

Scholarly Digital Edition

-

Building a Workflow



Diego Canali for **TALOS – AI4SSH**

ERA Chair in Artificial Intelligence for Humanities and Social Sciences

Rethymno, 6 November 2023

Building a Workflow





Building a Workflow

1. Standards for Philology: XML-TEI
2. EVT - Edition Visualization Technology
3. RIDE – Criteria for Reviewing Scholarly Digital Editions



1. Standards for Philology

- XML – eXtensible Markup Language
- TEI – Text Encoding Initiative
- Desiderata: Pierazzo 2019, common infrastructure and standard workflow



1. Standards for Philology

XML – eXtensible Markup Language

- A meta-language
- Syntax, but not vocabulary
- Extensible: you can “invent” your own vocabulary

<note>

<to>Marina</to>

<from>Diego</from>

<header>Reminder</header>

<body>Remember to water the
plants</body>

</note>



1. Standards for Philology

- An XML document is defined:
 - **“well formed”** → if properly structured (nested)
 - **“valid”** → against a Schema

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.
3 <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.
4   schematypens="http://purl.oclc.org/dsdl/schematron"?>
5 <TEI xmlns="http://www.tei-c.org/ns/1.0">
6   <teiHeader>
7     <fileDesc>
8       <titleStmt>
9         <title>Title</title>
10      </titleStmt>
11     <publicationStmt>
12       <p>Publication Information</p>
13     </publicationStmt>
14     <sourceDesc>
15       <p>Information about the source</p>
16     </sourceDesc>
17   </fileDesc>
18 </teiHeader>
19 <text>
20   <body>
21     <p>Some text here.</p>
22   </body>
23 </text>
24 </TEI>
```

Text Grid Author

C:\...\Untitled1.xml Validation succ... U+003C 1 : 1 Days left for evaluation: 27 11 new mess...

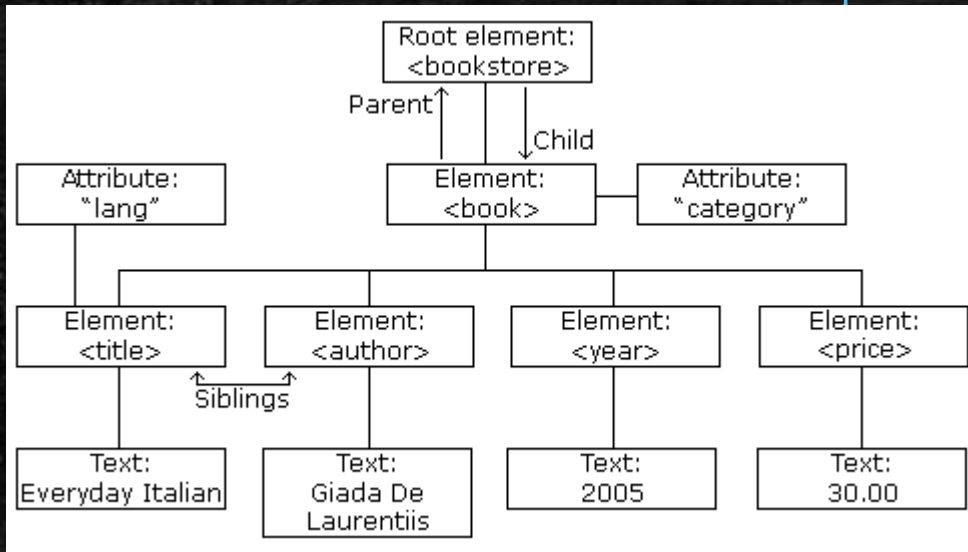


1. Standards for Philology



- XML Tree Structure → Hierarchy
- <https://www.w3schools.com/XML/default.asp>

- A simple example



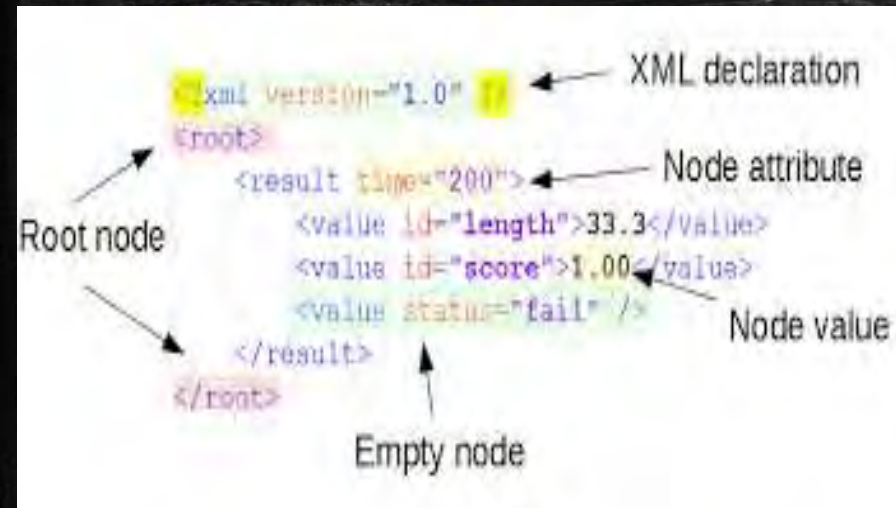
```
<?xml version="1.0" encoding="UTF-8"?>
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this
weekend!</body>
</note>
```



1. Standards for Philology

XML Documents Must Have:

- The Prolog (XML declaration)
- The Root Element
- All XML Elements Must Have an Opening and Closing Tag
- XML Tags are Case Sensitive
- XML Elements Must be Properly Nested
- XML Attribute Values Must Always be "Quoted"
- Entity References (& → &#x2013;)





1. Standards for Philology

HTML



CSS



XML

- Structure to an unstructured text
- Metadata
 - Of the text
 - Of the source(s)
 - Of the responsible(s)
 - ...
- BUT **no** info about **LAYOUT** → need different languages (*e.g.* CSS)
- Struggle: more than one formal language



1. Standards for Philology



- TEI P5 – Guidelines

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

- “The Text Encoding Initiative Consortium is an international organization whose mission is to develop and maintain **guidelines** for the **digital encoding** of literary and linguistic texts”
- “An **international and interdisciplinary standard** that is widely used by libraries, museums, publishers, and individual scholars to **represent** all kinds of textual material for online research and teaching”



< Text Encoding Initiative >

TEI P5 – Guidelines

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

[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.](#)



< Text Encoding Initiative >

- A Gentle introduction to TEI
- <https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>

v. A Gentle Introduction to XML

Table of contents

- v.1. What's Special about XML?
- v.2. Textual Structures
- v.3. XML Structures
- v.4. Validating a Document's Structure
- v.5. Complicating the Issue
- v.6. Attributes
- v.7. Other Components of an XML Document
- v.8. Putting It All Together



< Text Encoding Initiative >

1 The TEI Infrastructure

<https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html>

1.1 TEI Modules

