



ZENTRUM FÜR
INFORMATIONSMODELLIERUNG
AUSTRIAN CENTRE FOR
DIGITAL HUMANITIES

KARL-FRANZENS-UNIVERSITÄT GRAZ
UNIVERSITY OF GRAZ



Beyond TEI

Digital Editions with XPath & XSLT for the Web & in \LaTeX



Sarah Lang
Harvard, April/May 2022



Overview

- 1. The workshop
- 2. Motivation
- 3. Digital Scholarly Editing
- 4. Digital Archives

The workshop

Goals

1. get to know XPath & XSLT (and learn how to use it)
2. understand the role of XML/TEI, XPath and XSLT in Digital Editing
3. be able to use XSLT to generate HTML and \LaTeX output from TEI
4. Two days isn't enough for you to master XSLT!

Schedule

Day 1, morning XML, TEI and Digital Editing → repetition of the basics, making sure we're all on the same page, understanding why we're even learning XSLT.

Day 1, afternoon Navigating XML documents using XPath, introduction to HTML (& Bootstrap) and \LaTeX (& *reledmac*)

Day 2 Transforming XML documents into HTML & \LaTeX output formats using XSLT

Single point of entry for all workshop-related materials: [\TeX Ninja blogpost](#) & [Github Repository](#) ('additional resources' directory)

Introductions

Please introduce yourselves!

1. Name, pronouns, field/topic of study
2. Why did you come to this workshop?
3. Previous experience with Digital Humanities (DH) or editing?

Contact

- Twitter: @SarahALang_
@latex_ninja
- Website: sarahalang.com
latex-ninja.com
- Email: [sarah.lang@uni-graz.
at](mailto:sarah.lang@uni-graz.at)

Sarah Lang (she/they)

- originally from Germany, now in Graz (Austria)
- Studied Latin, French & History (teacher's education) in Graz & Montpellier (France), then Archaeology Bachelor, Master in Religious Studies & Philosophy
- got a DH certificate & started working at Zentrum für Informationsmodellierung (ZIM) / Centre for Information Modelling in Graz
 - Moral Weeklies/Spectators → gams.uni-graz.at/mws
 - Graz Repository of Ancient Fables (GRaF) → gams.uni-graz.at/graf
 - PhD thesis: Decoding alchemical *Decknamen* digitally. A Polysemantic Annotation and Machine Reasoning Algorithm for the Corpus of Iatrochymist Michael Maier (1568–1622)
- Now: teaching in Graz, Passau & Vienna; PostDoc in Graz. Research interests: history of science (alchemy), Neo-Latin, text mining and computer vision

Motivation

TEI, now what?

Why are we doing this workshop? The motivation from our abstract:

- [...] the Text Encoding Initiative (TEI) for XML has become the gold standard for scholarly editions of texts.
- But what happens after an edition is encoded in TEI?
- While it is an **ideal format for archiving digital data**, it is **less than ideal for viewing and interacting with the edited text**.

Goals for the next session

1. understand the terms from the abstract:
 - ✖ digital edition
 - ✖ single source principle
 - ✖ ~~creating different representations from our data~~ → (why) do we even need to create new presentations from our data? Aren't there tools for that?

Teaser i

GAMS ufas

Urfehdebücher der Stadt Basel – digitale Edition

Universität Basel

Ratsbücher O10, Urfehdebuch X
(1563-1569)

Ziterverschlag:
Urfehdebuch X der Stadt Basel (1563-1569), in: Die Urfehdebücher der Stadt Basel – digitale Edition, hg. v. Susanna Burghartz, Sonja Calm und Georg Vogeler Basel/Graz 2016. (Gelekt verändert am 31.1.2017);
hd:11471/1010.2.1.

Zeitraum:
03.01.1563 bis 04.12.1569

fol. 1v

[Aeno d[em]ni 1563]

Beginn Undatiert 1563 1564 1565 1566 1567 1568 1569 Ende

Home Projekt Urfehdebuch Kategorien Datenkorb (0) Erweiterte Suche Volltextsuche

Datum: 1564-02-25

Person

Bernier Dukler ein welischer von wegen das er sich mitt' win überleden, und einer denphale gesondhangt habt in usser der schweiz [Hof]tred gefangenschafft inn Toussuzia gefangen hofft auf den 25 tag Hornungs ussgelossen iuswalt ut moria

Tatbestand

abschleiden (Draengschafft)

Strafe

Orte

Schlagwort

zufolge/mische

Beginn Undatiert 1563 1564 1565 1566 1567 1568 1569 Ende

Teaser ii

CoReMA → Look at: <http://gams.uni-graz.at/corema>

This is the most recent version of this document. See change log

[269r]

