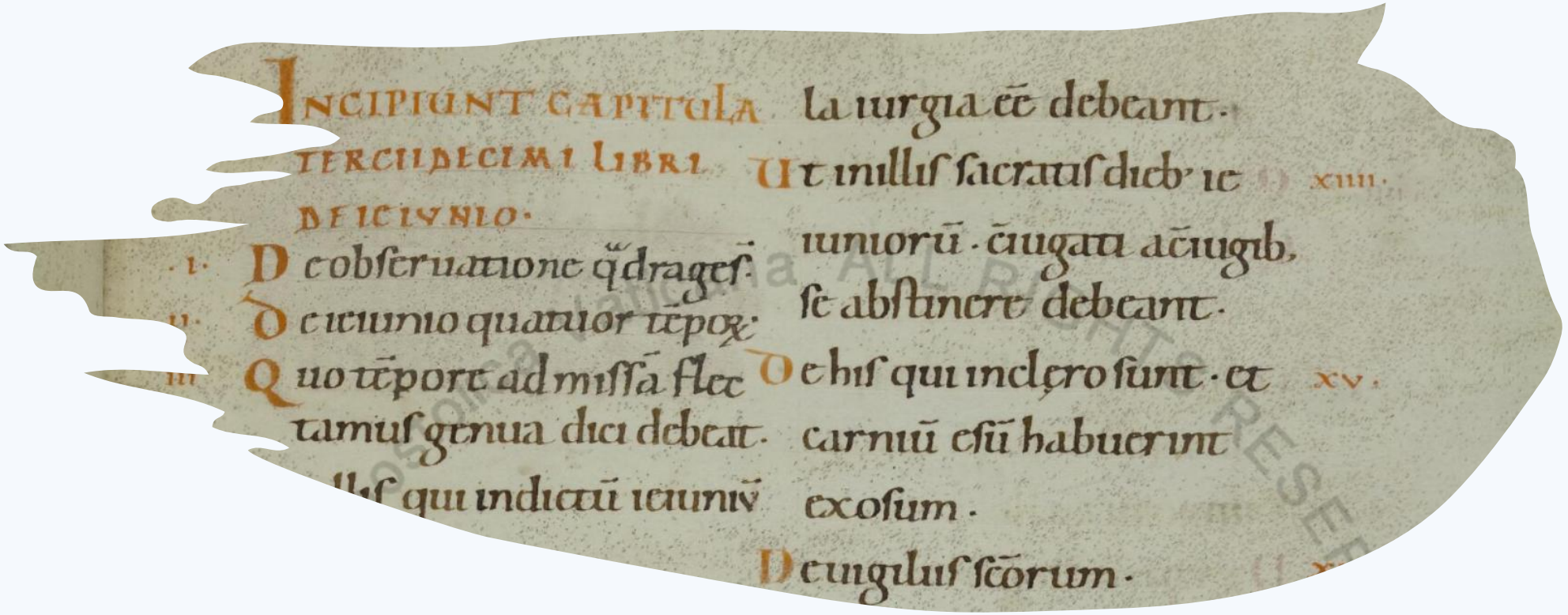


Transkribus

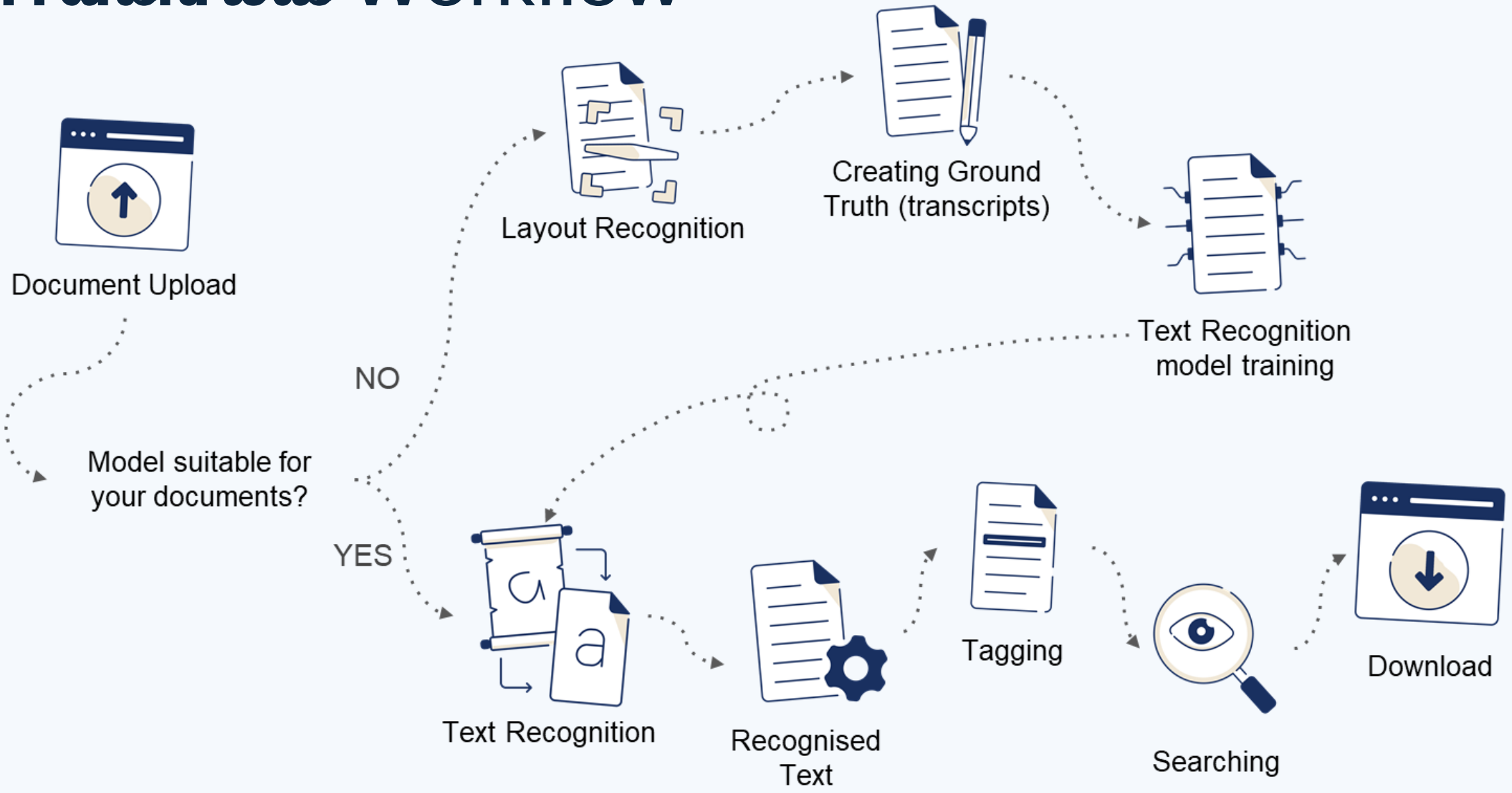
Data Export

