



**Staatsbibliothek
zu Berlin**
Preußischer Kulturbesitz

STAATSBIBLIOTHEK ZU BERLIN – PK

WS05 DIGITAL EDITIONS

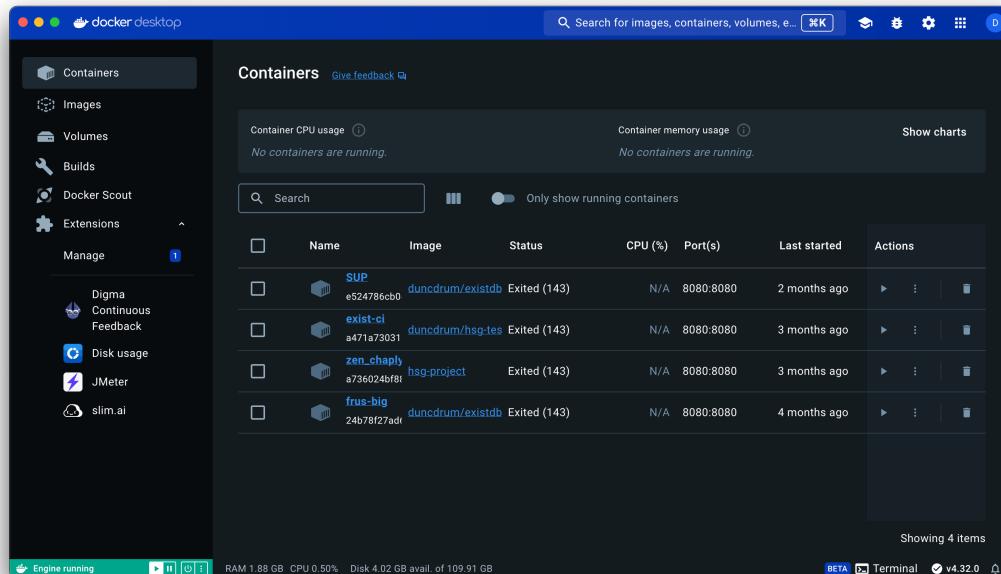
Duncan Paterson

- Duncan Paterson ([duncdrum](#))
 - 2022.11 – Now Subject Librarian Taiwan, Hong Kong, China at State Library Berlin
 - Until 2022 recent academic projects in Freiburg (READCHINA) and Heidelberg (HRA)
 - Freelancing Research Software Engineer 10+ years, FOSS contributions to Exist-db , TEI Guidelines, ...
 - Joined Jinntec in 2022.03
- Juri Leino ([line-o](#))
 - Member of the community group developing Xpath, Xquery and XSLT 4
 - Creator of several tools and libraries for exist-db
 - Exist-db core developer since 5 years
 - Working with Jinntec since 4 years
 - Working with exist solutions GmbH since 7 years
 - Freelance Software Engineer from 2003 until todayLars

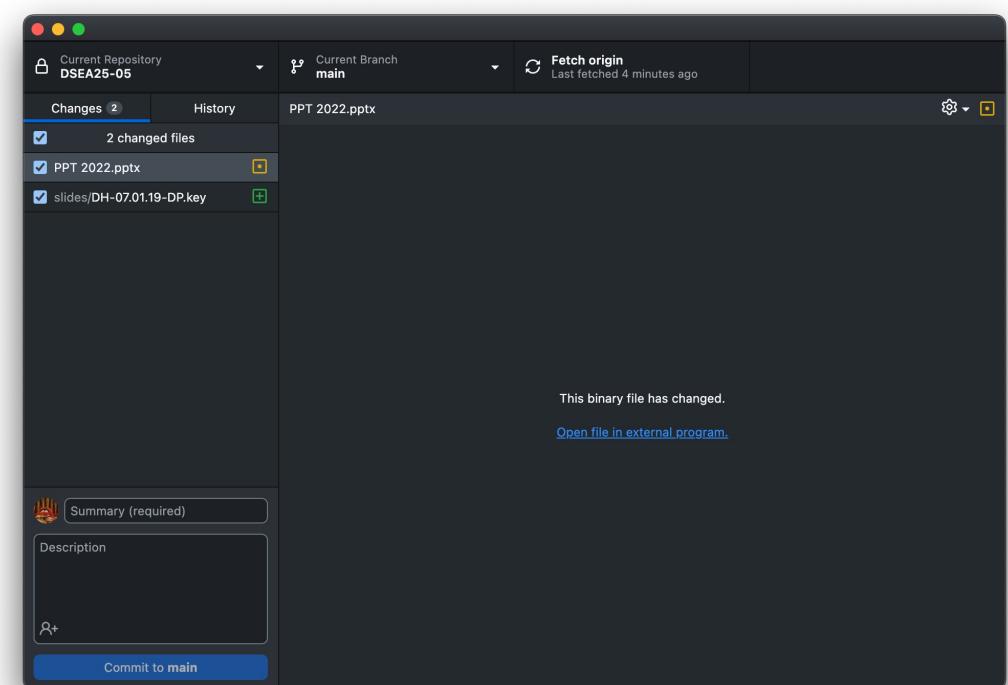


1. Introduction	09:00 – 10:00
1. Data Modelling and structures	
2. Types of editions	
3. Round Table: Plan your edition	
2. Hands On	10:00 – 10:30
1. Basic xml, first tei-xml	
COFFE BREAK	10:30 – 11:00
2. Xpath	11:00 – 11:30
3. TEI-publisher	11:30 – 12:30
1. Importing documents	
2. ODD editor	
3. Annotation editor	
4. App generator	
LUNCH BREAK	12:30 – 14:00

Docker



GitHub



Docker

```
docker pull ghcr.io/eeditiones/tei-publisher-app/teipublisher:master
```

```
docker run -dit -p 8080:8080 --name tp9 ghcr.io/eeditiones/tei-publisher-app/teipublisher:master
```

Shared folder

- <https://wg-cloud.gnz.mpg.de/s/djEyQyReqo8NNzg>

GitHub

- https://github.com/StaatsbibliothekBerlin/D-Sea24_WS05

Goals

- Understand the importance of modelling in Digital Humanities
- Understand the importance of standards in the context of the FAIR principles
- Gain a basic understanding of linked open data
- Understand modelling texts in TEI

According to McCarty (2005: 24):

- **Model:** either a representation of something for purposes of study, or a design for realising something new.
 - Simplified, idealised representation of real objects
- **Modeling:** the heuristic process of constructing and manipulating models.