Ein guet kuchenpueech
Fish as collared pork (Nr. I) [cooking
recipe] ¶

Wilwd machen ein schwein koppff (pig head)
von vischen (fish) So nym cherpfen (carp)¹ oder
schleyen (tench) oder perben (common barbel)
vnd schueb sye schon vnd schlach sy auf vnd
mach sy zw stuecken vnd thu sy In ein pfan
(pan) vnd gewß dar an halbs wasser (water)
halbs wein (wine) vnd daz dy pruee (water wine
broth) nuer ains vin gers hach gee uber dy visch

Search [] I ↑ X

Ingredient	Dish	Tool
Name	Ref.	

x = x

Description [] Recipes [] Transcription [] Full-Width [] Metadata []

Teaser iii

Furnace and Fugue (music in MEI, text and images)



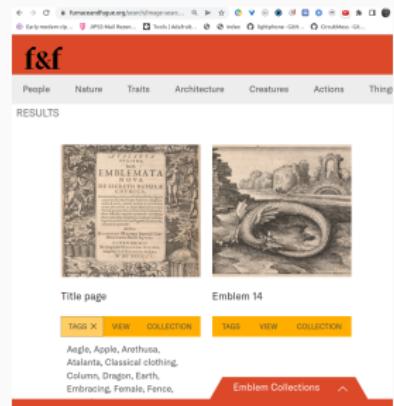
The screenshot shows a digital edition of a historical book. At the top, there's a header with navigation icons and the URL 'furnaceandfugue.org/stable/afugueinmei.html'. Below the header, the title 'EMBLEMATA L. De ferre Naturae' is visible, followed by the subtitle 'Portentum cum venientia in ventribus'. The main content area displays a woodcut-style illustration of a figure in a dynamic pose, possibly a deity or a personified nature figure, set against a background of trees and architectural structures. To the right of the image, there's a musical score in MEI format with three voices labeled 'Voice 1', 'Voice 2', and 'Voice 3'. Below the score, there's a section titled 'the first Epigram' with Latin text and its English translation. The Latin text reads:

Et subiectum pro x. t. e. p. / adiutori ab aliis
Fecit in hunc lemnos / et modice recte,
Uero invenit / quod non posset / invenire
Aer, manus, fore corporis, manus, pectus,
Ingenium, caput, fore ab aliis invenire,
Pecunia, dicitur, quod non posset / invenire.

The English translation is:

If heroes once expand his windy womb,
That the infant in the world alive may come,
His prowess will the Heroes all outdo
In art, in hand, in mind, and body too.
Think him no Cacus, nor abortive brood:
Nor yet Agrippa, for his star was good.

At the bottom of the page, there are search and zoom icons, and a footer with 'DISCUSSION' and 'Emblem Collections' buttons.



The screenshot shows a search results page for 'Emblems'. The header includes the f&f logo and links to 'People', 'Nature', 'Traits', 'Architecture', 'Creatures', 'Actions', and 'Thing'. Below the header, the word 'RESULTS' is displayed. Two emblem illustrations are shown side-by-side. The left one is titled 'Title page' and the right one is titled 'Emblem 14'. Each illustration has a 'TAGS' button, a 'VIEW' button, and a 'COLLECTION' button below it. The 'TAGS' for both are identical: 'Aegle, Apple, Arithusa, Alisma, Classical clothing, Column, Dragon, Earth, Embracing, Female, Fence.' A 'Emblem Collections' button is located at the bottom right of the results area.

From TEI to Edition

What did all those examples have in common?

- all are digital editions
- all are based on data in TEI-XML

That's what we want. – How do we get there?

Remember

- XML is great for capturing & long-term archiving data in its whole complexity
 - ...but also ends up chaotic with a lot of detail
 - theoretically human-readable
 - in practice, it's better to create a new representation of the data to read
- **Desired result:** view/presentation, visualization and interaction with the data as a website (i.e. HTML data)
- → we need to transform the data

TEI Publication Tools

There is a plethora of out-of-the-box solutions to make editions out of TEI data. → [TEI List of Resources for Publishing and Delivery Tools](#), for example:

1. teiPublisher
2. EVT
3. TEICHI
4. Versioning Machine
5. Boilerplate
6. Oxgarage
7. GAMS ...