1.2 Defining a TEI Schema

1.3 The TEI Class System

1.4 Macros

1.5 The TEI Infrastructure Module

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	17 Simple Analytic Mechanisms
certainty	Certainty and Uncertainty	21 Certainty, Precision, and Responsibility
core	Common Core	3 Elements Available in All TEI Documents
corpus	Metadata for Language Corpora	15 Language Corpora
dictionaries	Print Dictionaries	9 Dictionaries
drama	Performance Texts	7 Performance Texts
figures	Tables, Formulae, Figures	14 Tables, Formulæ, Graphics, and Notated Music
gaiji	Character and Glyph Documentation	5 Characters, Glyphs, and Writing Modes
header	Common Metadata	2 The TEI Header
iso-fs	Feature Structures	18 Feature Structures
linking	Linking, Segmentation, and Alignment	16 Linking, Segmentation, and Alignment
msdescription	Manuscript Description	10 Manuscript Description
namesdates	Names, Dates, People, and Places	13 Names, Dates, People, and Places
nets	Graphs, Networks, and Trees	19 Graphs, Networks, and Trees
spoken	Transcribed Speech	8 Transcriptions of Speech
tagdocs	Documentation Elements	22 Documentation Elements
tei	TEI Infrastructure	1 The TEI Infrastructure
textcrit	Text Criticism	12 Critical Apparatus
textstructure	Default Text Structure	4 Default Text Structure
transcr	Transcription of Primary Sources	11 Representation of Primary Sources
verse	Verse	6 Verse



< Text Encoding Initiative >

- Among the `<teiHeader>` elements, only `<fileDesc>` is *mandatory*

2 The TEI Header

<https://tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>

2.1 Organization of the TEI Header

2.2 The File Description

2.3 The Encoding Description

2.4 The Profile Description

2.5 Non-TEI Metadata

2.6 The Revision Description

2.7 Minimal and Recommended Headers

2.8 Note for Library Cataloguers

2.9 The TEI Header Module

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>
        <!-- title of the resource -->
      </title>
    </titleStmt>
    <publicationStmt>
      <p>
        <!-- Information about distribution of the resource -->
      </p>
    </publicationStmt>
    <sourceDesc>
      <p>
        <!-- Information about source from which the resource derives -->
      </p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```



< Text Encoding Initiative >

3 Elements Available in All TEI Documents

<https://tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COBICOR>

3.1 Paragraphs

3.2 Treatment of Punctuation

3.3 Highlighting and Quotation

3.4 Terms and Glosses, Ruby Annotations, and Equivalents and Descriptions

3.5 Simple Editorial Changes

3.6 Names, Numbers, Dates, Abbreviations, and Addresses

3.7 Simple Links and Cross-References

3.8 Lists

3.9 Notes, Annotation, and Indexing

3.10 Graphics and Other Non-textual Components

3.11 Reference Systems

3.12 Bibliographic Citations and References

3.13 Passages of Verse or Drama

3.14 Overview of the Core Module



< Text Encoding Initiative >

4 Default Text Structure

<https://tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>

4.1 Divisions of the Body

4.2 Elements Common to All Divisions

4.3 Grouped and Floating Texts

4.4 Virtual Divisions

4.5 Front Matter

4.6 Title Pages

4.7 Back Matter

4.8 Module for Default Text Structure

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
  </teiHeader>
  <text>
    <front>
      <!-- front matter for composite text -->
    </front>
    <group>
      <text>
        <front>
          <!-- front matter of first unitary text, if any -->
        </front>
        <body>
          <!-- body of first unitary text -->
        </body>
        <back>
          <!-- back matter of first unitary text, if any -->
        </back>
      </text>
      <text>
        <body>
          <!-- body of second unitary text -->
        </body>
      </text>
    </group>
    <back>
      <!-- back matter for composite text, if any -->
    </back>
  </text>
</TEI>
```