Classification in Flanders and Jannidis (2015):

- **Research-driven modelling:** to express specific research ideas in cases where data is being created to support the creator's own research needs
- **Curation-driven modelling:** to serve as an interchange format for some types of users and user communities where data is typically being created and modeled with someone else's needs in mind

- **Model of vs. model for** (McCarty 2005)
- **Modeling for production vs. modeling for understanding** (Eide 2014)

Why modelling?	Using a model	Creating a model
Modelling for production	Using TEI to encode and publish a document	Creating a module in the TEI Guidelines
Modelling for understanding	Using TEI encoding as a research method	Creating a module in the TEI Guidelines

- Modelling as the key methodological structure of Digital Humanities
- As computers depend on models to operate, we need models for:
 - Automatization of tasks during development
 - Hypothesis testing
 - Analysis

- According to (Pierazzo 2015)
 1. Organisation of entities such as texts, documents and works along with their relationships and how they were produced.
 2. Definition of the type and purpose for the production of a new edition, its implied community of users and the features best represent their various needs.

From data to theory: Wallis and Nelson (2011) and the **3A** perspective:

- **Annotation** consists of the application of a scheme to texts. Annotations may include structural markup, part-of-speech tagging, parsing, and numerous other representations.
- **Abstraction** consists of the translation (mapping) of terms in the scheme to terms in a theoretically motivated model or dataset. Abstraction typically includes linguist-directed search but may include e.g., rule-learning for parsers.
- **Analysis** consists of statistically probing, manipulating and generalising from the dataset. Analysis might include statistical evaluations, optimisation of rule-bases or knowledge discovery methods.

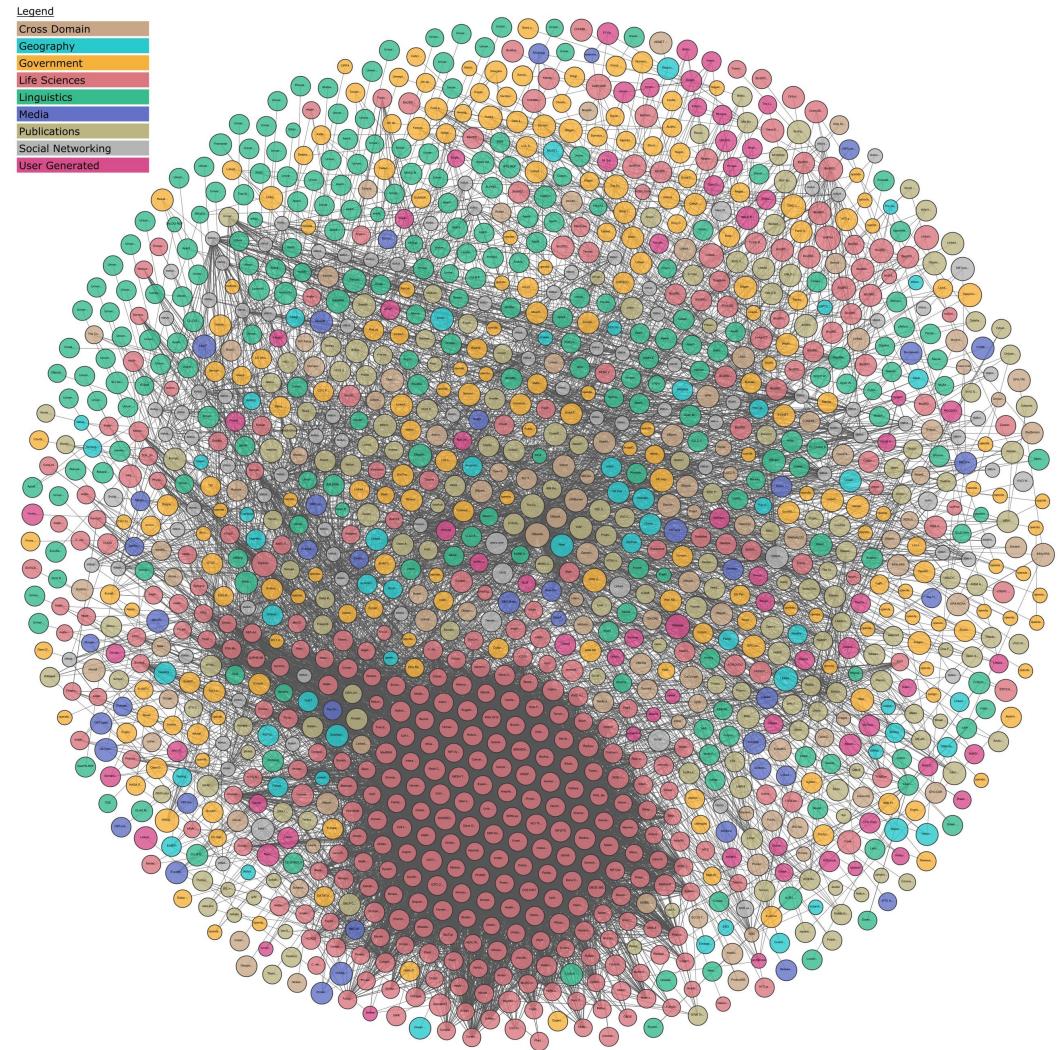
Goals

- Understand the importance of modelling in Digital Humanities
- Understand the importance of standards in the context of the FAIR principles
- Gain a basic understanding of linked open data
- Understand modelling texts in TEI

- Data standardisation vs data normalisation
- Standard: codified rules and guidelines for the creation, description and management of digital resources (Gill and Miller, 2002)
- Advantages:
 - Reusability
 - Interoperability
 - Long-term preservation (format/schema update)



- An approach to structure data representation that manifests as a Web of Data
- It enables:
 - Transparent access to data
 - Data integration across disparate data silos
 - Platform independent data connectivity