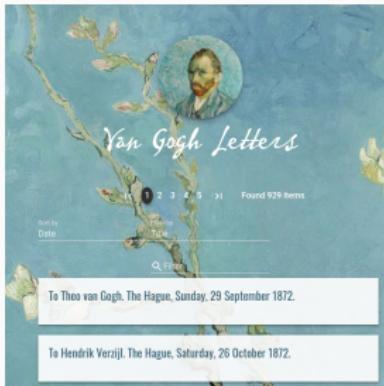
TEI Publisher: The Instant Publishing Toolbox

- <https://teipublisher.com/index.html>
- https://wiki.tei-c.org/index.php/TEI_Publisher
- <https://wiki.tei-c.org/index.php/TeiPublisher>
- <http://hisoma.huma-num.fr/exist/apps/tei-publisher/index.html>
- “Testimonial”: Florian Chiffolleau (2020), Blog Post: *Publication of my digital edition – Working with TEI Publisher.*



TEI Publisher: The Instant Publishing Toolbox ii

Demos: Van Gogh Letters ● Early English Books Online (EEBO)



This screenshot shows a detailed view of a letter from Theo van Gogh to his father. The header information includes the recipient ("To Theo van Gogh, The Hague, Sunday, 29 September 1872."), the date ("September 1872"), and the location ("The Hague, 29 September 1872"). The letter text is in Dutch:

[Wandel] Theo,
My dear Theo,
Dank voor je brief,
het deed mij
goeden dat je
weer goed
aangekomen zijt.
Ik heb je de eerste
dagen ge mist &
het was mij
vreemd je niet te
vinden als ik mind
dags thuis kwam.
Wij hebben
pechte dagen in

Notes:

- a Expression meaning 'to make the most of an opportunity'. In the context of this letter, it could also be meant literally.
- 1 Theo attended secondary school in Oosterwijk in the province of North Brabant. He walked the 6 km to school from his parents' house



Arcus recessum loco est Hierophatica Agrippa Graecis ergo modis cogit, ad Eumeniendas fabiorum apud tempora desumptas, deinceps latentes, et invenientes per suos instrumentos, augures, ex via figurata, et invenientia, et novellis amissis & eis invenientibus, propositis, deductis, vobis ut patet alegria, superem servitios fiducie & pretium encyclopediam eorum spuri electivis vertitis luce, auditoria, ratiocinio, et sensu, velut digna recompensae, ut illos expediti istud Michaelis Maioris Comitis Palii Caesaris, opere easius, Phil. & Med. Doc. Acc. Cesar. Mai. ipsorum inde.

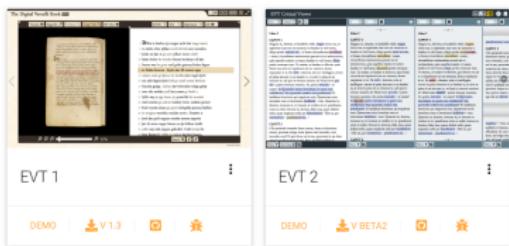
MICHAELIS MAIERI COMITIS PAL.
CAES. •RAE•ATIO AD LECTORES.

*S*i miremni lectori leversi, quid Case* milit exierit car buse tractu ARC ANISSIMORVM nomine insignitus post tot & tantos de Hierophatica, algorij, fabulis, theologicis, philosophicis, politicis, Chymicis & medicis perulgatos authores ediderim ac in publican mundi lucem nunc tandem sparserim, vobis que impensis de tot praeiuicatis ad dijjicendum inscriperim, postquam rem ipsam, mea studia ac mestrum, vestrum que canadore simul respserit, miris abe qae dubio doctiss. Inter Christianos, vt spinor, nullus est ad o' rerum omnium impensis, si modo sensu communicalat & prima literarum rudimentatenet, qui non aliisque libros Ethnikorum (quecumq; ingen

Edition Visualization Technology (EVT)

- <http://evt.labcd.unipi.it/> - <http://evt-project.sourceforge.net/>
- <https://visualizationtechnology.wordpress.com/>
- Roberto Rosselli Del Turco, Giancarlo Buomprisco, Chiara Di Pietro, Julia Kenny, Raffaele Masotti and Jacopo Pugliese: *Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions*, Journal of the Text Encoding Initiative 8 (2014/15): Selected Papers from the 2013 TEI Conference “TEI Processing: Workflows and Tools”; <https://doi.org/10.4000/jtei.1077>
- Del Turco, Roberto Rosselli. “Designing an Advanced Software Tool for Digital Scholarly Editions: The Inception and Development of EVT (Edition Visualization Technology).” *Textual Cultures*, vol. 12, no. 2, 2019, pp. 91–111. JSTOR, www.jstor.org/stable/26821538.

A light-weight, open source tool specifically designed to create digital editions from XML-encoded texts, freeing the scholar from the burden of web programming and enabling the final user to browse, explore and study digital editions by means of a user-friendly interface.