Michela Parma, JGU Mainz Universität
michela.parma@uni-mainz.de

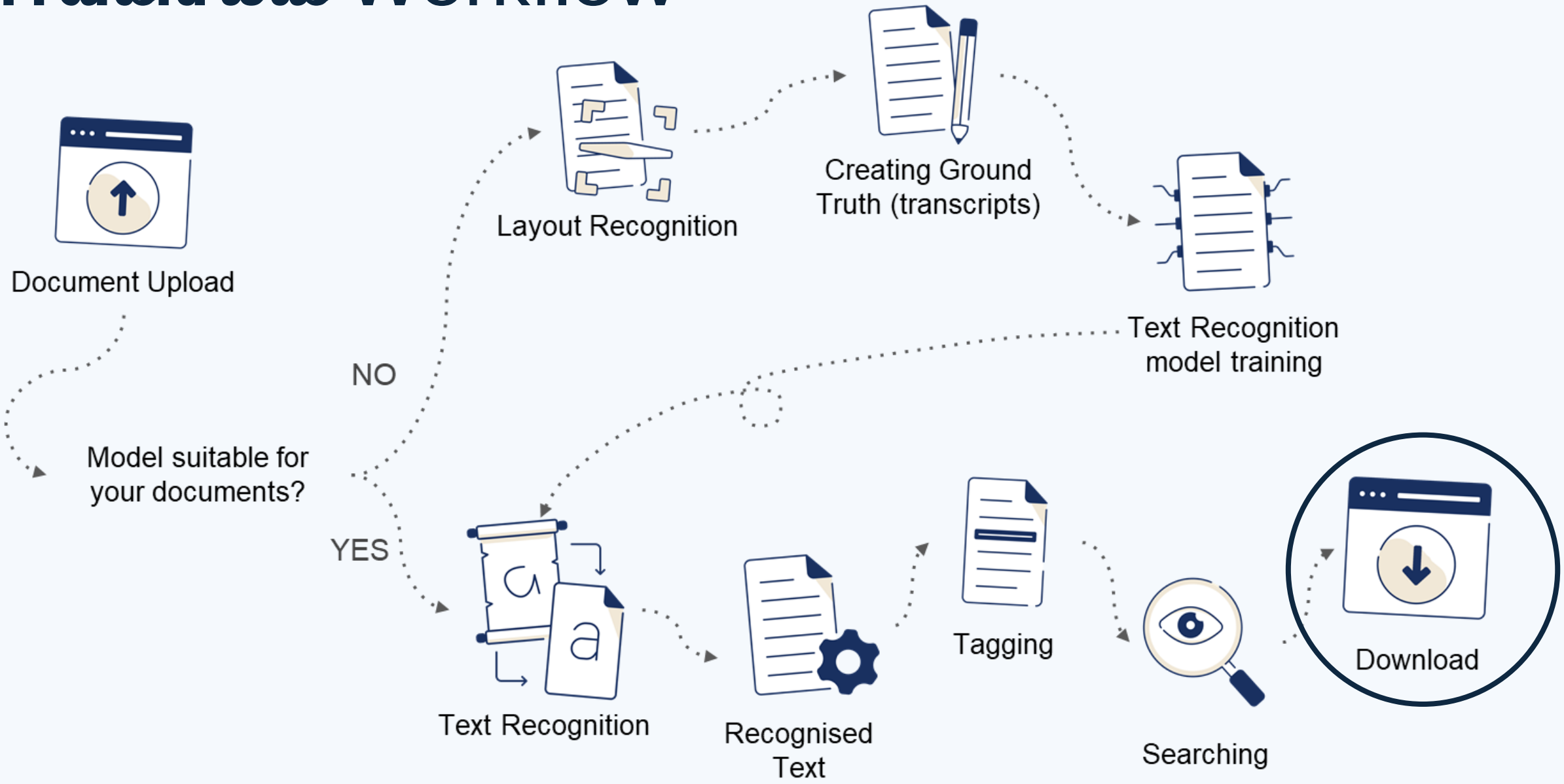
