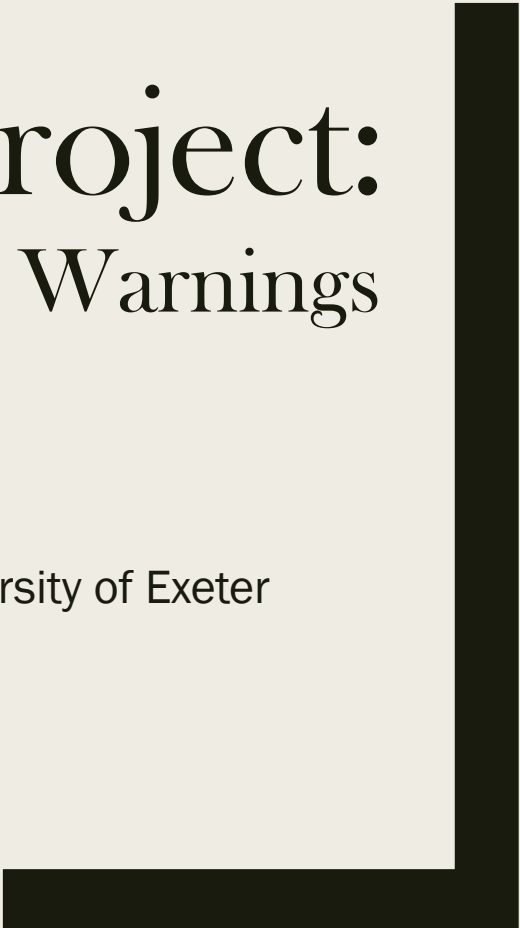




# Planning Your Project:

## Tips, Tricks, Standards And Warnings

Dr Lizzy Williamson  
Research Fellow in Digital Humanities, University of Exeter  
@earlymodernpost



- Project planning
- Documentation and encoding policies
- Directories and projects
- Standards and standardization

# What's in a research project?

- Research questions: why?
- Goals: where do you go?
- Non-goals: where do you stop?
- Methodology: what? How?
- Logistics: how? Who with? When?
- Audience: who for?
- An ending...

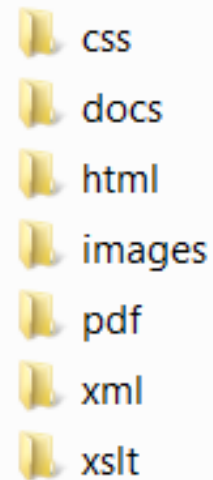
# What's in a digital edition?








- You have your transcriptions in XML – what else do you need?
- Publication options, user interface design, server space, continued support? These require funding...
- Managing without institutional support:
  - *a project of two parts?*
  - *Teach yourself how to transform XML?*
  - *Other publication options (e.g. in-built stylesheets, TEI publisher, contributing to larger projects, publish in specialist journal like <http://scholarlyediting.org/>)?*

# What's in a digital edition?

 ExampleProject\_June2019 >

- XML > your encoded transcriptions
- Schema > to constrain/validate your encoding
- XSLT > to transform your xml files
- HTML > for display online (live or static)
- PDF folder > you may want downloadable PDFs
- Images folder > you may have images to store/display
- Docs folder > to store internal project documentation
- CSS > to make things pretty
- XQuery > to query and extract your data
- js/jquery/json > to make things interactive/jazzy



-  css
-  docs
-  html
-  images
-  pdf
-  xml
-  xslt

# Starting your project

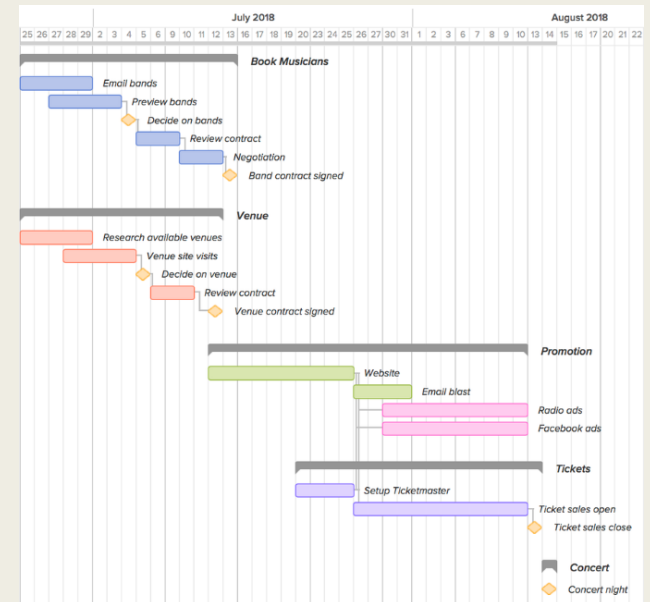
- Project assumptions: what, why, who for, how, when, who?
- Environment: Look at other editions, what do you like/dislike, what were their methods, view their xml/html? What already exists? Related materials to link to, projects whose materials overlap with yours?
- Are you clear on your research objects? What are the corpus limits? What are criteria for inclusion/exclusion? What is the meaning of your corpus? What happens if you discover additional texts?
- Are there ethical or legal issues? Personal data/images? Issue of copyright? Archival permissions?

