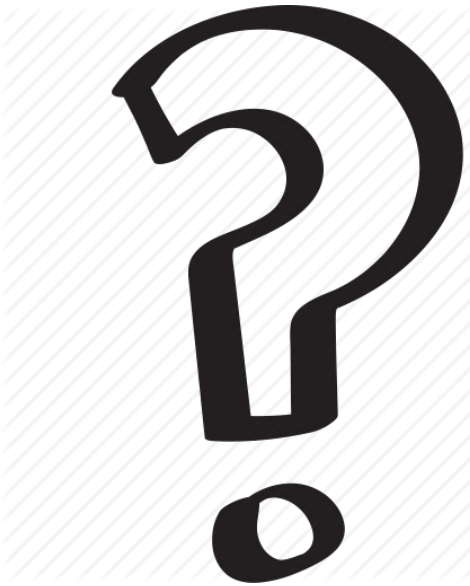
Participants:

P1

P2

P3

Questions



```
<workshop name="MiDA”>
        <instructors>
                    <name>
                ...
                    </name>
                     <name>
                ...
                    </name>
        </instructors>
        <participants>
          <name>
              ...
                    </name>
                     <name>
                ...
          </name>
        </participants>
</workshop>
```

```
<workshop name="Stylo">
        <teachers>
                    <name>
                ...
                     </name>
                    <name>
                ...
                    </name>
        </teachers>
        <students>
            <name>
              ...
                    </name>
                     <name>
                ...
                    </name>
        </students>
</workshop>
```

Need for a lingua franca: TEI (Text Encoding Initiative)

Concepts

XML

Syntax

TEI

Language:
vocabulary and grammar


```
<element>
    <element attribute="value">
        content
    </element>
</element>
```

```
<p>

<note type="foot">

<head>
```

TEI (Text Encoding Initiative)

- The TEI is an international and interdisciplinary standards project
- Established in 1987 to develop, maintain and promulgate hardware- and software-independent methods for encoding humanities data in electronic form.
- A markup language (a text encoding language)
  - Developed by an international consortium;
  - Free and open-source.
  - Both a community standard and a community research effort

TEI Guidelines

https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html

< Text Encoding Initiative >

P5: Guidelines for Electronic Text Encoding and Interchange

Version 3.5.0. Last updated on 29th January 2019, revision 3c0c64ec4

[English] [Deutsch] [Español] [Italiano] [Français] [日本語] [한국어] [中文]

Front Matter

Title

i. Releases of the TEI Guidelines

ii. Dedication

iii. Preface and Acknowledgments

⊞ iv. About These Guidelines

⊞ v. A Gentle Introduction to XML

⊞ vi. Languages and Character Sets

Back Matter

⊞ Appendix A Model Classes

Text Body

⊞ 1 The TEI Infrastructure

⊞ 2 The TEI Header

⊞ 3 Elements Available in All TEI Documents

⊞ 4 Default Text Structure

⊞ 5 Characters, Glyphs, and Writing Modes

⊞ 6 Verse

⊞ 7 Performance Texts

⊞ 8 Transcriptions of Speech

⊞ 9 Dictionaries

⊞ 10 Manuscript Description

TEI (Text Encoding Initiative)

