

# Supplemental Materials for Tag-Xplore: Interactive Exploration of Annotation Practices in Digital Editions

M. Blum,<sup>†1</sup> M. Sachdeva<sup>1</sup>, Y. Stricker<sup>2</sup>, R. Mumenthaler<sup>2,3</sup>, and J. Bernard<sup>1,3</sup>

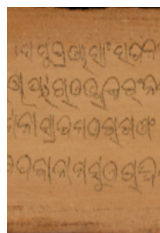
<sup>1</sup>University of Zürich, Zürich, Switzerland; <sup>2</sup>University Library of Zürich, Zürich, Switzerland; <sup>3</sup>Digital Society Initiative, Zürich, Switzerland

## Abstract

This document contains additional, supplemental material for Tag-Xplore. We provide details to the digital editions studied, the characteristics of TEI-XML documents, and the technical architecture of Tag-Xplore.

## 1. Digital Editions

The six digital editions studied in Tag-Xplore are coordinated by the Center for Digital Editions at the University of Zürich (ZDE). They span multiple research domains and are heterogeneous in their research context and scope.

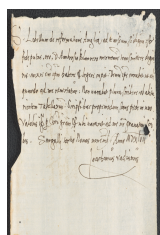


### Paippalāda Recension of the Atharvaveda

This digital critical edition of the Paippalāda-Samhitā (PS) of the Atharvaveda provides an English translation, a full morpho-lexical analysis, detailed linguistic and Indological comments, and links to Sanskrit literature.

Institute of Comparative Language Science and Institute of Asian and Oriental Studies UZH  
Asia, 1st millennium BCE

1 document; 75'301 total tags; 65 unique tags



### Bullinger Digital

The extensive collection of correspondence letters of Heinrich Bullinger provide valuable insights into the political events, theological disputes, economic conditions and everyday life in the Early Modern period.

Institut für Computerlinguistik und Institut für Schweizerische Reformationsgeschichte der Universität Zürich  
Europe, 16th century

13'159 documents; 1'569'899 total tags; 86 unique tags

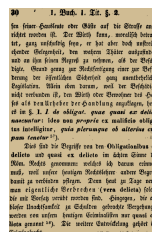


### Heinrich Wölfflin Collected Works

This digital edition project publishes the collected works of the art historian Heinrich Wölfflin, alongside commentary and additional contextualizing material.

Institute of Art History UZH  
Europe, 19th and 20th century

4 documents; 15'205 total tags; 103 unique tags



### iurisprudentia

The iurisprudentia project aims to improve accessibility of historical law documents. Its initial focus is on legislative works of german-speaking Europe.

Institute of Private Law UZH  
Europe, 18th to 20th century

352 documents; 534'847 total tags; 31 unique tags



### Königsfelden

A historic collection of documents from the Königsfelden monastery and its environment. It provides rich and diverse insight into topics such as economy, politics, and social movements over centuries.

Department of History UZH  
Europe, 13th to 17th century

1'547 documents; 838'583 total tags; 83 unique tags



### Robert Walser Critical Edition

This critical edition is a collection of works from the Swiss author Robert Walser. Among other things, it contains published book material, newspaper prints and manuscripts.

German Department UZH  
Europe, 19th and 20th century

20 documents; 94'271 total tags; 70 unique tags

<sup>†</sup> Corresponding Author

## 2. TEI-XML Document

All the documents in the introduced digital editions are annotated using the TEI-XML guidelines. As such, they are semi-structured XML documents. The space of annotation tags contains the 589 tags from the official TEI guidelines. Each tag belongs to one of the 19 modules, which is a categorization provided by the TEI guidelines. On some occasions, documents contain non-TEI-conformant tags. Across all six editions, there are five of these tags. These special tags are categorized under an additional *custom* module. Figure 1 shows an illustrative excerpt of a TEI-XML document from the *Robert Walser* edition.

```

<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmnt>
        <title> NZZ 30.10.1921: Tramfahrt </title>
      >
      <author> Robert Walser </author>
      <sponsor> KWA </sponsor>
      </titleStmnt>
      <publicationStmnt>
        <authority> Kritische Robert Walser-
          Ausgabe (KWA) </authority>
        <pubPlace> Basel </pubPlace>
        <date when="2013"> 2013 </date>
      </publicationStmnt>
      <sourceDesc> [...] </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text xml:id="P9">
    <body>
      [...]
      <head type="Texttitel" xml:id="N12862">
        Tramfahrt. </head>
      <byline> Von Robert Walser. </byline>
      <p xml:id="N1286E">
        Zu Fuß bin ich ein <app><lem> Landsknecht
          , Reisläufer </lem><rdg wit="KPo
          -1928-24"> Landsknecht </rdg></app>;
        <app> <lem> steigt </lem><rdg wit="KPo
          -1928-24"> steige </rdg></app> ich in
        <app><lem> ein </lem><rdg wit="KPo
          -1928-24"> eine </rdg></app> Tram
        [...]
      </p>
    </body>
  </text>
</TEI>
  
```

Figure 1: An example TEI-XML document from the *Robert Walser* edition. TEI-XML tags are highlighted in blue. One can observe the *teiHeader* containing metadata about the document and the *text* part containing the digitized content.

## 3. Technical Architecture

Tag-Xplore contains non-visual analysis components. This includes a data preprocessing pipeline, database storage, dimensionality reduction components, and an API for communication between the frontend and backend.

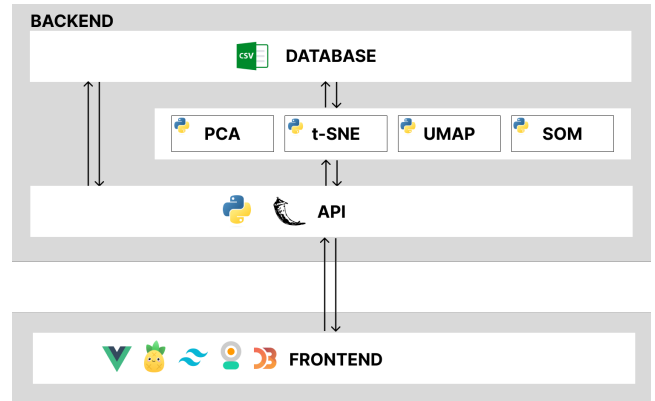


Figure 2: The high-level architecture of Tag-Xplore. The separate frontend and backend components exchange data over HTTP. Inside the backend, the API calls various helper functions and fetches data directly from the database.