TeICHI – Bringing TEI Lite to Drupal

- TEICHI: a TEI lite integration into Drupal (<http://teichi.org>)
- Sebastian Pape, Christof Schöch and Lutz Wegner: *TEICHI and the Tools Paradox. Developing a Publishing Framework for Digital Editions*, Journal of the Text Encoding Initiative 2 (2012): Selected Papers from the 2010 TEI Conference; <https://doi.org/10.4000/jtei.432>
- TEICHI wiki

Accueil Introduction Essai sur le récit Dossier critique Aide

Essai sur le récit

:: Introduction
L'introduction présente Bérardier de Bataut et l'*Essai sur le récit* ainsi que la présente édition électronique.

:: L'Essai sur le récit
L'édition électronique annotée de l'*Essai sur le récit* de Bérardier de Bataut permet l'affichage alternatif d'un texte de lecture ou d'une transcription linéaire.

:: Dossier critique
Le dossier critique donne des informations bio-bibliographiques et fournit la documentation éditoriale et technique du projet d'édition.

:: Accès à l'aide
L'aide fournit des renseignements sur les possibilités de navigation et les fonctionnalités du site.

Référence bibliographique
François-Joseph Bérardier de Bataut, *Essai sur le récit, ou Entretiens sur la manière de raconter* (Paris : Charles-Pierre Berton, 1776). Édition électronique sous la direction de Christof Schöch, 2010. URL : <http://www.berardier.org>. (Version 0.6, 12/2010.)

The image shows the title page of the 1776 edition of 'Essai sur le récit'. The title is 'ESSAI SUR LE RÉCIT, OU ENTRETIENS SUR LA MANIÈRE DE RACONTER'. Below the title, it says 'PAR M. F. J. BERARDIER DE BATAUT, Essai sur Théâtre, Récit de Notre-Dame de Jeunesse & autres Projets d'Éditions de l'Université de Paris.' There is also some smaller text at the bottom right.

The Versioning Machine

- Versioning Machine (<http://v-machine.org/>)
- The TEI Versioning Machine Versioning Machine (Duke)

The screenshot shows a comparison between two versions of the poem "Faith is a fine invention" by Emily Dickinson. The interface includes a logo for "V.M. 5.0", a title bar with the poem's name, and a toolbar with buttons for "7 Total Versions", "Line numbers", and "Bibliographic panel".

Bibliographic Information	Version a660: A 660, verse embedded in letter to	Version h201: H 201, fascicle version of poem.
<p>"Faith is a fine invention"</p> <p>by Emily Dickinson</p> <p>Original Source</p> <p>See Witness List.</p> <p>Witness List</p> <ul style="list-style-type: none">Witness a660: A 660, verse embedded in letter to Samuel Bowles.Witness h201: H 201, fascicle version of poem.Witness h72: H 72, fascicle version of poem.Witness p1891: Published as poem XXX in the second volume of Todd and Higginson's Poems of Emily Dickinson.Witness 11894: Letter to Samuel Bowles published in Todd's edition of Dickinson's letters.Witness cp32: Published as poem LVI in Martha Dickinson Bianchi's Complete Poems of Emily Dickinson.	<p>1 "Faith" is a fine invention 2 When Gentlemen can see - 3 But Microscopes are prudent 4 In an Emergency .</p>	<p>1 Faith is a fine invention 2 For Gentlemen who see - 3 But Microscopes are prudent 4 In an Emergency !</p>

TEI Boilerplate

- <http://teiboilerplate.org/>
- Boilerplate (<http://dcl.slis.indiana.edu/teibp/>)
- Demo
- TEI Boilerplate Wiki
- Slides on TEI Boilerplate
- Exercise on TEI Boilerplate
- Poster on TEI Boilerplate

TEI Boilerplate Examples & Features

Nested <div> elements and headings

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam gravida, urna eget venenatis mollis, ipsum tellus hendrerit tellus, sed vestibulum urna quam eget enim. Pellentesque odio lacus, laoreet quis commodo eu, vehicula congue purus.

Nested <div> elements and headings

Vivamus rhoncus erat mi. Vestibulum sapien sapien, pulvinar id vehicula in, porta tristique ligula. Sed commodo diam sit amet dui posuere luctus. Morbi tempus sem adipiscing nisl molestie pharetra.

Nested <div> elements and headings

Sed fermentum dui vel enim consectetur malesuada. Sed a mi felis. Curabitur sit amet mi augue, ac volutpat leo. Maecenas vitae condimentum orci. Morbi non quam sed neque tincidunt tempus posuere sit amet purus.

Nested <div> elements and headings

Nested <div> elements and headings

Nested <div> elements and headings



Oxgarage

- REST-based transformation web service for TEI documents (<https://oxgarage.tei-c.org/>)
 - allows you to transform a number of (markup-based) formats to TEI or create them from TEI (`.html`, `.tex`, `.docx`, ...)
 - Based on TEI base stylesheets (<http://www.tei-c.org/Tools/Stylesheets/>)
- ✓ very useful tool
- ✗ not all TEI elements taken into account, not customizable

Oxgarage

...offers a set of standard stylesheets to convert between TEI-encoded XML documents and other documents, mostly markup-based, such as `.docx`, `.html` or `.tex`.