# Project planning

- What do you need to do to your research objects?  
Break this into stages.
- Articulate workflow
  - *Encoding **cycle**: document analysis, schema development, mark up documents*
  - *Where does transcription fit in? During/prior to encoding? OCR and/or manual? Trial transcription to help timing estimates.*
  - *Publication **cycle**: prepare texts, process texts, publish texts*
  - *Iterative!*

# Project planning

- Alternate viewpoints throughout: overview, detail
- What are your milestones and dependencies? Gantt chart? Identify flexibility/rigidity across your processes and goals.
- Soft launches, launches
- Sunsetting and archiving





# Exercise 1

What are you own:

- project assumptions: what/why/who for/how? (includes when/in what order/who with)
- research objects
- work stages

Discuss these with your neighbour. It can be an imaginary or speculative project; if you can't think of anything use example of discovering a new archive of letters by Emily Bronte and wanting to photograph, transcribe, encode, and put them online.

# Documentation

- Before, during, after:
  - *Planning your encoding: policy and schema*
  - *Adapting to new features*
  - *Documenting and archiving*
- Readme files
  - *Use for your own working practice and when sharing and archiving*
  - *Use platform agnostic file formats (plaintext, markdown)*

# Documentation: commenting

- Comment your code: leave brief descriptions of what each part does – really useful for yourself and others!
- Comments in your encoding: leave findable markers like @todo, or questions coupled with initials of a team member (QLW)

```
function setup_map_geonames(geonames_id, zoom)
{
    // Get data from api.geonames.org
    $.getJSON("http://api.geonames.org/getJSON?formatted=true&geonameId=" + geonames_id

    .done(function(data)
    {
        // Set location
        var location = [parseFloat(data.lat), parseFloat(data.lng)];

        // Draw map and set zoom/bounds
        if(data.bbox == undefined)
        {
            map = L.map("map").setView(location, zoom, {maxZoom: maxZoom})
```

`<p>`Possible types include: maintext (for the primary textual sections of the document), witness (for witness statements), invocation, cipher, superscription (for endorsements), quran (for Qur'anic verses), vnotes (for verso notes)`</p>`

`<!-- QNC: what do ciphers look like? Are they a textual division? Do we need to consider a tabular layout? -->`

```
.fitBounds([[data.bbox["north"], data.bbox["west"]],[data.bbox["south"], dat
```

# Documentation: Encoding policy

- It is essential to articulate and record your editorial and encoding decisions.
- Your approach will depend on the size and complexity of the project
- Value of sharing your documentation

## Options:

- End-of-project documentation in TEI header (<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-encodingDesc.html>),
- Narrative notes for your own reference
- Write Encoding Guidelines in TEI...


# Exercise: Writing an Encoding Policy in TEI

- Start with document analysis: What are the structural units and thematic entities in your texts?
- Here you can document how you will encode: TEI header, structural mark-up, thematic mark-up, other entities like people and places...

# Exercise: Writing an Encoding Policy in TEI

- Save the example project folder on your desktop
- In oXygen, create xml file (TEI P5 all) and save inside your project folder
- Use structural elements: <div>, <head>, <p> etc
- Use elements for marking up examples, elements, attributes etc:  
<egXML xmlns="http://www.tei-c.org/ns/Examples">, <gi>, <att>, <val>, <tag>

# Exercise: Writing an Encoding Policy in TEI

- Transform using in-built stylesheets...
- Click top toolbar icon: 
- TEI P5 XHTML, Duplicate.
- Name it 'Encoding Policy'
- Can look at Parameters, Output instructions
- Click 'ok', 'Apply associated'

# Documentation: File naming

- human and machine readable
- sort and search manually, computationally easy to process
- Avoid: special characters, spaces, periods
- Meaningful: relate to project and specific text
- For example, use shelfmark, unique identifier, dates, consistently abbreviated name:  
dhe-hl-h709.xml      fl\_mathura\_1644\_2671-by-9a.xml
- Same applies to xml:ids!