Kassel, 29.01.2025

Exemplar Vatikanstadt 586 - Book 13

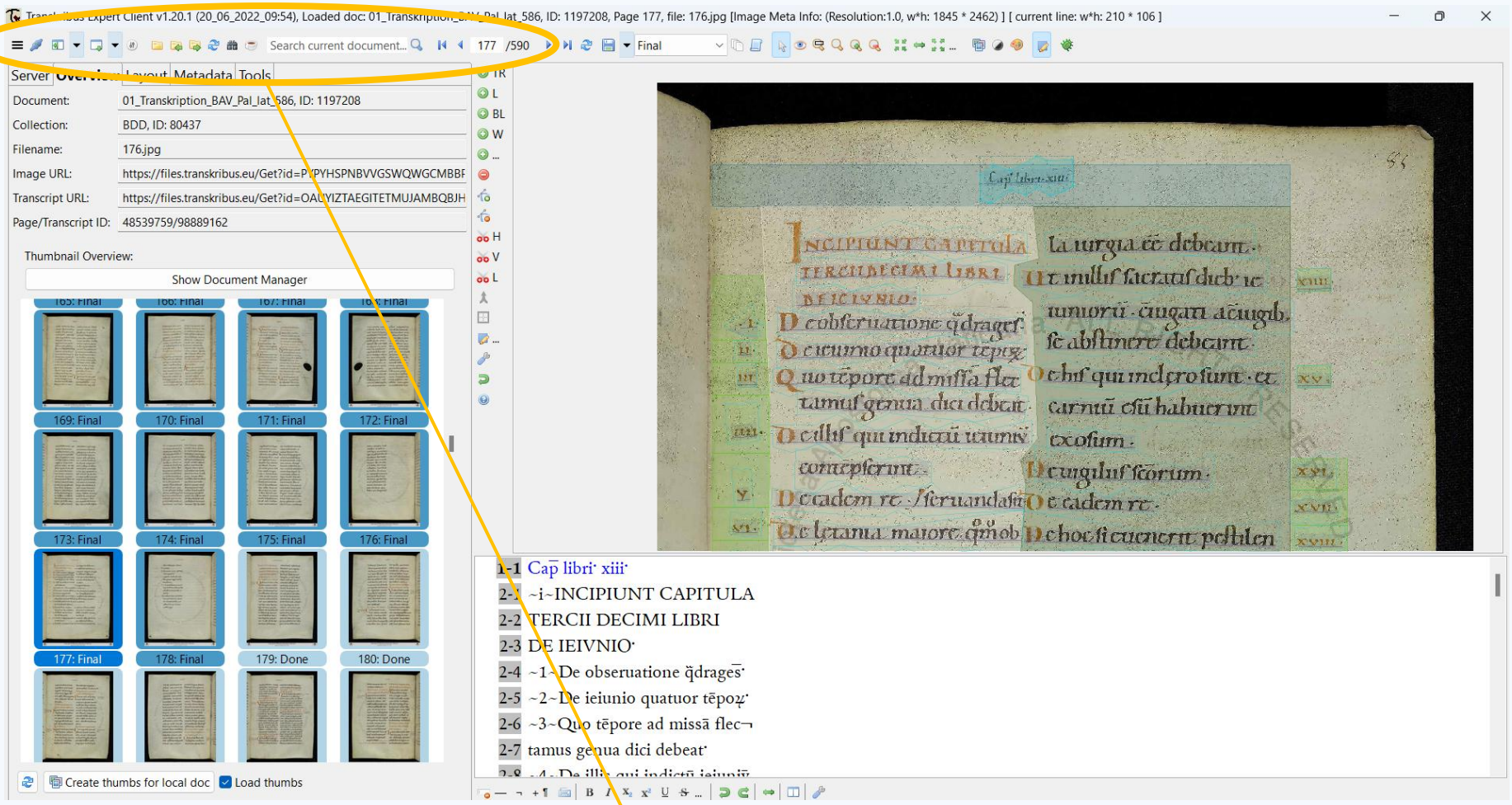