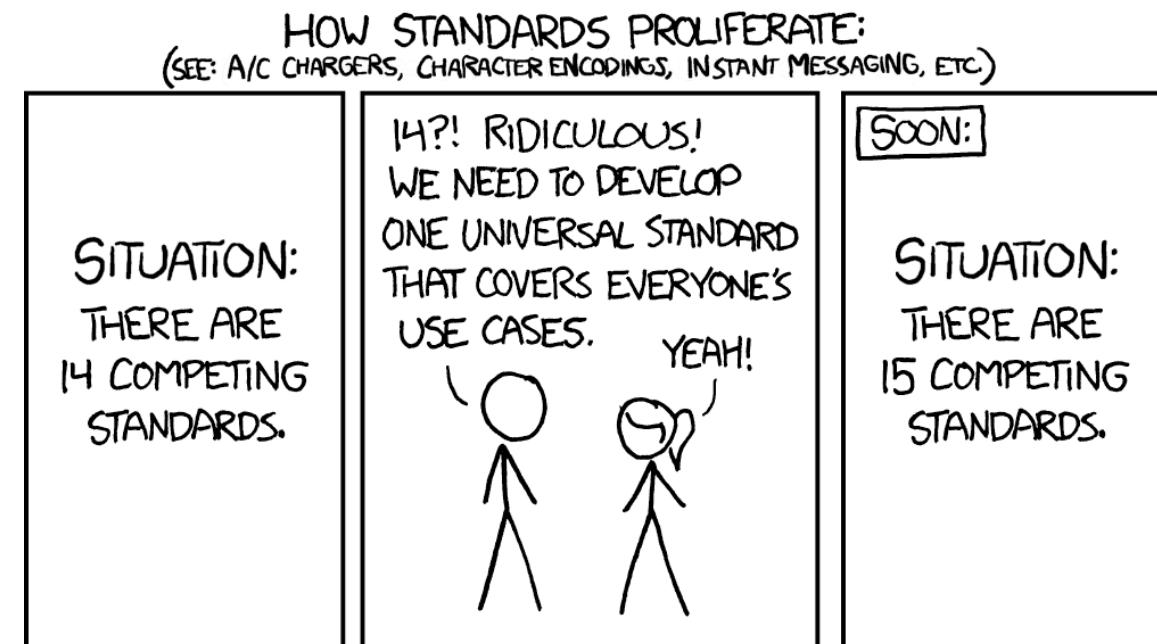
The Linked Open Data Cloud from lod-cloud.net



Extremely domain-specific and based on research questions

Tips:

- Look at what your peers are doing
- Look at what other disciplines are doing
- Don't write your own



Source: [XKCD](#)

Goals

- Understand the importance of modelling in Digital Humanities
- Understand the importance of standards in the context of the FAIR principles
- Gain a basic understanding of linked open data
- Understand modelling texts in TEI

Text Encoding Initiative

- Set of guidelines to encode cultural documents
- International Consortium that maintains and develops said guidelines
- Community of projects and researchers that implement the TEI guidelines



XML

TEI

Concepts



Syntax

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

Language: vocabulary and grammar

```
<p>
<note type="foot">
<head>
```

[Bauman & Flanders \(2007\)](#)

- Definition of vocabulary
 - Formalization of restrictions
 - Design of output
- 
- TEI customization

[Bauman & Flanders \(2007\)](#)

 <Text Encoding Initiative>

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines — English  Search

TEI: Guidelines for Electronic Text Encoding and Interchange

P5 Version 4.7.0. Last updated on 16th November 2023, revision e5dd73ed0

[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](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- Appendix F [Bibliography](#)
- Appendix G [Deprecations](#)
- Appendix H [Prefatory Notes](#)
- Appendix I [Colophon](#)

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](#)
- 11 [Representation of Primary Sources](#)
- 12 [Critical Apparatus](#)
- 13 [Names, Dates, People, and Places](#)
- 14 [Tables, Formulae, Graphics, and Notated Music](#)
- 15 [Language Corpora](#)
- 16 [Linking, Segmentation, and Alignment](#)
- 17 [Simple Analytic Mechanisms](#)
- 18 [Feature Structures](#)
- 19 [Graphs, Networks, and Trees](#)
- 20 [Non-hierarchical Structures](#)
- 21 [Certainty, Precision, and Responsibility](#)
- 22 [Documentation Elements](#)
- 23 [Using the TEI](#)

TEI sourcecode

- [Getting and Using the TEI Sources.](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

Goals

- Gain a basic understanding of the different types of digital scholarly editions (DSE)

- Documentary edition (focus on transcribing)
 - Diplomatic (exact reproduction)
 - Normalized (e.g. punctuated)
- Facsimile editions (faithful image reproduction)
- Annotated Edition (focus on commentary and notes)
- Genetic Editions (focus on composition and development)
 - Make authorial revisions and annotations accessible
- Critical Edition
 - Apparatus criticus for textual variants and editorial interventions
 - Includes multiple witnesses of a text
 - Scholarly commentary

- Introduce your edition project
- What is the purpose of your edition?
- Which textual features do you intend to model?
- How do you plan on modelling them?

Eide, Øyvind. 2014. “Ontologies, Data Modeling, and TEI.” Journal of the Text Encoding Initiative, Issue 8 (December). doi:10.4000/jtei.1191.

Flanders, Julia, and Fotis Jannidis. 2015. “Knowledge Organization and Data Modeling in the Humanities.” [Link].

Gill, Tony, y Paul Miller. 2002. “Re-Inventing the Wheel? Standards, Interoperability and Digital Cultural Content”. D-Lib Magazine 8 (1). doi:10.1045/january2002-gill.

McCarty, Willard. 2005. Humanities Computing. Basingstoke: Palgrave Macmillan. doi:10.1057/9780230504219

Wallis, S., and G. Nelson. 2001. “Knowledge discovery in grammatically analysed corpora”. Data Mining and Knowledge Discovery, 5: 307–340.

Pierazzo, Elena. 2015. Digital Scholarly Editing: Theories, Models and Methods. London ; New York: Routledge.