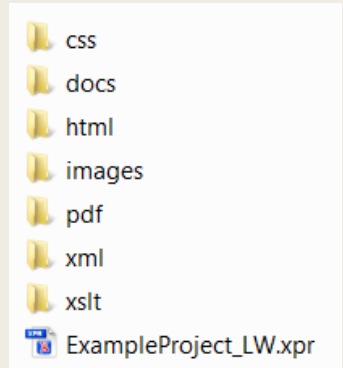
```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="dhe-hl-h709">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Letter from Thomas Hardy to Dorset County Museum
        <author ref="#ps:TH">Thomas Hardy</author>
        <!--The editor of the tei file-->
```



# Directories and Projects

- Your folder structure should also be logical and consistent
- Think about your processes, research objects, and entities – keep these well-ordered and nested appropriately
- Orderly nesting makes it easier to point to different folders when you want to reference your schema, point to your css, set up automatic transformations using your xslt, and so on

U:/ExampleProject\_June2019/docs/EncodingPolicy.html



# Directories and Projects

- Setting up a .xpr project file in oXygen helps you manage files grouped under the same XML project – it makes it easier to work with single or multiple XML editions/projects
- Exercise: set up a project file
  - *Project – New project – navigate to your project folder. Simple!*
- Project panel allows you to
  - *Group physical folders and files (mirrors of actual files in local system)*
  - *Create logical folders (virtual groupings, create and delete without affecting the physical files)*
  - *validate or transform whole folder (right click folder, ‘validate’)*
  - *toggle easily between projects*

# Directories and Projects

- Try setting up a transformation scenario for your XML folder in your project panel
  - *Right click XML folder in project panel, configure transformation*
  - *New XML transformation with XSLT*
  - *XSL URL: click folder icon, point to the file in XSLT folder*
  - *Output tab, save as `../..../html/${cfn}.html`*

Break it down – what does the filepath mean?

*/ = path separator*

*. = self, present context*

*.. = parent*

*\${cfn} = current filename, for each file being transformed*

The same logic will come up when you do XPath, XSLT...

# Tips for working with multiple files

- Bulk validation and transformation on folders via project panel
- Find and Replace in Files (Find, icon, right click folder) (be very careful!)
- Find differences between two files: Tools > Compare files, set source and target, move changes from one to the other
- Check spelling in files (depending on your texts!)
- Set code templates for frequent snippets: Options > Preferences > go to Editor > Templates > Code Templates. You can store these to specific projects, share them etc.

# Directories and Projects

- Secure working

- *Have master files with regular backups, separate from published/working copies.*
- *Versioning systems: single repository for your files, you 'check out' a local copy, make changes, then 'check in' those changes so they are merged to the master files.*
- *For example, Git, svn (tools – svn client), sharepoint (window – show view – sharepoint browser)*

# Standardization and linking: People

- Brief review: Could tag as name in your text, could reference an identifier within that document
- Shared entities benefit from having a separate authority list: one place to add additional detail, change content etc.
- Who gets an entry? Primary participants (authors, correspondents etc), anyone mentioned? Fictional people?
- Person reference in text, with @ref:  
`<persName ref="#VirginiaWoolf">V Woolf</persName>`  
`<persName ref="#ps:VirginiaWoolf">V Woolf</persName>`
- Look at person record in person file, with xml:id  
`<person xml:id="VirginiaWoolf">...`

# LOD and VIAF: Linked Open Data and Virtual International Authority File

Citation list in person record:

```
<listBibl type="semantic">  
  <bibl type="viaf" n="39385478"/>  
</listBibl>  
<listBibl type="general">  
  <bibl type="wikipedia" n="Virginia_Woolf"/>  
</listBibl>
```

<https://viaf.org/viaf/39385478/>

[https://en.wikipedia.org/wiki/Virginia\\_Woolf](https://en.wikipedia.org/wiki/Virginia_Woolf)

Other links and references: ODNB, paid resources, EM People module...

