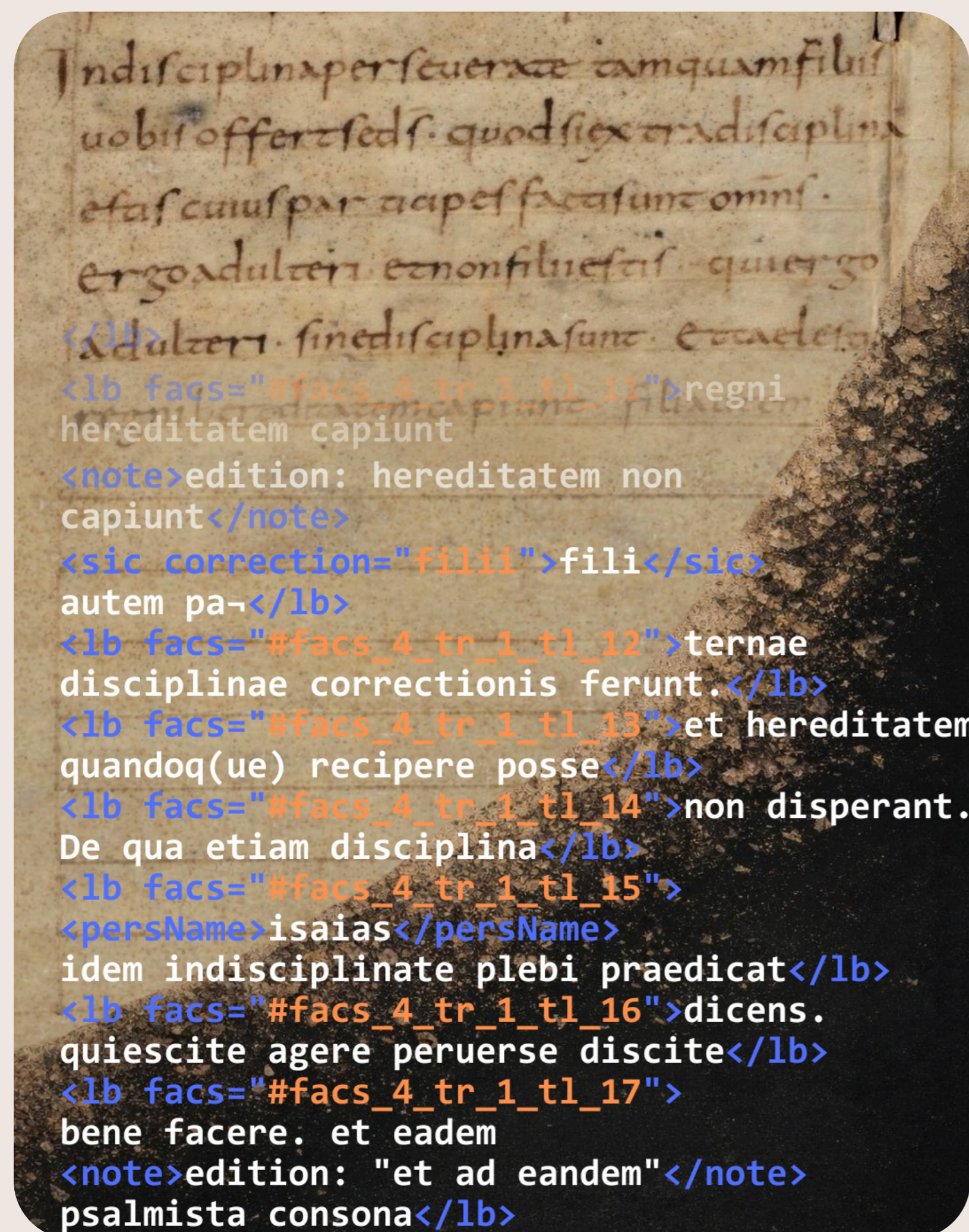




From Fragments to Findings

Research infrastructures for digital humanities



Digitisation

- Online Access
- Digital Reconstruction



Processing

- Image Pre-Processing
- OCR/HTR
- Text Encoding (TEI-XML)



AI Analysis

- Data Post-Processing
- NER
- Automated Data Extraction
- Semantic Text Comparison



Publishing

- Digital Edition
- Long-Term Preservation

Sources → Data → Knowledge

Historical documents

Libraries house many untapped treasures for research within their historical collections. Literally a treasure trove are fragment boxes which house remainders of books, which haven't survived in their original form but are important witnesses of the texts they carry. With these sometimes quite small puzzle pieces we can complete our picture of the past - and in some cases even the original document.

Digital Methods

To support computational research, digitized historical documents are first converted into machine-readable formats. This process utilizes automatic text recognition (OCR/HTR) and TEI-XML encoding to establish a versatile textual foundation. The resulting structured output allows for the application of AI-powered Natural Language Processing (NLP) methods, including Named Entity Recognition (NER), automated metadata extraction, and computational text comparison, to enable data-driven analysis provided by research-oriented library services for new research opportunities.

Library Infrastructure

Digital editions open up new possibilities to explore historical texts that go beyond those supplied by printed editions. While many digital editions offer special functionalities, which are at the cutting edge of technical development at the time, they often don't age well. Usually created within a project context there is no one to maintain and update them. The University Library of Tübingen is on its way to establish in collaboration with the Department of Medieval History the TEI-Publisher as a sustainable infrastructure solution, which is durable and easy to maintain. The content of these editions is archived so that it will be preserved for the long term and can be reused in other contexts.