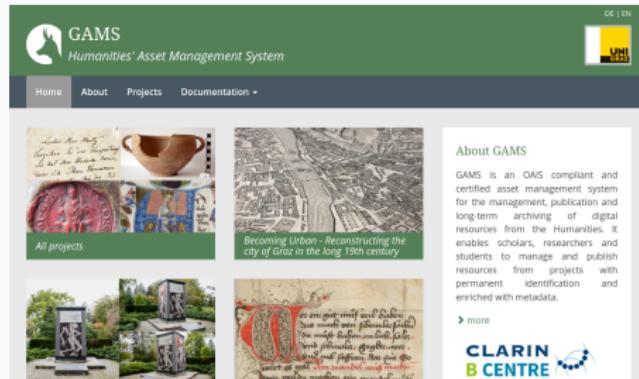
However, not all elements are covered and you have no control over how exactly they are processed.

Tools make things easier superficially but can come at a cost: potential bugs, less than ideal usability, lack of control and customizability:

→ to customize we need to write our own transformation using XSL stylesheets

GAMS (Humanities Asset Management System) i

- GAMS
(Geisteswissenschaftliches Asset Management System = AMS for the Humanities)
- based on **FEDORA** (*Flexible Extensible Digital Object Repository Architecture*) = infrastructure dedicated to the persistent archival and management of resources considered to be worthy of long-term preservation.



<https://gams.uni-graz.at>

GAMS (Humanities Asset Management System) ii

- user access provided through the **Cirilo** client
- **functionalities:** object creation and management, versioning, normalization & standards, choice of data formats.
- offers a plethora of **pre-defined content models** for data such as TEI, MEI, LIDO, SKOS, ontologies, R code and story lines
- → offers publication pipelines but is highly customizable
- more info:
<http://gams.uni-graz.at/doku>



Long-term archiving heritage data i

Long-term preservation

denotes the process of maintaining, curating and keeping data usable over a long period of time (10+ years).

Key functionalities of long-term archiving architectures include

- persistent identification,
- versioning,
- support of different data formats,
- management of associated metadata,
- data export and retrieval,
- security and scalability.

Special emphasis is placed on

- sustainability,
- citability &
- guarantee of long-term access to the contained resources.

XML is great for long-term archiving!

Consequently, data formats and software used for preservation should follow **open source & non-proprietary standards**; data is ideally encoded in an **unicode XML format**:

- plain text files are small
- human- & machine-readable
- recognized standard stable since 1998
- more on XML later...

Examples of data quality in GAMS i

The screenshot shows a search result for the year 1563. The top navigation bar includes links for Home, Projekt, Urfehdebuch, Kategorien, Datenkorb (0), Erweiterte Suche, Volletextsuche, and a search input field. The main content area displays a card for 'Ratsbücher O10, Urfehdenbuch X (1563-1569)'. The card includes a thumbnail, a title, a 'Zitierschlag' section with a URL, and a 'Zeitraum' section indicating the period from 03.01.1563 to 04.12.1569. Below this is a large preview image labeled 'fol. 1v'.

fol. 1v

[Anno d[omi]ni] 1563

Beginn Ende

← ◀ ▶ ⌂ Nicht sicher | gams.uni-graz.at/archive/objects/o:graf.2387/urfehdebuch/defObject/getDC? Early modern c... JAPSS-Mail k... Tools | Addthis... Index lightphone-QB... CircusMe... GIG... Stellenaangebote... Digital Humanis...

The screenshot shows a detailed view of a dataset. The top navigation bar includes links for GAMS, Home, About, Projects, and Documentation. The main content area has two panels. The left panel displays 'Phaedr. 1,18' and 'Grazer Reppositorium antiker Fabeln'. It lists 'Creator(s)' as phaedrus, 'Contributor(s)' as Ursula Gärtn., Lukas Spielberg, Sarah Lang, Lukas Werzer, Martina Hüttemeyer, Katharina Schwarz, Nora Kohlhofer, Ulrike Kalwoda-Bauer, Ursula Gärtn., BIG Rein, Förderverein Sparkling Science, BMBWF, and 'Publisher(s)' as Institut für Antike, FB Klassische Philologie, Karl-Franzens-Universität Graz. A 'Description' section states that the project aims to provide a direct embedding of partner schools, a scientifically founded and didactically refined text portal for ancient fables. A 'More metadata' button is present.