Exercise: add a new person entry to the people list, including an authority file reference. Choose one that is relevant to your project or imagined project

# Standardization and linking: Places

- Same logic applies to places: <placeName> reference in text with @ref, <place> record in places file with xml:id
- Create a places file in metadata folder, use <listPlace>
- Can have multiple names, can have language and date attributes (@notAfter)
- <location> for geographical coordinates, for an address; how granular?
- Shared identifiers: Geonames, urban gazetteers (Pelagios)
- Can use Geonames API to create map view, e.g. [Cotton Famine](#)

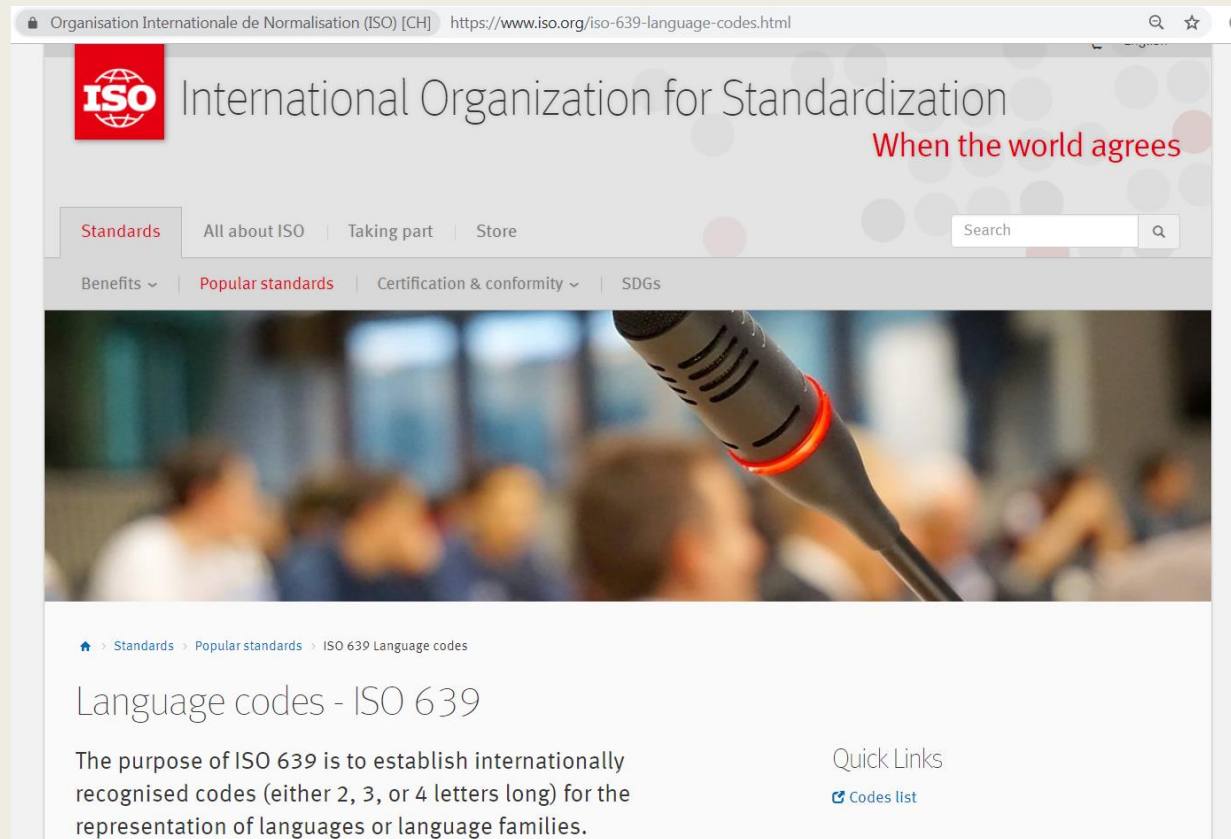
```
<listBibl type="semantic">  
  <bibl type="geonames" n="4930956"/>  
</listBibl>
```



# Standardization and linking

- What entities recur in your texts? Do you have additional information you'd like to add to a separate record?
- Works mentioned/bibliography?
- Add a list of project participants? Add reference to <resp> statements in TEI header.