< Text Encoding Initiative >

10 Manuscript Description

<https://tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>

10.1 Overview

10.2 The Manuscript Description Element

10.3 Phrase-level Elements

10.4 The Manuscript Identifier

10.5 The Manuscript Heading

10.6 Intellectual Content

10.7 Physical Description

10.8 History

10.9 Additional Information

10.10 Manuscript Parts

10.11 Manuscript Fragments

10.12 Module for Manuscript Description



< Text Encoding Initiative >

11 Representation of Primary Sources

<https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>

11.1 Digital Facsimiles

11.2 Combining Transcription with Facsimile

11.3 Scope of Transcriptions

11.4 Aspects of Layout

11.5 Transcription and Ruby

11.6 Headers, Footers, and Similar Matter

11.7 Identifying Changes and Revisions

11.8 Other Primary Source Features not Covered in these Guidelines

11.9 Module for Transcription of Primary Sources



< Text Encoding Initiative >

12 Critical Apparatus

<https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>

12.1 The Apparatus Entry, Readings, and Witnesses

12.2 Linking the Apparatus to the Text

12.3 Using Apparatus Elements in Transcriptions

12.4 Strategies for Encoding Variation

12.5 Module for Critical Apparatus

```
<p n="4">
  <!--here we find the first direct speech-->
  <said who="#Καλ" direct="true">Εἰς καλόν γε ὑμῖν συντετύχηκα· ἐστὶν γὰρ μέλλω
    <persName ref="#Αυτο">Αὐτόλυκον</persName> καὶ τὸν πατέρα αὐτοῦ. οἶμαι
    οὖν πολὺ ἂν τὴν κατασκευὴν μοι λαμπροτέραν φανῆναι εἰ ἀνδράσιν <app>
      <lem wit="#Aristides #A">ἐκκεκαθαρμένοις</lem>
      <rdg wit="#B #E #H1 #Ha #C #D #F #H2 #G" type="graphic"
        >ἐγκεκαθαρμένοις</rdg>
    </app> τὰς ψυχὰς ὥσπερ ὑμῖν ὁ ἀνδρῶν κεκοσμημένος εἶη μᾶλλον ἢ εἰ στρατηγοῖς
    καὶ <app>
      <lem>ἰππάρχοις</lem>
      <rdg wit="codd." type="graphic" select="ἰππάρχαις"/>
      <rdg wit="#B" type="substantial">ὑπάρχαις</rdg>
    </app> καὶ <app>
      <lem resp="Dindorf">σπουδαρχίαις</lem>
      <rdg wit="codd.">σπουδάρχαις</rdg>
      <rdg resp="Mehler">σπουδαρχίδαίς</rdg>
    </app>.</said></p>
```



1. Standards for Philology

DESIDERATA for Scholarly Digital Editions

- Distinction
 - *Specialized Editions*
 - *Serial Editions*

- *Define:*
 - *Common infrastructure*
 - *Common workflow and standards for:*
 - *Encoding*
 - *Hosting*
 - *Displaying*

from **Pierazzo E.** (2019) - *Quale infrastruttura per le edizioni digitali?*



2. EVT - Edition Visualization Technology

- How to combine
 - *Serial Editions*
 - *Common infrastructure*
 - *Common workflow and standards?*

Edition Visualization Technology



Edition Visualization Technology

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.

The Digital Vercelli Book CC BY-SA

EVT 1

DEMO V 1.3

EVT Critical Viewer

EVT 2

DEMO V BETA2



Edition Visualization Technology

- EVT was born in the context of the *Digital Vercelli Book* project (first edition January 2014)
- It has evolved in a tool suitable to fit different texts and needs.
- The continuous development and need to adapt it to different types of documents and TEI-encoded texts has shifted the development focus towards creation of a more general tool for the web publication of TEI-based documents, able to cater for multiple use cases.





Edition Visualization Technology

EVT 2: main features

Critical edition support. Enlarged critical apparatus, variant heat map, witnesses collation and variant filtering are some of the main features developed for the critical edition support.

Bookmark. Direct reference to the current view of the web application, considering view mode, current document, page and edition level, eventual collated witnesses and selected apparatus entry.

High level of **customization.** The editor can customize both the user interface layout and the appearance of the graphical components.



Go to the demo online version

http://evt.labcd.unipi.it/demo/evt2-beta2/avicenna/index.html#/readingTxt?d=doc_1&p=C-112v&s=text-body-div&e=critical



Edition Visualization Technology

Born to **display diplomatic editions** → EVT 1

Open source, it can be **used, adapted, modified** by the user

EVT 2 → AngularJS

EVT 3 → Angular 9+ (2021)