Dataset

Object type: cm:TEI
Permalink: https://gams.uni-graz.at/o:graf.2387
Handle: No handle assigned
Publication date: 2020-07-22T16:42:03.702Z
Last modified: 2020-09-08T12:04:50.176Z
License: Resource licensed under Creative Commons BY-NC 4.0

Versions

o:graf.2387/TEI_SOURCE (current version)

View object

HTML | Fedora profile

Export Metadata

Dublin Core

Files in this object (current version)

TEI_SOURCE tei+xml 29.1 KB View Download

Download selected files (zip) Download all files (zip)

← <http://gams.uni-graz.at/o:ufbas.1563>

- **top image:** info on recommended citation, downloadable source data in XML & RDF
- **below:** ‘data view’ of the Dublin Core (DC) metadata for an XML object in GAMS (ensuring citability, etc.)

← <http://gams.uni-graz.at/archive/objects/o:graf.2387/methods/sdef:Object/getDC?>

Examples of data quality in GAMS ii

- GAMS isn't an out-of-the-box tool – it's something between a **Content Management System (CMS)**, a **repository** (long-term archiving) and a **publication platform**.
- The XSLT transformations applied to files in GAMS are custom but there are **wippets** (standard javascript functionalities → *widget + snippet*, example here) and **standard templates** which can be reused.
- as a publication platform: presentation and long-term archiving, allowing for persistent access to the data, versioning, etc.

Examples of data quality in GAMS iii

One more example: <https://gams.uni-graz.at/beurb>

Metadata view on XML data

Nicht sicher | gams.uni-graz.at/beurb.bei.1910

... PDF-Mit-Resen... Tools | About... Index | Smartphone-G... Gridview-G... Stellmärkte... Digital-Human... Stellmärkte... Jobs | EAON-... Th...

Quellen • Raum-zeitliche Navigation • Analyse • Projekt •

Illustrierter Führer durch die steiermärkische Landeshauptstadt Graz

METS Source TEI SOURCE

Über den Reisebericht Faksimile & Text Karte & Text



Titel	Illustrierter Führer durch die steiermärkische Landeshauptstadt Graz
Autor	Walter von Semetkowsky
Datum	1910
Ort	Graz
Verlag	Uv. Mosers Buchhandlung (J. Meyerhoff) k.k. Hof-Buchhändler
Umfang	Seite 1-83
Permalink	https://gams.uni-graz.at/o/beurb.tel.1910

Map view linked to edited text

Nicht sicher | gams.uni-graz.at/beurb.bei.1910/beitragsmap.html?view=map

... PDF-Mit-Resen... Tools | About... Index | Smartphone-G... Gridview-G... Stellmärkte... Digital-Human... Stellmärkte... Jobs | EAON-... Th...

Quellen • Raum-zeitliche Navigation • Analyse • Projekt •

Über den Reisebericht Faksimile & Text Karte & Text



Legende
- Gebäude
- Straßen
- verkehrsregulierungen
- ausbaustufen
- Wasser
- Plätze
- Städtebau

[1] Praktische Zeiteinteilung.
Wer einen Tag in Graz verbringen möchte, beginnt sich vormittags vor allem in den Schätzen der Elektrischen Sehenswürdigkeiten. Erst im späten Vormittag kann ein Rundgang durch die innere Stadt (S. 27) und ein Rundgang durch die innere Stadt (S. 30). Nachmittags pflegt man einen Ausflug zum Hünstein (S. 78) zu machen; von der Höhe hat eine herrliche Aussicht auf die Stadt und ihre Umgebung; wenn es dem Fremden paßt, läßt sich im Anschluß daran mit Benützung der elektrischen Bahn auch Maria-Trost (S. 78) besuchen. Für den Abend sind vielleicht noch die Vorstellungen des auch verworfenen Anforderungen entsprechenden Stadttheaters (S. 10) zu empfehlen.

Wer zwei oder mehr Tage bleibt, kann den nächsten Vormittag einer eingehenderen Besichtigung der inneren Stadt und der Museen, des Landhauses, Landes-Zeughäuses und Joanneums widmen.

Both views are created ‘on-the-fly’ from the same XML file (=single source principle, more on that later...).

Practice!

Use <https://oxgarage.tei-c.org/> to transform your data to HTML. What do you notice? Do you like it or would you need to customize?

Digital Scholarly Editing

What is a Digital Scholarly Edition? i

Digital editions (as per Sahle 2016):

- discipline-independent & not exclusive to text
- overcome the limitations of print (financial, inter-/ multimediality, etc.)
- follow a digital paradigm like printed editions follow the print paradigm (“shaped by the technical limitations and cultural practices”)
- “A digitised edition is not a digital edition.” → it’s not about the storage medium (just being on the internet doesn’t make it a digital edition!)
- “A digital edition cannot be given in print without significant loss of content and functionality.”
 - different views, interactivity, searchability
 - facsimiles alongside diplomatic transcription and reading versions
 - all generated from the same source data: **single source principle** → How? Data transformation!
 - ≠ Digital Archive, text corpus or facsimile → has critical engagement

What is a Digital Scholarly Edition? ii

We need to start at the beginning:

What is a scholarly edition?