Transkribus Workflow

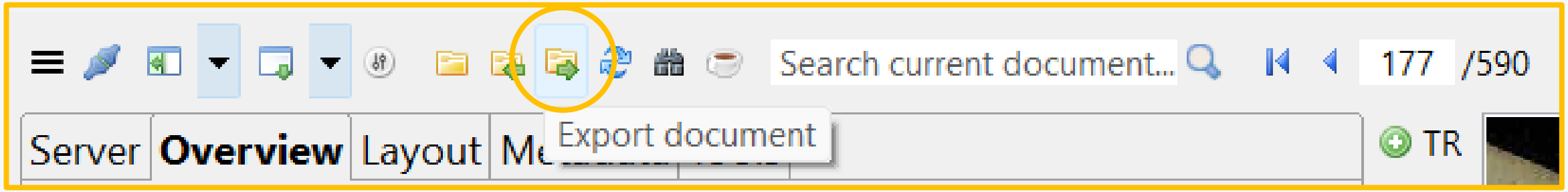


Transkribus Workflow

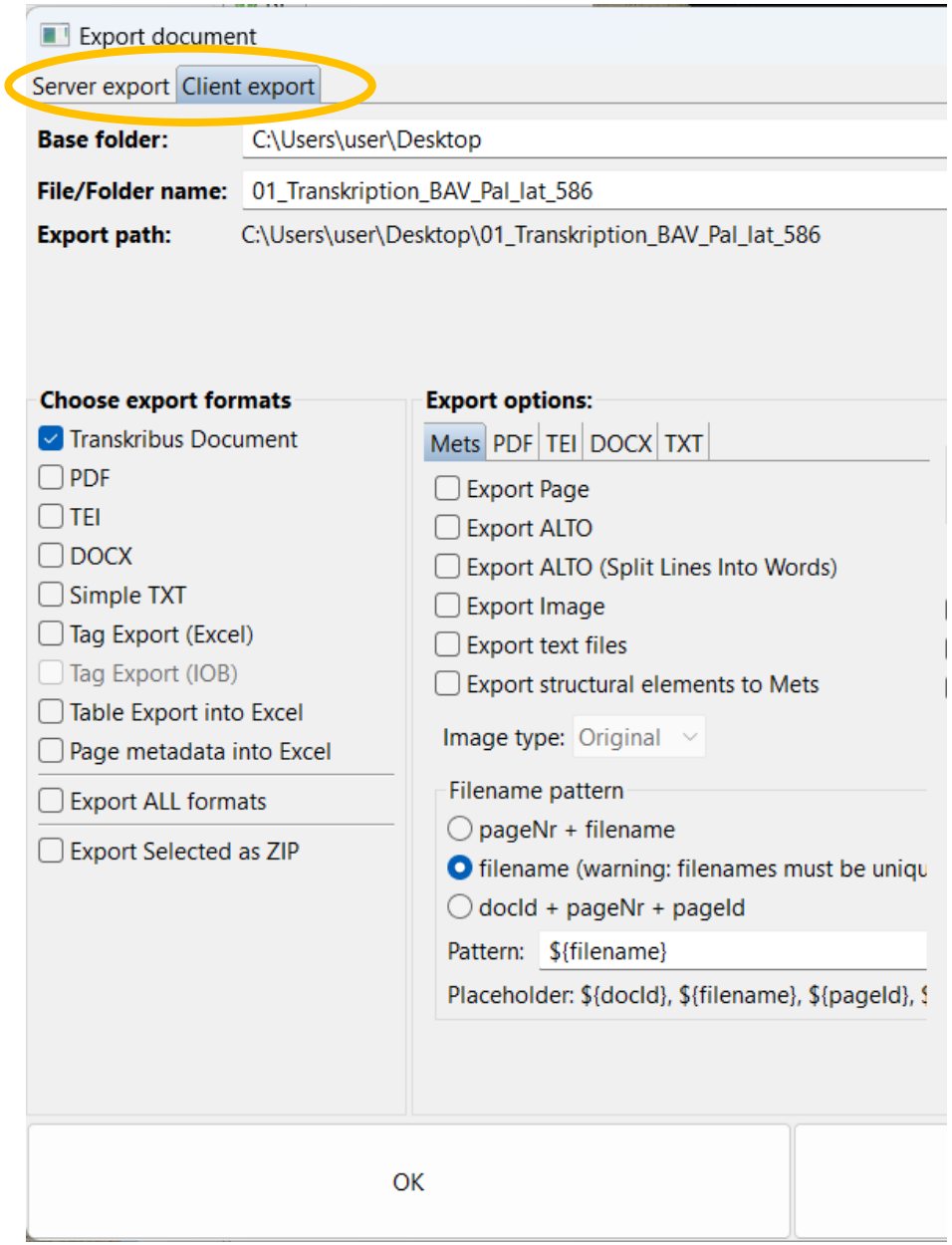


Export Document Button





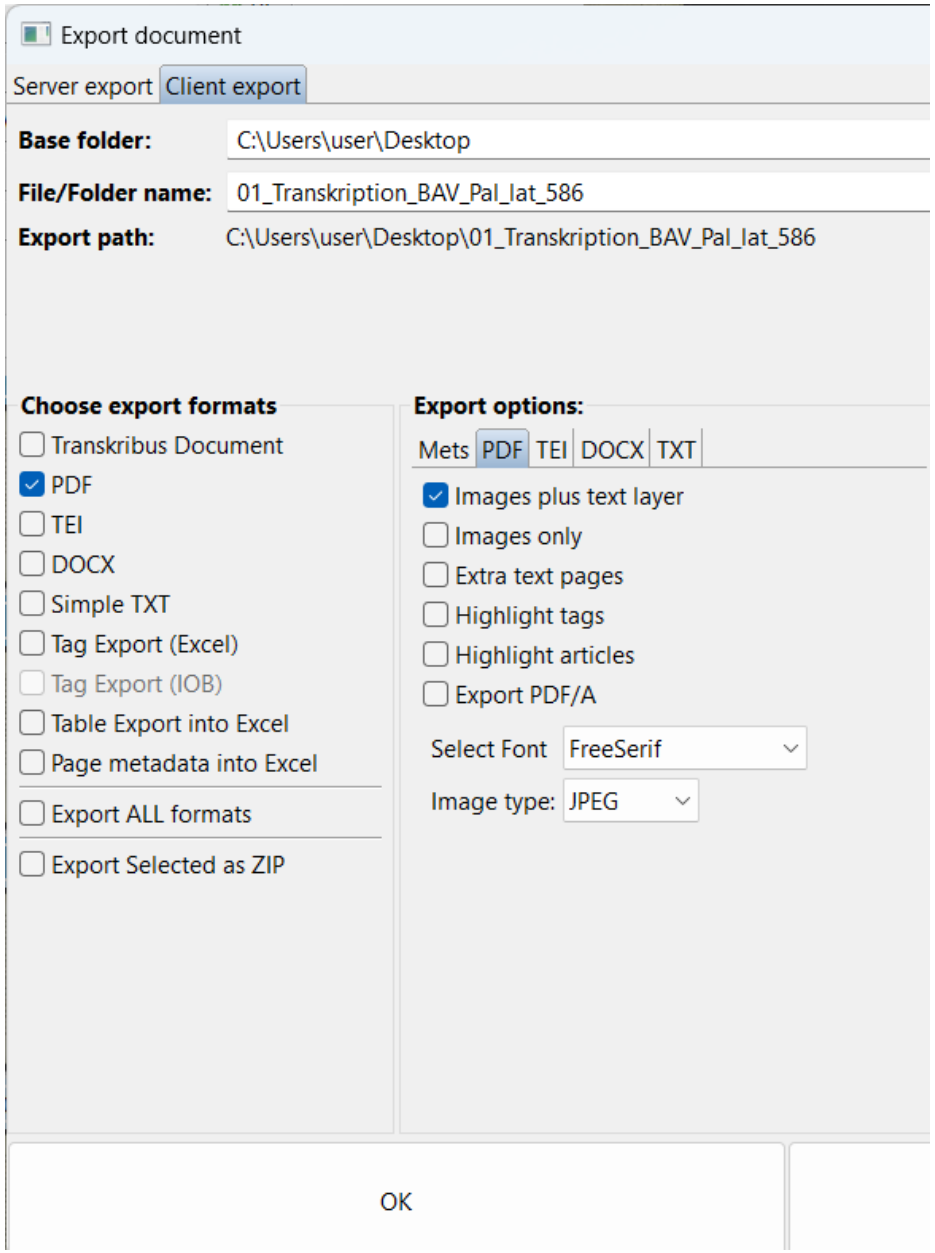
https://help.transkribus.org/downloading



Transkribus Document – Mets (Metadata Encoding and Transmission Standard)