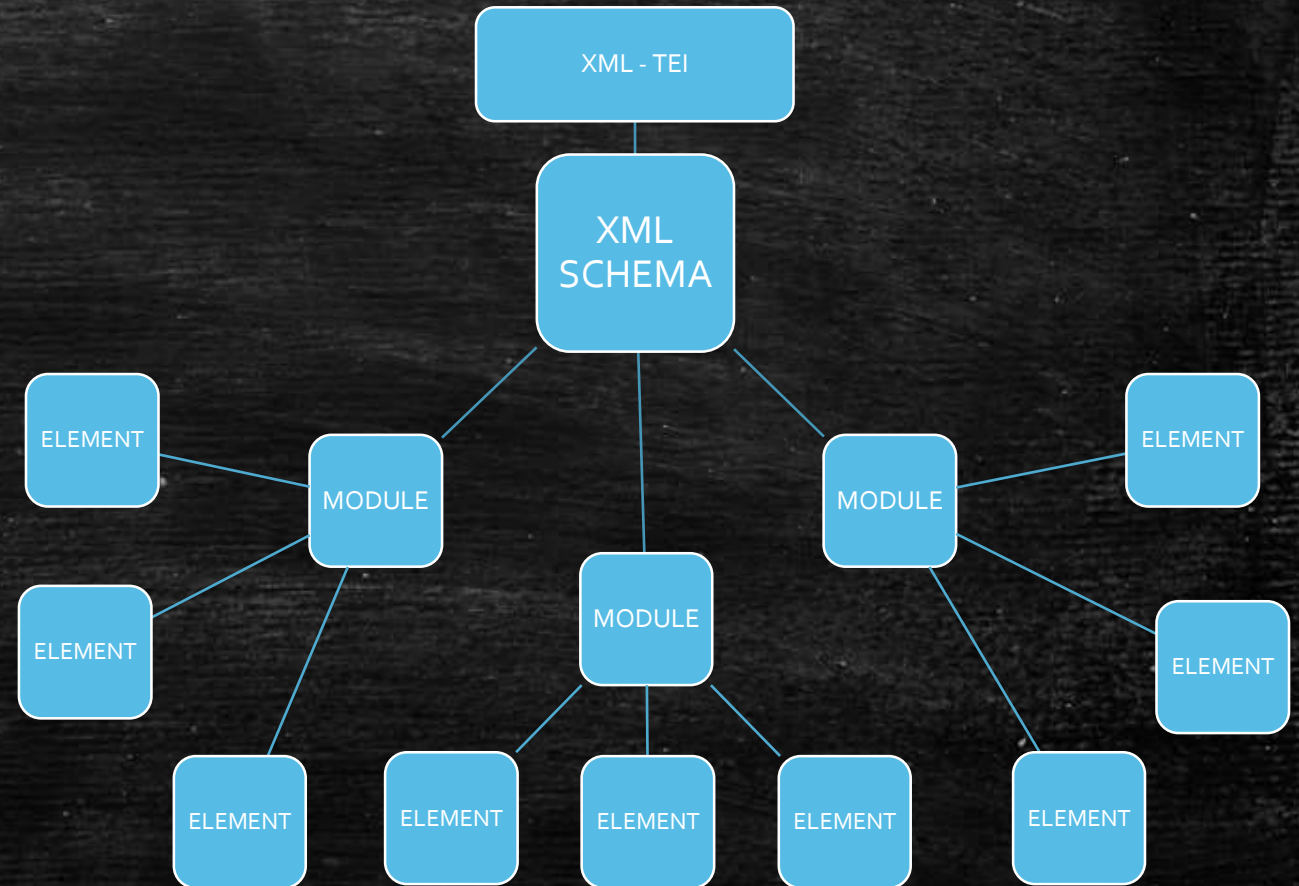
- Configurable layout
- It works automatically (if correctly configured)
- Clear layout
- Simple navigation
- Document data are separated from the software



Edition Visualization Technology

TEI markup for Critical Editions

- Remember:
- XML-TEI is a **descriptive** kind of mark up (not procedural)
- XML documents are independent from hardware and software
- XML-TEI is both **human** and **machine readable**
- Its about the **semantic structure** of a document (not its display)





Edition Visualization Technology

Indispensable modules for critical editions:

- core
- msdescription
- transcr
- textcrit
- analysis
- gaiji

(others not mentatory, but frequent modules are linking, figures, namesdates, verse, certainty)

[analysis] Simple analytic mechanisms

[c](#) [cl](#) [interc](#) [InternGrp](#) [m](#) [rc](#) [chr](#) [s](#) [soan](#) [soanGrp](#) [w](#)

[certainty] Certainty, Precision, and Responsibility

[certainty](#) [precision](#) [resoons](#)

[core] Elements common to all TEI documents

[abbr](#) [add](#) [addri](#) [line](#) [address](#) [analytic](#) [author](#) [bibl](#) [biblScope](#) [biblStruct](#) [binaryObject](#) [cb](#) [choice](#) [cfr](#) [citedRange](#) [corr](#) [date](#) [del](#) [deac](#) [distinct](#) [divGen](#) [editor](#) [ellipsis](#) [email](#) [emph](#) [expan](#) [foreign](#) [gap](#) [gb](#) [glass](#) [graphic](#) [head](#) [headItem](#) [headLabel](#) [hi](#) [imprint](#) [index](#) [item](#) [l](#) [label](#) [lb](#) [lg](#) [list](#) [listBibl](#) [measure](#) [measureGrp](#) [media](#) [meeting](#) [mentioned](#) [milestone](#) [monogr](#) [name](#) [note](#) [noteGrp](#) [num](#) [orig](#) [p](#) [pb](#) [postBox](#) [postCode](#) [ptr](#) [subPlace](#) [publisher](#) [u](#) [quote](#) [cb](#) [ref](#) [reg](#) [relatedItem](#) [reso](#) [respStnt](#) [is](#) [it](#) [tuby](#) [said](#) [series](#) [sic](#) [soCalled](#) [so](#) [speaker](#) [stage](#) [street](#) [teiCorous](#) [term](#) [textLang](#) [time](#) [title](#) [unclear](#) [unit](#)

[gaiji] Character and glyph documentation

[char](#) [charDecl](#) [g](#) [glyph](#) [localProp](#) [mapping](#) [unicodeProp](#) [unihanProp](#)