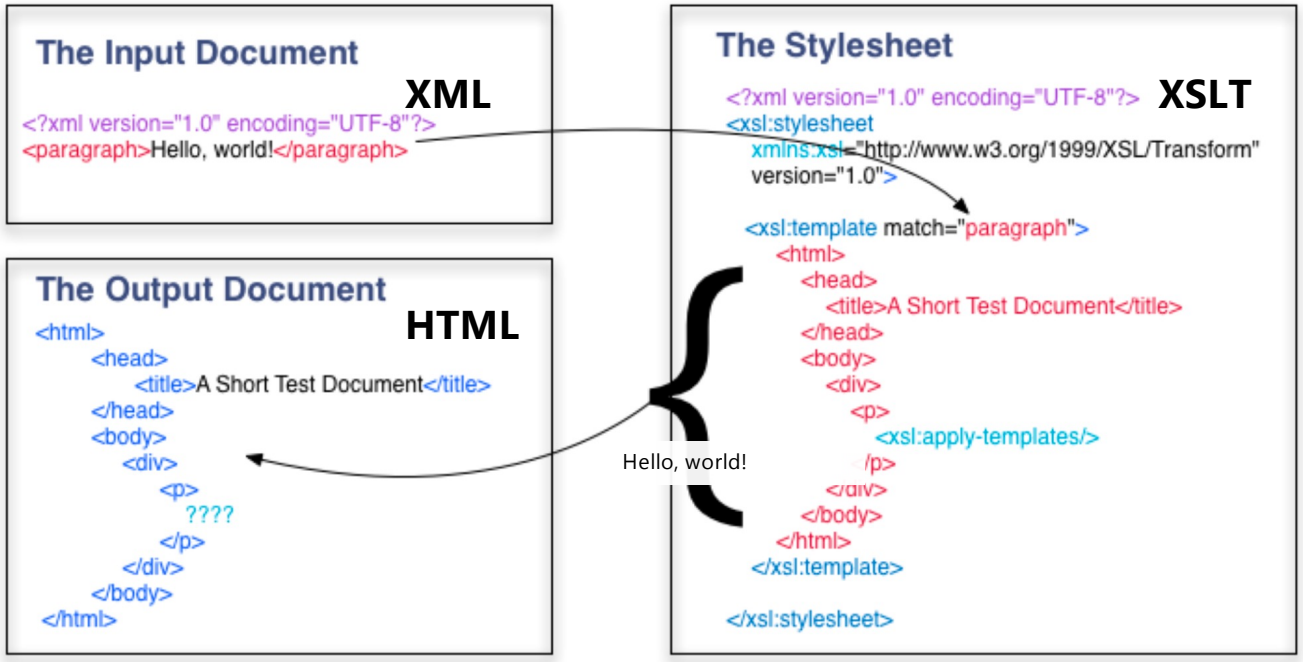
- The TEI Guidelines help us model our research materials:
  - Sustainably
  - Sharably
  - Analytically
  - Formally

Exercise 3: XML-based manuscript description

- Go to Workshop's Google Folder
- Go to sub-folder Exercises/Week 1/Day 2 - Part 1
- Download the XML file named **AM_30_fol.xml** to your drive
- Open **AM_30_fol.xml** with Oxygen XML Editor

- Which elements do you see in the file? Can you guess what information they contain?

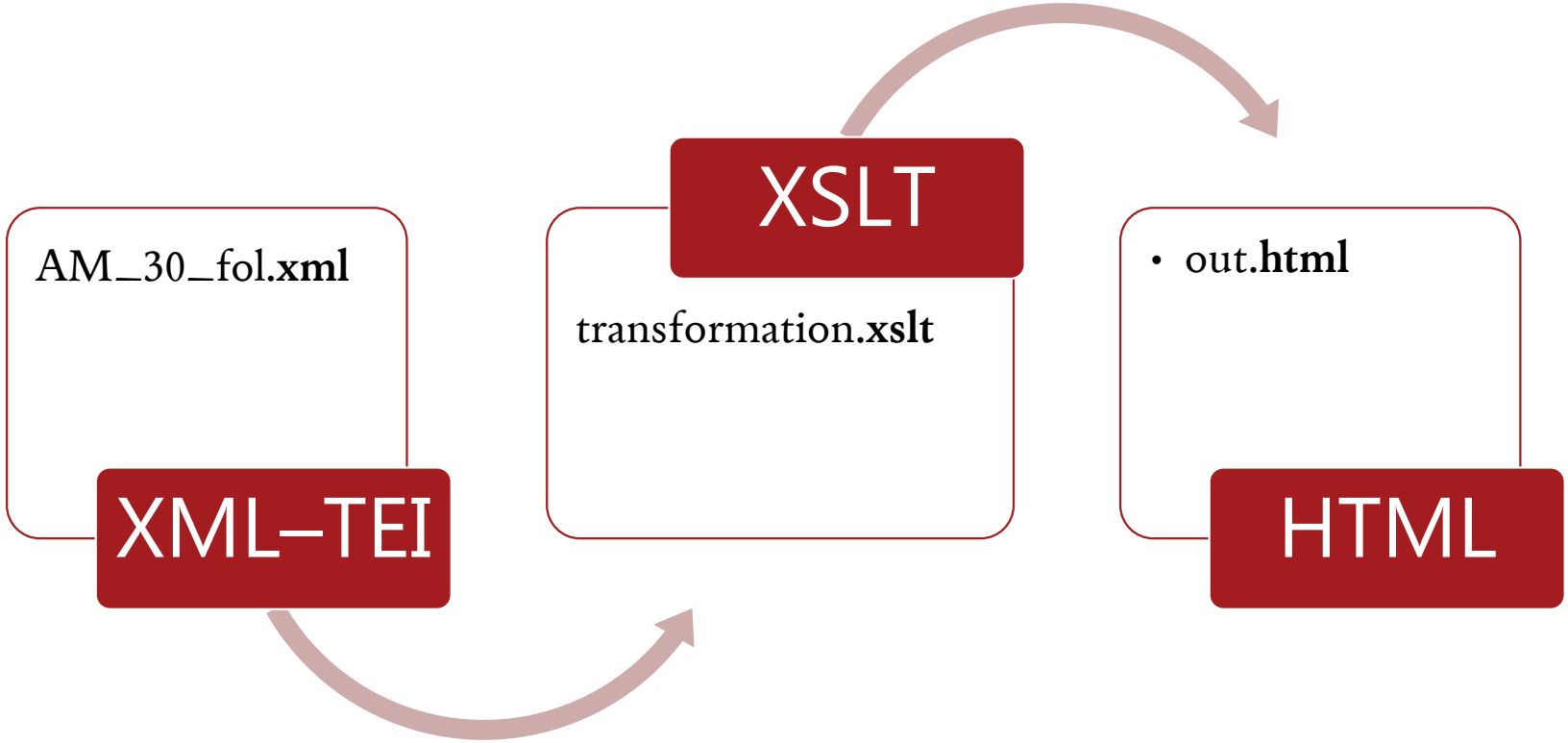
Basic Workflow: From Description to Viewer
XSLT (Extensible Stylesheet Language Transformations)