Goals

- Understand the basic structure of XML
- Understand the difference between well-formedness and validity
- Understand the basic structure of TEI documents
- Encode a text in TEI

Create a first document with one "root" element

- Open eXide
- Create a new XML document in eXide
- Press "New" in the top row of buttons
- then "Create" in the following dialog

<root/>

- Root node
- Unary tag
- Element

Can also be written
as:

<root></root>

- Empty tag

Now create a new element "other"

```
<root>  
    <other />  
</root>
```

- Well-formed
- open tag
- Closing tag
- Child element
- Parent element

Add an attribute "id" to the "other" element with any value

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

```
<root>
    <other id="special"/>
</root>
```

- attributes are children of the element _and_ its parent elements

- Add an element "child" to the root node.
- Add an element "other" as `_its_ child` and set its text content to "test".

```
<root>
    <other id="special"/>
    <child>
        <other>test</other>
    </child>
</root>
```

- Save this document in eXide

- File > Save > my-doc.xml

- The X in XML stands for extensible.
- Namespaces can be used to extend existing encodings with new elements and to validate documents against a schema
- `xmlns=http://www.tei-c.org/ns/1.0`

Well-formed:

- compliant with XML rules

Valid:

- Correct use of the vocabulary: elements and attributes are available in that language
- Correct use of grammar: elements are used in the proper place following the defined order

samples/tei_empty.xml

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```

2.4 encode a poem

`samples/wuyeti.txt`

鳥夜啼

黃云城邊烏欲栖
歸飛啞啞枝上啼
机中織錦秦川女
碧紗如煙隔窗語
停梭悵然憶遠人
獨宿孤房淚如雨

2.3 basic TEI document

samples/wuyeti.txt

烏夜啼

黃云城邊烏欲栖
歸飛啞啞枝上啼
机中織錦秦川女
碧紗如煙隔窗語
停梭悵然憶遠人
獨宿孤房淚如雨

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>烏夜啼 tei example</title>
        <author>李白</author>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body xml:lang="zh">
      <div type="poem">
        <head>烏夜啼</head>
        <lg>
          <l>黃云城邊烏欲栖</l>
          <l>歸飛啞啞枝上啼</l>
          <l>机中織錦秦川女</l>
          <l>碧紗如煙隔窗語</l>
          <l>停梭悵然憶遠人</l>
          <l>獨宿孤房淚如雨</l>
        </lg>
      </div>
    </body>
  </text>
</TEI>
```

2.3 a different poem

samples/zhanzheng_jiaoxiangqu.txt

2.3 basic TEI document

samples/zhanzheng_jiaoxiangqu.txt

BREAK TIME (30min)

Goals

- Learn how to select an xml element using xpath
- Learn how to navigate within an xml document

- - / selects the _children_ of a node
- - // selects the _descendants_ of a node
- - . represents the _current_ node
- - .. represents the _parent_ node
- - <a>/b will select the b element inside the a node

- Create a new XQuery document in eXide
- Using the "New XQuery"-button
- copy and paste the contents of your first document in your new XQuery file
- either after `xquery version "3.1";` or replace the contents entirely

- select the "other" element that is the direct child of "root"

■

Hint: use /

```
<root>
  <other id="special"/>
  <child>
    <other>test</other>
  </child>
</root>
/other
```

- Select all descendant "other" elements of "root"
- Hint: use //

2.2 Xpath Exercise 8

```
<root>
  <other id="special"/>
  <child>
    <other>test</other>
  </child>
</root>
//other
```

- Select all text inside the root element
 - – with `text()` you can select the textual contents of an element

2.2 Xpath Exercise 8

```
<root>
  <other id="special"/>
  <child>
    <other>test</other>
  </child>
</root>
/text()
```

- Select the id attribute of "other" elements



Hint: use // with / and @id

2.2 Xpath Exercise 8

```
<root>
  <other id="special"/>
  <child>
    <other>test</other>
  </child>
</root>
//other/@id
```

- Predicates are in square brackets following a selection or node (set)
- <a><c/> [CONDITION]
- /b [CONDITION]



- Select all "other" elements with an id-attribute "root"
-
- Hint: use @id in a predicate

2.2 Xpath Exercise 8