[namesdates] Names and dates

[addName](#) [affiliation](#) [age](#) [birth](#) [bloc](#) [climate](#) [country](#) [death](#) [district](#) [education](#) [event](#) [faith](#) [floruit](#) [forename](#) [genName](#) [gender](#) [geo](#) [geoFeat](#) [geopName](#) [langKnowledge](#) [langKnown](#) [listEvent](#) [listNym](#) [listObject](#) [listOrg](#) [listPerson](#) [listPlace](#) [listRelation](#) [location](#) [nameLink](#) [nationality](#) [nym](#) [object](#) [objectIdentifier](#) [objectName](#) [occupation](#) [offset](#) [org](#) [scpName](#) [persName](#) [persPronouns](#) [person](#) [personGrp](#) [persone](#) [place](#) [placeName](#) [population](#) [region](#) [relation](#) [residence](#) [roleName](#) [settlement](#) [sex](#) [soccsStatus](#) [state](#) [surname](#) [terrain](#) [trait](#)

[textcrit] Critical Apparatus

[app](#) [lacunaEnd](#) [lacunaStart](#) [lem](#) [listApp](#) [listWit](#) [rds](#) [rdsGrp](#) [variantEncoding](#) [wit](#) [witDetail](#) [witEnd](#) [witStart](#) [witness](#)

[transcr] Transcription of primary sources

[addSoan](#) [am](#) [damage](#) [damageSpan](#) [delSpan](#) [ex](#) [facsimile](#) [fw](#) [handNotes](#) [handShift](#) [line](#) [listTranspose](#) [metamark](#) [mod](#) [path](#) [redo](#) [restore](#) [retrace](#) [seal](#) [sourceDoc](#) [space](#) [subst](#) [substJoin](#) [supplied](#) [surface](#) [surfaceGrp](#) [surplus](#) [transpose](#) [undo](#) [zone](#)

[verse] Verse structures

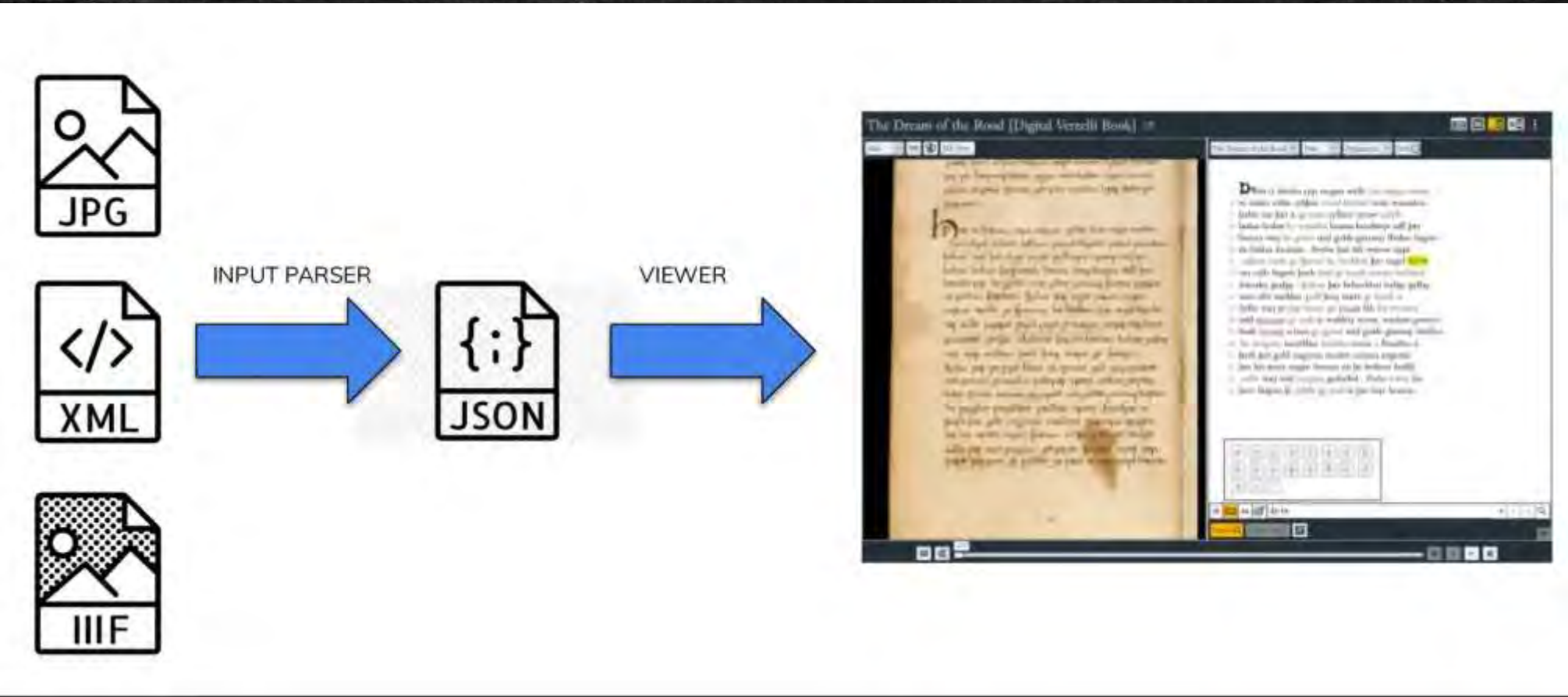
[msdescription] Manuscript Description

[accMat](#) [acquisition](#) [additional](#) [additions](#) [adminInfo](#) [altIdentifier](#) [binding](#) [bindingDesc](#) [catchwords](#) [collation](#) [collection](#) [colophon](#) [condition](#) [custEvent](#) [custodialHist](#) [decoDesc](#) [decoNote](#) [depth](#) [dim](#) [dimensions](#) [explicit](#) [filliation](#) [finalRubric](#) [foliation](#) [handDesc](#) [height](#) [heraldry](#) [history](#) [incipit](#) [institution](#) [layout](#) [layoutDesc](#) [locus](#) [locusGrp](#) [material](#) [msContents](#) [msDesc](#) [msFrag](#) [msIdentifier](#) [msItem](#) [msItemStruct](#) [msName](#) [msPart](#) [musicNotation](#) [objectDesc](#) [objectType](#) [origDate](#) [origPlace](#) [origin](#) [physDesc](#) [provenance](#) [recordHist](#) [repository](#) [rubric](#) [scriptDesc](#) [seal](#) [sealDesc](#) [secFol](#) [signatures](#) [source](#) [stamp](#) [summary](#) [support](#) [supportDesc](#) [surrogates](#) [typeDesc](#) [typeNote](#) [watermark](#) [width](#)