# Standardization: languages and dates



The screenshot shows a web browser window displaying the ISO website. The address bar shows the URL: <https://www.iso.org/iso-639-language-codes.html>. The page header features the ISO logo (a red square with a white globe and the letters 'ISO') and the text 'International Organization for Standardization' and 'When the world agrees'. Below the header is a navigation menu with links for 'Standards', 'All about ISO', 'Taking part', and 'Store'. A search bar is also present. The main content area features a large image of a microphone in the foreground, with a blurred background of people in a meeting. Below the image, the breadcrumb trail reads: [Home](#) > [Standards](#) > [Popular standards](#) > [ISO 639 Language codes](#). The main heading is 'Language codes - ISO 639'. The introductory text states: 'The purpose of ISO 639 is to establish internationally recognised codes (either 2, 3, or 4 letters long) for the representation of languages or language families.' To the right, there is a 'Quick Links' section with a link to 'Codes list'.

Organisation Internationale de Normalisation (ISO) [CH] <https://www.iso.org/iso-639-language-codes.html>

**ISO** International Organization for Standardization  
When the world agrees

Standards | All about ISO | Taking part | Store

Search

Benefits | Popular standards | Certification & conformity | SDGs

[Home](#) > [Standards](#) > [Popular standards](#) > [ISO 639 Language codes](#)

## Language codes - ISO 639

The purpose of ISO 639 is to establish internationally recognised codes (either 2, 3, or 4 letters long) for the representation of languages or language families.

Quick Links

- [Codes list](#)

# Standardization: languages and dates

- ISO codes, @lang, <foreign> element
- Right-to-left languages in oXygen
- Unicode/special characters (Edit -> Insert from character map)
- Different calendars, @when attribute for standardization

```
<div xml:lang="per" type="maintext">
  <p n="1"><persName ref="#ps:nathu">ناتھو</persName> ولد <persName ref="#ps:maqbul">مقبول</persName> ساکن بلده <placeName ref="#pl:burhanpur">برہانپور</placeName> ام چون بر زمین چودھری <persName ref="#ps:chaudhri_purshottam_das">پرشتوم داس</persName>
  <lb n="2"/><persName ref="#ps:Alamraj">عالم راج</persName> و <persName ref="#ps:Mohan">موهن</persName> خانہ کھپرل و چاردیواری انداختہ بودم عملہ آنرا قیمت نموده بحضور
  <persName ref="Thakur">تھ کر</persName> (؟) <lb n="3"/>دادم <lb n="4"/>در قبض و تصرف <measure unit="rupee" quantity="13">سیزده روپیہ</measure> مبلغ
  <lb n="5"/>دادم کہ ثانی الحال بوقت نوشتہ خود در آوردم این چند کلمہ بطریق سند ...
  <date calendar="hijri-qamri" when="1661-01-10">9</date> بکار آید تحریر فی التاریخ</p>
</div>
<div xml:lang="hin" type="maintext">
  <p>
    <lb n="1"/> लिखतं शे <persName>नाथु</persName> वलद शेख <persName>मकबुल</persName> सा
    <placeName>बुरहानपुर</placeName> आगे
    <lb n="2"/> चौ <persName>परसोताम दास</persName>
    कि जमी पर घर खपरेल छार देवा-
    <lb n="3"/>ली दाली थी अमला उसका मोल की जा हजुर
  </p>
</div>
```

# Standardization: original and modernized spelling

- Early modern spellings – use fuzzy searching, or regularize?
- `<choice><orig>olde</orig><reg>old</reg></choice>`
- Tools like morphadorner, VARD 2



# Sunsetting

- Nothing lasts forever: distinguish data from manifestation
- Releasing texts, files, code: download, Github, institutional repository
- Creative Commons licensing: legal specifics, human readable, machine readable. Include icon on site.  
<https://creativecommons.org/>
- Sharing, licensing, documenting and archiving = these make your work sustainable



**Attribution**  
**CC BY**



**Attribution-ShareAlike**  
**CC BY-SA**



**Attribution-NoDerivs**  
**CC BY-ND**



**Attribution-NonCommercial**  
**CC BY-NC**



**Attribution-NonCommercial-ShareAlike**  
**CC BY-NC-SA**



**Attribution-NonCommercial-NoDerivs**  
**CC BY-NC-ND**



# Questions? Clarifications? Comments? Experiences?

Dr Lizzy Williamson  
Research Fellow in Digital Humanities, University of Exeter  
@earlymodernpost