- originally stems from ecdotics (*ecdotica, ecdotique, Editorik / Editionswissenschaft*) → tradition of Philology
- **goal = textual criticism:** reconstruct the original version of a text transmitted to us via textual witnesses (*Lachmannian paradigm*)

Sahle 2016

“ A scholarly edition is the critical representation of historic documents.
Edition ist die erschließende Wiedergabe historischer Dokumente. ”

What is a Digital Scholarly Edition? iii

Criteria from Sahle 2016:

1. **Representation:** recoding a document and its transformation (in the same or other media)
 - Visually: image reproduction, facsimile
 - Textually: transcription.
 - varying degrees of closeness/abstraction with regard to the original document
2. representation \neq presentation
3. **Critical engagement** (based on scholarly agenda). “Critical engagement without representation is not an edition—but an examination, a catalogue or a description.”
 - textual criticism
 - historic criticism
 - bibliographic criticism
 - material criticism
 - visual criticism
 - etc.

What is a Digital Scholarly Edition? iv

4. **Documents:** “every non-abstract object that is the subject of an edition can be called a document.”
5. **Historic:** editions “explain what is not evident to the present-day reader. In short, they bridge a distance in time, a historical difference. Texts that are created today do not need to be critically edited. They can speak for themselves.”

Sahle 2016

“ A representation without [critical] treatment or the addition of information is not an edition—but a facsimile, a reproduction or—nowadays—a digital archive or library.

Critical representation as a compound notion of editing aims at the reconstruction and reproduction of texts and as such addresses their material and visual dimension as well as their abstract and intentional dimension(Sahle 2016). ”

What is a Digital Scholarly Edition? v

Questions to ask if you want to know if it's a Digital Scholarly Edition:

1. **Is there a full representation** of the subject in question?
2. **Is it critical?** → processing rules stated and applied, scholarly knowledge included to make the document easier to understand, regarding material, document genesis/creation, context and reception?
3. **Is the edition of academic quality?** → transparent and rigorous edition process, responsibilities stated, enables future research on a reliable basis.
4. **Does the edition follow a digital paradigm?** → makes use of the possibilities of the digital, not printable without major loss of content or functionality

What is a Digital Scholarly Edition? vi

Ideally, a DSE should also implement the FAIR criteria

Findable. in library catalogs, discovery systems or repositories, with a persistent identifier

Accessible. free for any user, no access restrictions? (from open access to usability, language selection, etc.)

Interoperable. data in standardized and widely used format (for example TEI or similar standards), allowing for reuse and data exchange outside of the project.

Reusable. data accessible (individual download, aggregate download, repository, API)? Licenses allowing reuse. Data creation, modelling and processing documented adequately so that others can make sense of it?

- RIDE (review journal for digital editions and resources)
- RIDE Criteria for Reviewing Digital Editions and Resources

Edition workflows and schools i

The following slides are adapted from Georg Vogeler's slides on digital editing.

Traditional Workflow of Philological Editing

- **Heuristics:** Find your textual witnesses
- **Transcription:** transfer the text into your preferred alphabet (from the original/from a photo)
- **Collation:** Compare the textual witnesses
- **Recensio:** Evaluate the variants and create a stemma (commenting)
- Write your introduction
- Typesetting
- Create an index (referring to pages)
- Print and distribution

In the digital realm:

- edition
- beta version
- just TEI data publication
- hybrid (web and print) publication
- minimal edition (adaptation from minimal computing)

A digital edition is social, iterative, a process...

Edition workflows and schools ii

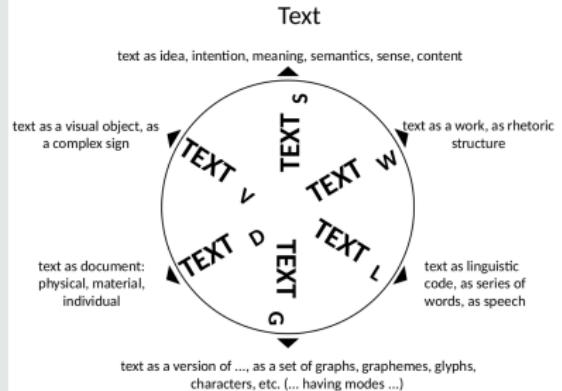
„Edition“

- a particular version of a book
- a particular version of a product
- all the copies of a book that are printed or published at one time

„Historical-critical“ Edition

- Documentation of the history of the text transmission in the “apparatus of variants”
- Critical evaluation of the textual transmission

Patrick Sahle's text wheel



Karl Lachmann (1793–1851)

Editorial intervention: Philological knowledge (linguistics, style) allows to emend fragmentary textual transmission.