Edition Visualization Technology

- How EVT manages data





Edition Visualization Technology

- **Mark up Methods:**
 - Location-referenced method
 - Double-end-point-attached method
 - *Parallel segmentation method* →
- The chosen method must be specified in the `<teiHeader>`, within the element `<variantEncoding>`.

- Advantages

- In-line mark up → more intuitive than double-end-point attached
- Possibility to nest the variants
- Possibility to reconstruct the text of all the witnesses

- Disadvantages

- Variants cannot overlap
- The apparatus is in-line, which means it is not in a different, external part from the text
- Not supposed to be used with complex traditions



Edition Visualization Technology

- Set of elements and attributes:

- <app>
- <lem>
- <rdg>
 - @wit="#"
 - @type
 - @cause
 - @varSeq
 - @resp
 - @ref
 - @hand
- <witness>
- <listWit>

```
<p n="3"> ἰδὼν δὲ ὁμοῦ ὄντας <persName ref="#Σοκ">  
  <app>  
    <lem>Σωκράτην</lem>  
    <rdg wit="#F" type="orthographic">σοκράτη</rdg>  
  </app></persName> τε καὶ <persName ref="#Κρι">Κριτόβουλον</persName> καὶ  
  <persName ref="#Ερμ">Ἐρμογένην</persName> καὶ <persName ref="#Αντι">  
  <app>  
    <lem>Ἀντισθένην</lem>  
    <rdg wit="#F #H2" type="orthographic">Ἀντισθένη</rdg>  
  </app></persName> καὶ <persName ref="#Καρ">Χαρμίδην</persName>.
```

```
<p n="4">  
  <!--here we find the first direct speech-->  
  <said who="#Καλ" direct="true">Εἶς καλὸν γε ὑμῖν συντετύχηκα· ἐστὶν γὰρ μέλλω  
    <persName ref="#Αυτο">Αὐτόλυκον</persName> καὶ τὸν πατέρα αὐτοῦ. οἶμαι  
    οὖν πολὺ ἂν τὴν κατασκευὴν μοι λαμπροτέραν φανῆναι εἰ ἀνδράσιν <app>  
      <lem wit="#Aristides #A">ἔκκεκαθαρμένοις</lem>  
      <rdg wit="#B #E #H1 #Ha #C #D #F #H2 #G" type="graphic">  
        >ἔγκεκαθαρμένοις</rdg>  
    </app> τὰς ψυχὰς ὡσπερ ὑμῖν ὁ ἀνδρῶν κεκοσμημένος εἴη μᾶλλον ἢ εἰ στρατηγοῖς  
    καὶ <app>  
      <lem>ἰππάρχοις</lem>  
      <rdg wit="codd." type="graphic" select="ἰππάρχαις"/>  
      <rdg wit="#B" type="substantial">ὑπάρχαις</rdg>  
    </app> καὶ <app>  
      <lem resp="Dindorf">σπουδαρχίαις</lem>  
      <rdg wit="codd.">σπουδάρχαις</rdg>  
      <rdg resp="Mehler">σπουδαρχίδαίς</rdg>  
    </app>.</said></p>
```



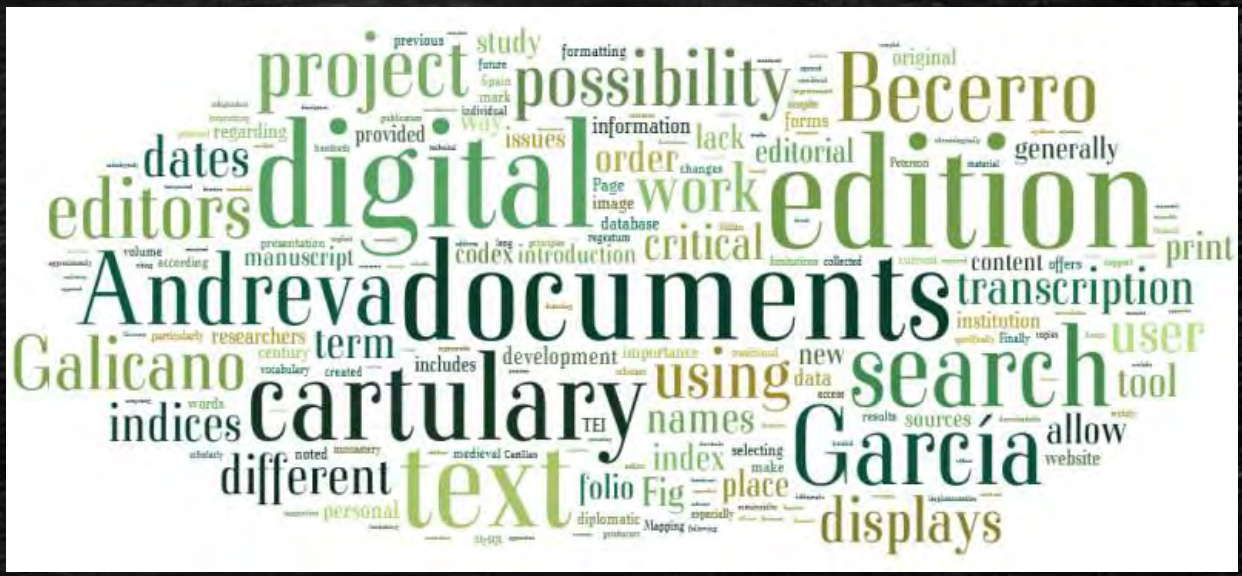
3. RIDE - Criteria for Reviewing Scholarly Digital Editions