Exercise 4: With XSLT from XML to HTML

- Go to Workshop's Google Folder
- Go to sub-folder Exercises/Week 1/Day 2 - Part 1
- Download the files called **Day_2_transformation.xsl** & **Day_2_Guidelines_Transformation_Scenario_XML_HTML.docx**
- Convert your TEI-XML to HTML with XSLT, following the guidelines

Transformation



COPENHAGEN, DEN ARNAMAGNÆANSKE SAMLING, AM 30 FOL.

SHORT DESCRIPTION

AM 30 fol. is a paper manuscript in folio format comprising 56 leaves gathered into five quires of six conjoint leaves each, with the exception of the last quire, with consists of four conjoint leaves only. The whole codex is made of one type of relatively thick laid paper with a "bull's head" watermark. AM 30 fol. contains two main texts dealing with the Christianization of the Slavs.

CONTENTS

- *Chronica Slavorum by Helmold of Bosau*
- *Chronica Slavorum by Arnold of Lübeck*

Initial HTML file: AM_30_fol.html

AM_30_fol.xml

TEI teiHeader fileDesc titleStmt

```
19                <idno>AM 30 fol.</idno>
20             </msIdentifier>
21             <msContents>
22                <summary>AM 30 fol. is a paper manuscript
23                   gathered into five quires of six conjoi
24                   the last quire, with consists of four c
25                   made of one type of relatively thick la
26                   AM 30 fol. contains two main texts deal
27                   Slavs.</summary>
28                <msItem>
29                   <title>Chronica Slavorum</title>
30                   <author>Helmold of Bosau</author>
31                </msItem>
32                <msItem>
33                   <title>Chronica Slavorum</title>
```

Text Grid Author

XML file

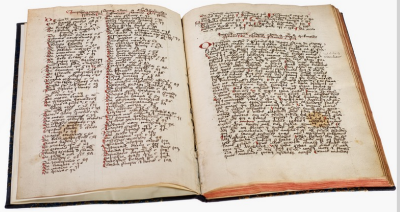
COPENHAGEN, DEN ARNAMAGNÆANSKE SAMLING, AM 30 FOL.

SHORT DESCRIPTION

AM 30 fol. is a paper manuscript in folio format comprising 56 leaves gathered into five quires of six conjoint leaves each, with the exception of the last quire, with consists of four conjoint leaves only. The whole codex is made of one type of relatively thick laid paper with a "bull's head" watermark. AM 30 fol. contains two main texts dealing with the Christianization of the Slavs.

CONTENTS:

- *Chronica Slavorum by Helmold of Bosau*
- *Chronica Slavorum by Arnold of Lübeck*



Result of transformation of XML file with XSLT: out.html

Additional Exercises

- Go to subfolder Exercises/Week 1/Day 2 - Part 1
- Follow the instruction in Day_2_Additional_Exercises

Questions