Editorial Schools

- Lachmann ('Historisch-Kritische Ausgabe', stemmatology)
- Reading text vs. critical edition = reader-oriented pragmatism / modernised edition / 'reading text'
- Last authorized edition
- First edition / *editio princeps*
- Main manuscript (Bédier 1937, 'Leithandschrift')
- 'diplomatic edition / transcription'
- Typographic/photographic facsimile
- 'documentary editing' (Tanselle)
- variorum edition
- genetic edition / *critique génétique* (Gabler)
- Copy-Text-Theory (Greg 1950/3, Bowers 1976, 1978)
- New Philology (Stackmann 1964, 1993, 1994, 1999; Ruh 1978, Cerquiglini 1989)

Digital Archives

Digital long-term archiving

- ensuring the authentic and sustainable availability of digital resources on the level of the bitstream and on a semantic level
- integral principle to every form of sustainable data storage
- begins with the data production in a sustainable data format (and ideally, following a recognized data standard)
- requires standardization of data formats and archiving workflows
- serves both the dissemination as well as the preservation of digital content
- not just about technical solutions but also institutional stability and policies

A digital archive

- is more than a mere collection of scanned book pages or digitized images, etc.
- if offers metadata, norm data and controlled vocabularies

Digital Archives

Digital Archive

- organized collection of digital objects (text, images, audio, video and multimedia streams)
- digital objects are described by standards both in terms of contents (e.g. TEI) and bibliographically (e.g. Dublin Core)
- published sustainably using interfaces, services and APIs (e.g. OAI-PMH)
- digital objects have unique, persistent and citable identifiers (e.g. DOI, URN, PURL, PID)
- authenticity of objects is checked by means of digital signatures or checksums ('Is the number of bytes in the object still the same it used to be?')

Trustworthy Digital Archives

as defined by the Research Libraries Group (RLG)

- secure organizational structure and legal status
- financial sustainability
- technological and procedural aptness
- ensuring data and system security
- documentation and transparency
- conformity with the OAIS standard

→ retro-digitizing objects but also standards for new (born digital) resources.

Open Archival Information System Reference Mode

OAIS

Generic model for the organization of a digital archive
→ developed 1995–2002 by the Consultative Committee for Space Data Systems (CCSDS)

Tasks for Digital Archives according to OAIS

1. data ingest
2. archival storage
3. data management
4. system administration
5. preservation planning
6. access

What's the difference between a Digital Archive and a Digital Scholarly Edition? (Sahle 2007)



Digital Archives vs. Digital Editions i

Digital documents

Material objects are the target of digitization but digitization doesn't reproduce them – it represents them in a digital format. A digital document is a view on the original material object.

Archive

Traditionally an archive is an ordered collection of documents with the goal of documenting them, preserving them in the long-term and making them accessible. → this function isn't only carried out by actual archives but also by museums, libraries and other cultural heritage institutions. Traditional archive material doesn't need to be represented because its physical objects can be accessed directly.

(we have already def'd editions)

Editions are often based in archival material. The edition isn't a storage device, it is a publication of a historical source and the editorial work done on it. While the edition contains lots of editorial work and enrichment, archival documents are usually original and largely unprocessed. A non-digital archive isn't a form of publication in and of itself – a digital archive *is* a form of publication, too, like a digital edition → blurs the lines a bit.

We could say that **the difference lies in the depth of data enrichment and editorial work.** → once a presentation form is provided, data from a digital archive becomes a digital edition.

Digital Archive

A main goal of a digital archive is to preserve and publish a specific choice of documents as a collection (ideally, representative and well-balanced). The digital objects aren't necessarily direct representations but may have undergone editorial intervention (e.g. normalized orthography). Its documents should be uniform in how they are encoded and processed. In the case of the single source principle, the edition is generated dynamically ('on the fly') from the archived source data. (If any edition is provided).

References

- [1] Patrick Sahle. "Digitales Archiv und Digitale Edition. Anmerkungen zur Begriffsklärung". In: *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien*. Ed. by Michael Stolz. Zürich, 2007, pp. 64–84.
- [2] Patrick Sahle. "What is a Scholarly Digital Edition?" In: *Digital Scholarly Editing: Theories and Practices*. Ed. by Matthew J. Driscoll and Elena Pierazzo. Cambridge: Open Book Publisher, 2016, pp. 19–39. URL:
<https://books.openedition.org/obp/3381>.