```
<root>
  <other id="special"/>
  <child>
    <other>test</other>
  </child>
</root>
//other/[@id]
```

- Select the parent of the "other" element without an id-attribute
- Hint: use `not()` in a predicate and `./.` to select the parent

2.2 Xpath Exercise 11

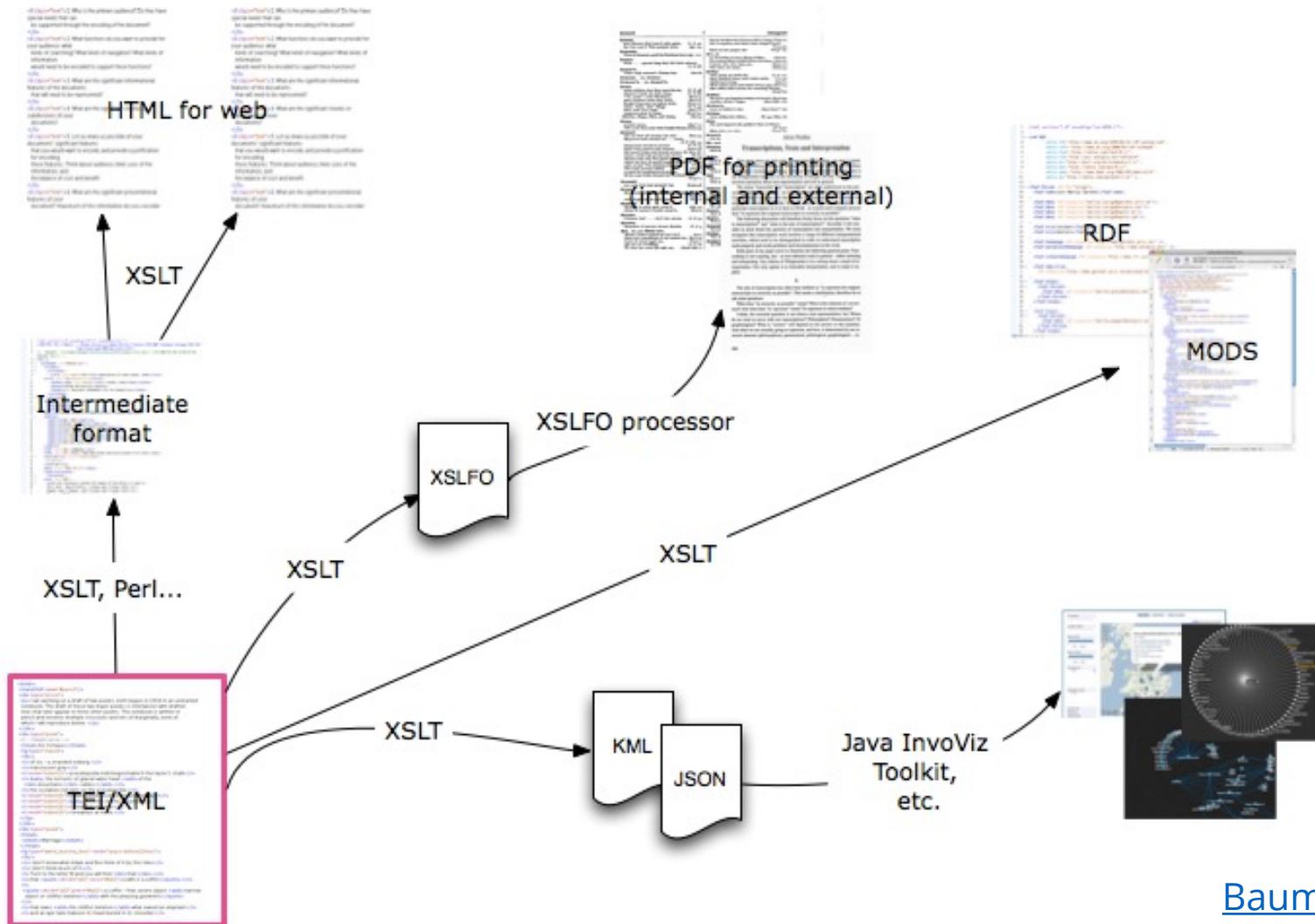
```
<root>
  <other id="special"/>
  <child>
    <other>test</other>
  </child>
</root>
//other/[not(@id)]/..
```

- Loading a document and selecting all other elements
- Hint: use `doc("/db/my-doc.xml")` to load the document we created earlier

2.2 Xpath Exercise 13

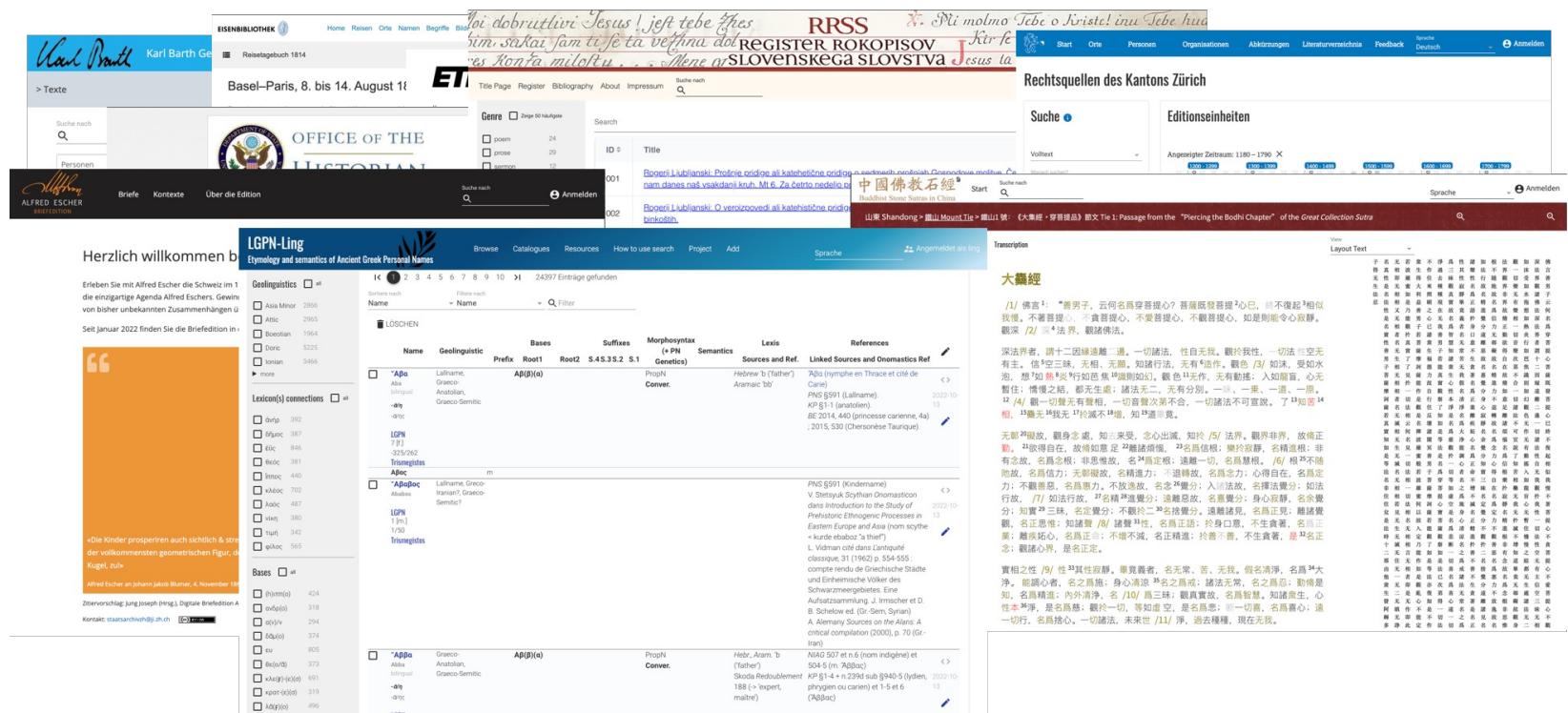
```
doc("/db/my-doc.xml")//other
```

1.1 TEI Single Source



Bauman & Flanders (2007)

An edition of medieval manuscripts differs from a genetic or epigraphic edition, a correspondence edition from a dictionary... but they all share ~80% of functional requirements



The collage includes the following components:

- Eisenbibliothek:** A screenshot showing a search interface for "Reisebuch 1814" from Basel-Paris, dated August 1811.
- Office of the Lüstodian:** A screenshot of a digital edition interface for Alfred Escher's *Briefe*.
- LGPN-Ling:** A detailed screenshot of the LGPN-Ling database, showing results for "Geolinguistics" and "Lexicon(s) connections". It displays tables for Name, Geolinguistic, Prefix, Bases, Suffixes, Morphosyntax (+ PN Genetics), Semantics, Lexis, Sources and Ref., Linked Sources and Onomastics Ref., and References. Specific entries for "Alba" and "Alloc" are shown.
- REGISTER ROKOPISOV:** A screenshot of the Register Rokopisov, showing a title page in Slovenskega Slovstva.
- Kirle:** A screenshot of the Kirle platform, showing a search interface for "Rechtsquellen des Kantons Zürich" (Legal Sources of the Canton of Zurich) from 1180-1790.
- 中国佛教石经:** A screenshot of the Chinese Buddhist Stone Inscriptions database.
- ETI:** A screenshot of the ETI (Electronic Text Initiative) platform, showing a search interface for "Title Page", "Register", "Bibliography", "About", "Impressum", and "Suche nach".
- Rechtsquellen des Kantons Zürich:** A detailed screenshot of the legal sources database, showing a search interface for "Suche" and "Editionseinheiten" (Edition Units), with filters for "Volltext" and "Auszugster Zeitraum: 1180 – 1790".
- 其他:** A screenshot of a page titled "Herzlich willkommen bei der Staatlichen Bibliothek zu Berlin" (Welcome to the State Library of Berlin).

Starting Point

- Encoding to be guided by your research questions, not technical considerations
- „One size fits all“ – rarely works in the humanities ;-)
- Every project is unique, though technical requirements overlap
- Limits interoperability and long term preservation? TEI has an answer ...

Guiding Principles

- open source
- empowering the editors & fun for developers
- rooted in standards
- meant to last
- designed for sharing and reuse

Beyond Encoding

- Use Standards Where You Can
- Backend: Open API
- Frontend: Web Components
- Mix and match as you like

Docker