Institut für Dokumentologie und Editorik

<https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/>

"Edition is the comprehensive reproduction of historical documents."





Institut für Dokumentologie und Editorik



“RIDE is a review journal dedicated to digital editions and resources and providing a critical forum for their discussion”

“RIDE reviewers are encouraged not only to discuss the traditional achievements and problems of editions, corpora, etc. in general, but also to consider evolving methodology and its technical implications”



IDE has created a comprehensive catalog of criteria that should be used as the basis for the discussions:

- Criteria for Reviewing Scholarly Digital Editions →
- Criteria for Reviewing Text Collections
- Criteria for Reviewing Tools and Environments for Digital Scholarly Editing

▪ **Criteria for Reviewing Scholarly Digital Editions**

- **Patrick Sahle**; in collaboration with Georg Vogeler and the members of the IDE; Version 1.1, June 2014
- Major contributions to this English version by: Misha Broughton, James Cummings, Franz Fischer, Philipp Steinkrüger, Walter Scholger.
- Published volumes: 17, May 2023



PRELIMINARY REMARKS

- DEFINITION: what can be considered a Scientific Digital Edition?

three requisites:

1. A justification of the editorial method adopted and a clear description of the rules that guided the edition --> these information are given in the XML Schema;
 2. Compliance with scholarly requirements towards content and quality, which includes that the self-stated rules are followed;
 3. An editorial concept that is not restricted to the technological limitations of print technology but that realizes a "digital paradigm".
- BEST PRACTICES: basic set of requirements which should be addressed by an SDE, either by satisfying them or by justifying their non-application. But keep in mind that reviews of SDEs should be seen as yet another contribution to ongoing methodological discussions --> continuous updates.



APPLICATION

- General criteria --> applicable to all kinds of Editions
- Specific criteria --> only for specific kind of SDEs
- Space for personal taste: --> not everything can be systematically explained, e.g. the choice of what to encode

PARAMETERS: to review an SDE we have to consider the circumstances:

1. goals
2. financial resources
3. the run-time of the project
4. existing resources



- **NB:** an SDE is very likely an ongoing process: at the moment it's reviewed, what's the status of the edition, if it's ready, if it can be considered ready for research purpose...? it's a good practice to leave a professional comment in the editorial process, and a suggestion for improvement
- **NB:** the reviewer also should be qualified, and be able to disguise about:
 - methodological scope of the edition
 - its implementation



1. OPENING THE REVIEW

- 1.1 the reviewer: provide your academic background
- 1.2 bibliographic identification of the reviewed SDE: title, responsible editors, other responsible people or institutions, dates of publication, address...
- 1.3 general introduction: subject, scope, context of the SDE; its relationship with printed or original sources; desiderata it's addressed to...
- 1.4 general parameters: editors, institutions, responsibilities; relation with other projects; financial, personnel and time resources; implementation...
- 1.5 transparency: are the general parameters easily accessible?



2. SUBJECT AND CONTENT OF THE EDITION

2.1 selection: how relevant is the SDE to current and future research?

2.2 previous and project's achievements: why is the SDE important? what is new?

2.3 content: what does the SDE publish? is there anything important missing? is the omission explained and/or justified?



3. AIMS AND METHODS

- 3.1 documentation: is there a description of aims and methods of the SDE? if not, is it self evident?
- 3.2 scholarly objectives: what academic questions does the SDE answer to? what's the field of research?
- 3.3 mission: what does the SDE want to accomplish? what does it promise explicitly? what's the target audience?
- 3.4 method: which editorial school does the SDE follow? which methodological approach? is it focused more on "works" or "documents"?



- 3.5 representation of documents and texts: how are the documents and text represented? what's the role and quality of digital images? do we have information about the transcription rules? is the transcription "document-centric" or "text-centric"? are amendments and a reconstruction of the ideal text available?
- 3.6 text criticism, indexing and commentary: what kind of textual criticism is practiced? (stemma, detailed description of manuscripts..) what kinds of indexing, commentary and description of documents and texts are applied?
- 3.7 data modelling: How is the editorial method technically implemented? What data model is applied? Is the documentation of the data model sufficient? Which data formats are used? Does the SDE follow common standards (e.g. TEI guidelines)? If yes, is the data modelling documented through a formal schema available on the SDE's site?



4. PUBLICATION AND PRESENTATION

- 4.1 technical infrastructure: which technologies are used for the publication of the SDE and why?
- 4.2 interface usability: is it user friendly even without a massive preliminary training? Can the user quickly identify the purpose, the content and the main access methods of the SDE?
- 4.3 browse: Is it possible to browse through the entirety of the content? Is browsing access easy to understand and allow for fast access to any part of the content?
- 4.4 search: Is there a simple and/or a complex search interface? How can you constrain your search?
- 4.5 indices: Are the indexes, commentary or description used in the presentation of the content?