```
docker pull ghcr.io/eeditiones/tei-publisher-app/teipublisher:master
```

```
docker run -dit -p 8080:8080 --name tp9 ghcr.io/eeditiones/tei-publisher-app/teipublisher:master
```

Shared folder

- <https://wg-cloud.gnz.mpg.de/s/djEyQyReqo8NNzg>

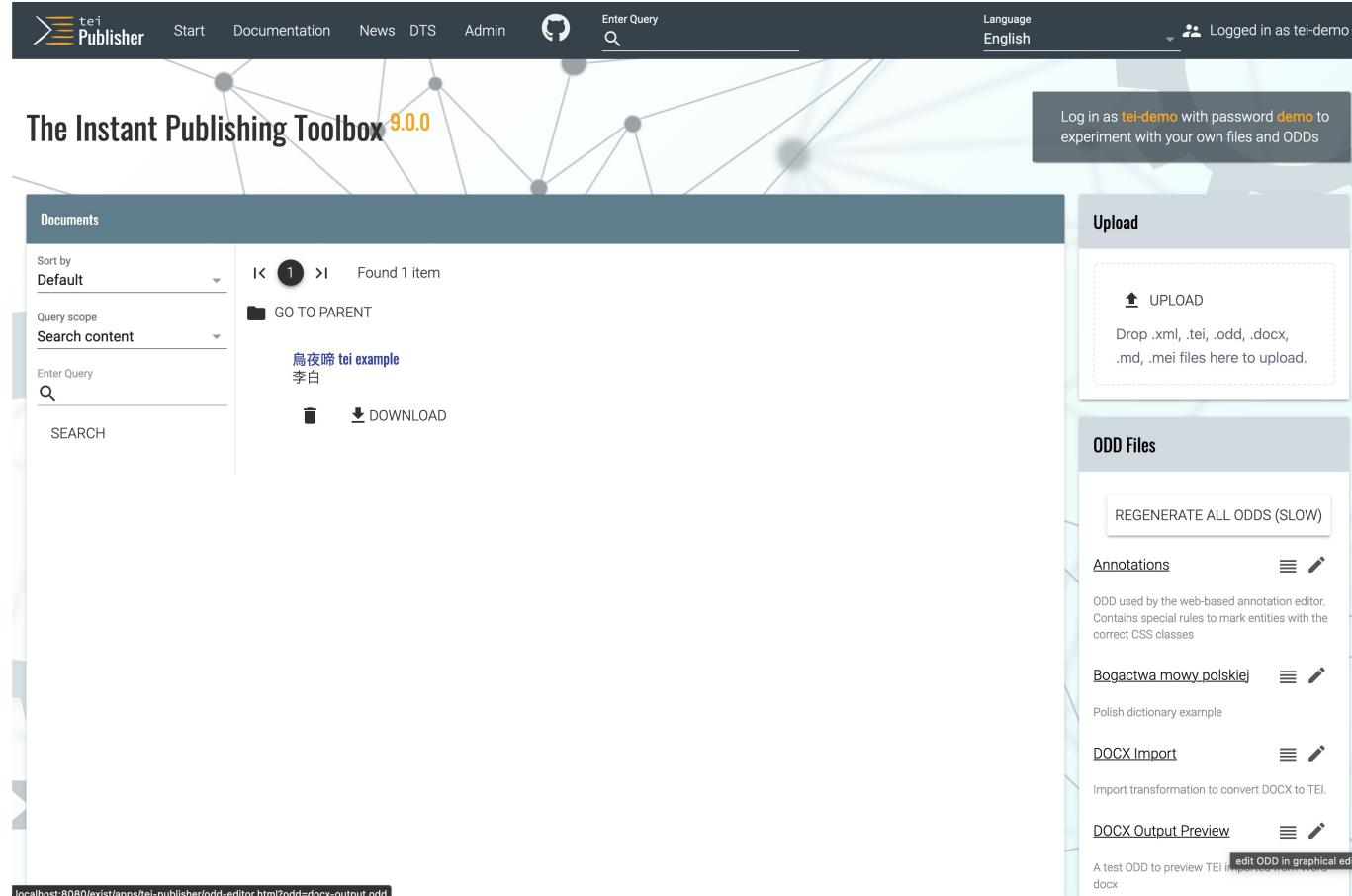
GitHub

- https://github.com/StaatsbibliothekBerlin/D-Sea24_WS05

Goals

- Import tei file (wuyeti.xml) into tei-publisher
- Create a new ODD
- Modify the display (vertical layout) in ODD editor
- Annotate Document via Annotation Editor
- Generate a standalone expat application

3.1 TP import



The Instant Publishing Toolbox 9.0.0

Documents

Sort by Default

Query scope Search content

Enter Query

SEARCH

Found 1 item

GO TO PARENT

烏夜啼 tei example
李白

DOWNLOAD

Upload

UPLOAD

Drop .xml, .tei, .odd, .docx, .md, .mei files here to upload.

ODD Files

REGENERATE ALL ODDS (SLOW)

Annotations

Bogactwa mowy polskiej

DOCX Import

DOCX Output Preview

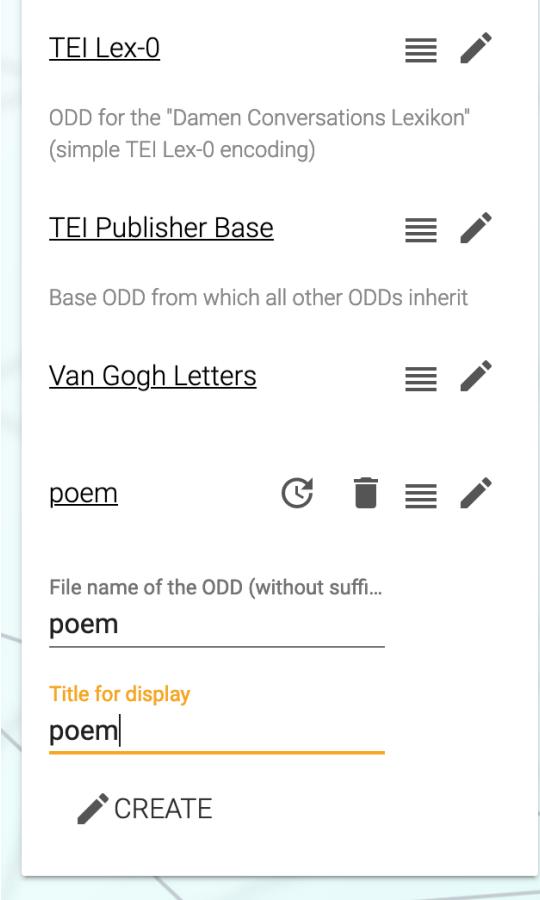
A test ODD to preview TEI imports from Microsoft Word docx

localhost:8080/exist/apps/tei-publisher/odd-editor.html?odd-docx-output.odd

Goals

- Import tei file (wuyeti.xml) into tei-publisher
- Create a new ODD
- Modify the display (vertical layout) in ODD editor
- Annotate Document via Annotation Editor
- Generate a standalone expat application

3.2 TP create ODD



TEI Lex-0 ≡ 

ODD for the "Damen Conversations Lexikon"
(simple TEI Lex-0 encoding)

TEI Publisher Base ≡ 

Base ODD from which all other ODDs inherit

Van Gogh Letters ≡ 

poem    

File name of the ODD (without suffi...
poem

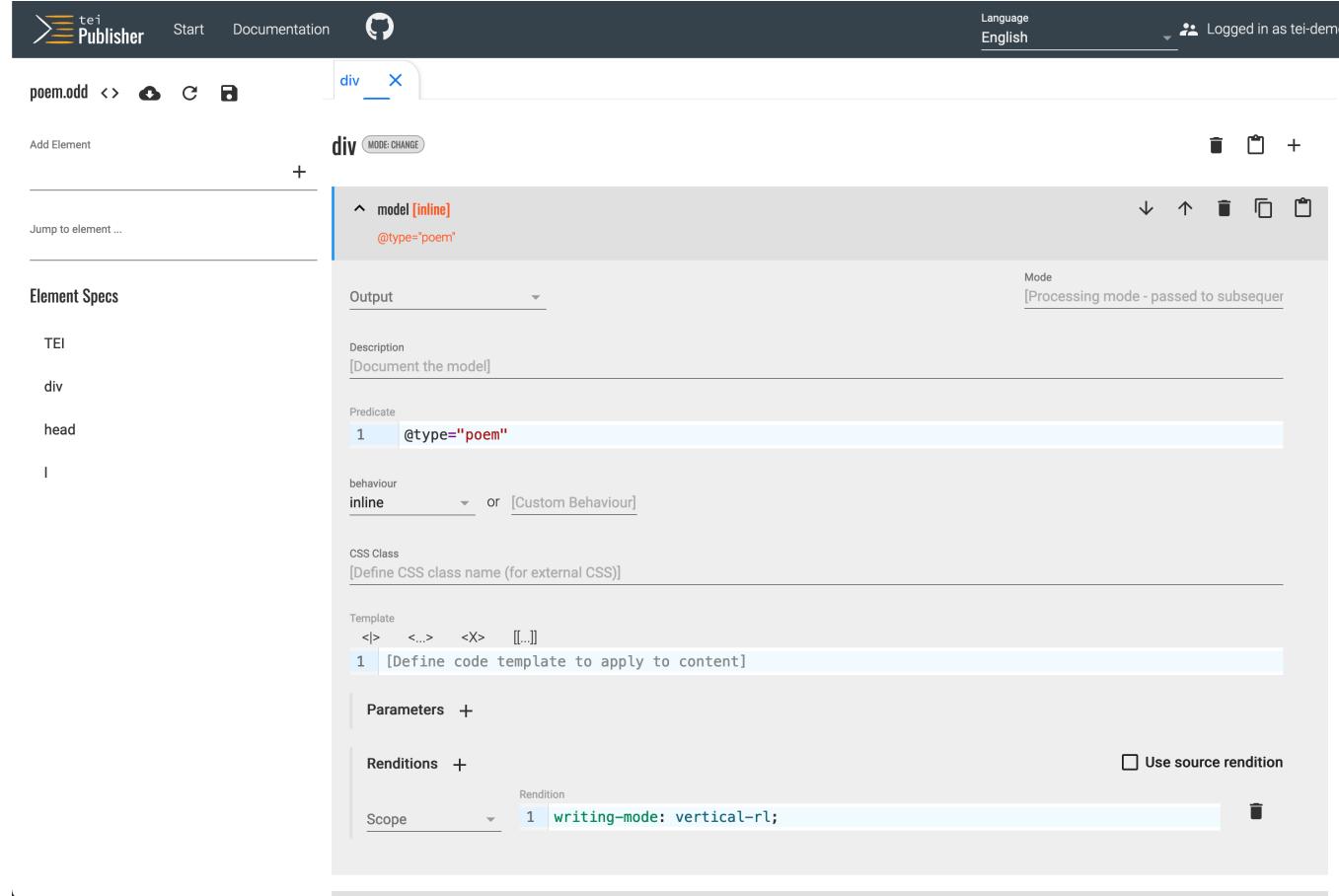
Title for display
poem

 CREATE

Goals

- Import tei file (wuyeti.xml) into tei-publisher
- Create a new ODD
- Modify the display (vertical layout) in ODD editor
- Annotate Document via Annotation Editor
- Generate a standalone expat application

3.3 tei-publisher



The screenshot shows the tei-publisher application interface. At the top, there is a navigation bar with links for 'Start' and 'Documentation', and a GitHub icon. On the right side of the top bar, it says 'Logged in as tei-demo'. Below the navigation bar, the main workspace displays a 'div' element spec for 'poem' type.

Add Element: A dropdown menu with a '+' button and a 'Jump to element ...' input field.

Element Specs: A sidebar listing various element specifications: TEI, div, head, and others.