- 4.6 quality of the presentation: are the images of sufficient quality for research purpose? are there significant errors in the transcription? is there a critical commentary on the textual tradition or the interpretation of the texts? can the user change the presentation of the material, e.g. from a diplomatic transcription to a normalized version?
- 4.7 metadata for description of and interlinkage between objects in the edition: How are the various constituent parts and objects of the edition described? Are they described clearly and comprehensively? Are the single parts interlinked? Are different text surrogates linked (e.g. text and image)? Are there internal links to further contextual information? Are the single parts linked to external resources?
- 4.8 identification and citation: Are there persistent identifiers for the objects of the SDE? Which level of the content structure do they address? Which resolving mechanisms and naming systems are used? Does the SDE supply citation guidelines?
- 4.9 technical interfaces: Are there technical interfaces which allow the reuse of the data of the SDE in other contexts? Can you harvest or download the data?
- 4.10 social integration: Does the SDE integrate with social media and / or virtual research platforms, easily allowing sharing/discussion of particular parts? Does the project have a social media presence in their community?



- 4.11 Spin offs and export formats: Are there alternative formats available, e.g. a printable version or digital formats for specific reading devices (e-Books, mobile devices etc.)?
- 4.12 Access to basic data: Is the basic or underlying data of the edition accessible (e.g. in XML) and if so, how? Is it provided for each single object and/or for the whole SDE? Is the access part of the SDE's user interface or part of an external repository? If you cannot access the basic data, is a justification provided?
- 4.13 Rights and licences: Does the SDE provide sufficient information on rights and restrictions for the reuse of different parts of the SDE (e.g. images, transcriptions, editorial comments)? Does the SDE utilize a rights model feasible for scholarly reuse of the data? Is a specific licence model (e.g. Creative Commons) in use?
- 4.14 Additional features: Does the SDE provide features that merit special attention because they are particularly useful and/or unusual? Think of visualisations, interactivity, image manipulation, options for annotations, commentary notes and personalisation etc.
- 4.15 Documentation and associated texts: Does the SDE provide an introduction or explanatory texts? Is there a help system? Is there sufficient documentation of the project, the edition, and the technical implementation of the SDE? Are the source and the selection of the material described? Are the editorial principles extensively and clearly explained?
- 4.16 Long term use: What are the SDE's prospects for long term use? Is the edition complete or does it promise further modifications and additions? Is there institutional support for the curation and sustainment of the SDE? Is the basic data archived? Is there a plan to provide continuous access to the presentation?



5. CONCLUSIONS

- 5.1 Terminology: Can you classify the project as an "SDE", and if so, by what definition of an SDE? How would you describe the digital resource as the outcome of an editorial project briefly? If the published results do not fulfill some of the minimal requirements (such as the documentation of the textual tradition, rule based representation, transparency of the editorial decisions, scholarly quality), you should point out that the edition does not conform to scholarly standards and therefore should not be considered an SDE.
- 5.2 Realisation of aims: To what extent has the SDE successfully accomplished its original aims?
- 5.3 Fulfillment of general requirements: Does the project fulfill the requirements of a state of the art SDE? Does it fulfill the two basic requests for creators of a SDE: "1: State what you do and act accordingly. 2: Keep to the common scholarly standards." Is the edition sufficiently documented? Is it citable and transparent? How is the quality of the content (images, texts, indexing, commentary, context information)?



- 5.4 Contribution of the SDE to scholarship: What does the SDE contribute to current scholarship in its target field? What does the SDE contribute to best practices in digital scholarly editing in general? What does the SDE accomplish which surpasses the possibilities of a printed edition?
- 5.5 Particularities: Which features merit special attention for noteworthiness and/or innovation, even if they are beyond the scope of these general criteria?
- 5.6 Usability, usefulness, quality: Is the SDE easy to use? Is it a useful contribution to a specific field of research? How would you describe its academic quality?
- 5.7 Suggestions for improvement: If the project is not complete and finished, what should be considered for further improvement? What would be nice and useful additions? What would be the most desirable steps after and outside an already terminated project?



Minimal bibliography and websites

Articles:

Pierazzo 2019 - *Quale infrastruttura per le edizioni digitali. Dalla tecnologia all'etica*

Rosselli del Turco 2019 - *Designing Advanced Software Tool for Digital Scholarly Editions*

Websites:

- **EVT Workshop at AIUCD2021**
 - Rosselli Del Turco R. - Introduzione a EVT Edition Visualization Technology. <https://www.youtube.com/watch?v=vAn2G18Nbgw&t=126s>
 - Spinelli F. - Markup TEI per edizioni critiche. <https://www.youtube.com/watch?v=pnDZRMV3doo&t=2857s>
 - Cerretini G. - Installazione, configurazione e uso di EVT. <https://www.youtube.com/watch?v=qRP8aJ5WyPA>
- **DariahTeach** - Text Encoding and the Text Encoding Initiative. <https://teach.dariah.eu/course/view.php?id=23>
- **W3Schools** - Tutorial for XML language. <https://www.w3schools.com/XML/default.asp>
- **TEI Guidelines**. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- **EVT Edition Visualization Technology**. <http://evt.labcd.unipi.it>
- **IDE - RIDE** - Criteria for Reviewing Scholarly Digital Editions. <https://www.i-d-e.de/publikationen/ride/>

Thank you for the attention



Diego Canali for **TALOS – AI4SSH**
ERA Chair in Artificial Intelligence for Humanities and Social Sciences
Rethymno, 6 November 2023