div [Mode CHANGE]: The selected element spec. It includes:

- model [inline]**: Red text indicating the current mode.
- @type="poem"**: A predicate entry.
- Output**: Mode: Processing mode - passed to subsequent.
- Description**: Document the model.
- Predicate**: 1 @type="poem".
- behaviour**: inline or Custom Behaviour.
- CSS Class**: Define CSS class name (for external CSS).
- Template**: A template with placeholder code: <|> <...> <X> [[..]].
- Parameters**: +
- Renditions**: +
- Scope**: Rendition 1 writing-mode: vertical-rl;
- Use source rendition**: A checkbox.

Goals

- Import tei file (wuyeti.xml) into tei-publisher
- Create a new ODD
- Modify the display (vertical layout) in ODD editor
- Annotate Document via Annotation Editor
- Generate a standalone expat application

Friedrich Leibnütz (Wikipedia sample)

Friedrich Leibnütz

Friedrich wurde als Sohn des Stadt- und Bergschreibers in Altenberg Ambrosius Leibnütz (* 14. April 1569 in Berggießhübel; † 28. Mai 1617 in [Altenberg](#)) und dessen am 23. August 1596 in Königstein geheirateten Frau Anna Deuerlin (* 14. April 1617 in Altenberg) geboren. Sein Großvater Christoph Leibnütz (* 10. September 1537 in Rochlitz; † 28. Mai 1587 in Pirna) hatte als Land- und Erbrichter unter Johann Wilhelm von Sachsen-Weimar gedient, wurde 1576 vom Kurfürst August von Sachsen als Schösser in Pirna eingesetzt und hatte am 27. Januar 1578 in Dresden seine Frau Barbara von Kahlenburg (* 11. Dezember 1539 in Kalundborg; † 11. Juli 1577 in Pirna) geheiratet, die aus einem adligen Geschlecht in Jütland stammte.

Nach anfänglichem Schulbesuch frequentierte Friedrich ab 1612 die kurfürstlich sächsische Landesschule St. Afra in Meißen. Nach sechs Jahren zog er nach [Leipzig](#) [place](#), wo er sich zunächst als Hauslehrer der Kinder des Oberhofgerichtsnotars Stephan Hoffmann seinen Unterhalt verdiente. Am 21. Oktober 1618 begann er an der Universität [Leipzig](#) [place](#) ein Studium der philosophischen und juristischen Wissenschaften. Hier wurde der damalige Professor der Ethik und Notar der Universität Leipzig Johann Müller (* in Frohburg; † 1628 in Leipzig [place](#)) sein führender Lehrer. Am 6. März 1619 erwarb er sich das Baccalaureat und wurde am 30. Januar 1622 Magister der Philosophie. Zuvor wurde er am 19. März 1621 als Adjunkt an die philosophische Fakultät angestellt, am 20. Oktober 1624 wurde er Notar der Universität [Leipzig](#) [organization](#) und am 26. November 1628 Professor für [Moralphilosophie](#) [term](#).

Annotation Details

Reference

Search/edit reference

Search entity

Q Altenberg +

- Altenberg local place
- Germany, Saxony
- Altenberg geonames.org place
- fourth-order administrative division - Saxony, Germany
- Altenberg geonames.org place
- hill - Hesse, Germany
- Altenberg geonames.org place
- hill - Hesse, Germany
- Altenberg geonames.org place
- hill - Hesse, Germany
- Altenberg bei Leibnitz geonames.org place
- populated place - Styria, Austria
- Altenberg an der Rax geonames.org place
- historical third-order administrative division - Styria, Austria
- Geising geonames.org place
- populated place - Saxony, Germany
- Altenberg bei Linz geonames.org place

Goals

- Import tei file (wuyeti.xml) into tei-publisher
- Create a new ODD
- Modify the display (vertical layout) in ODD editor
- Annotate Document via Annotation Editor
- Generate a standalone expat application

tei Publisher Start Documentation 

Language English  Logged in as tei-demo

TEI Lex-0
 TEI Publisher Base
 Van Gogh Letters
 poem

URL to uniquely identify the app
<https://e-editiones.org/apps/my-simple-app>

Short name to be used in URLs and file names
Short name

Name of subcollection to hold TEI documents
data

Title of the app which will be shown e.g. in the dashboard
Title of the app

Choose the HTML template to be used as default
HTML Template
Default single text layout

Choose what is shown by default when browsing text: a single page or an entire division. Display by page requires that the TEI is properly marked up with <tei:pb> tags.

Default View
By division (chapter/section...)

Define the smallest block on which a full text index is created. For documents organized into divisions, choosing 'Create on division' is usually best. If there are no divisions, choose 'Create on text'.

Default Full Text Index
Create on division

User account for administrative tasks. The user will be created if it does not yet exist.
The user account who will own the app.
User

Password for the user owning this app.
Password



Links

- <https://readchina.github.io/shouchaoben/>
- <https://www.cbeta.org/>
- <https://www.kanripo.org/>
- <https://www.stonesutras.org/home.html> (tei-publisher)
- <https://www.collexpersee.eu/projet/chi-know-po-corpus/> (WIP)
- http://dev.dila.edu.tw/Dunhuang_Manuscripts/
<https://zenodo.org/records/1133490>