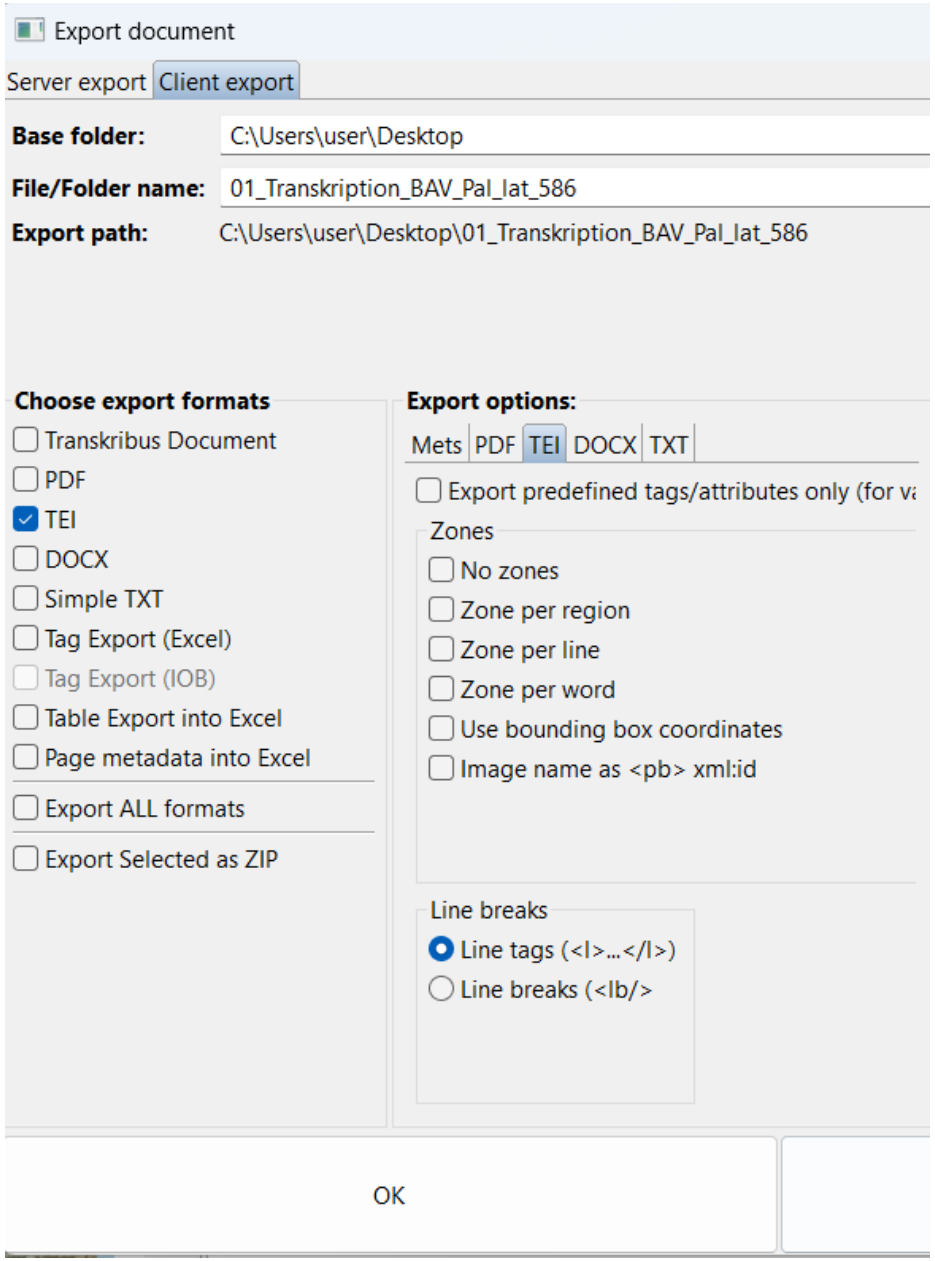
- Page
- ALTO (Analyzed Layout and Text Object)
- Image
- Text files (from each page)
- Structural elements to Mets (metadata.xml)

Mets https://www.loc.gov/standards/mets/
ALTO https://www.loc.gov/standards/alto/



PDF

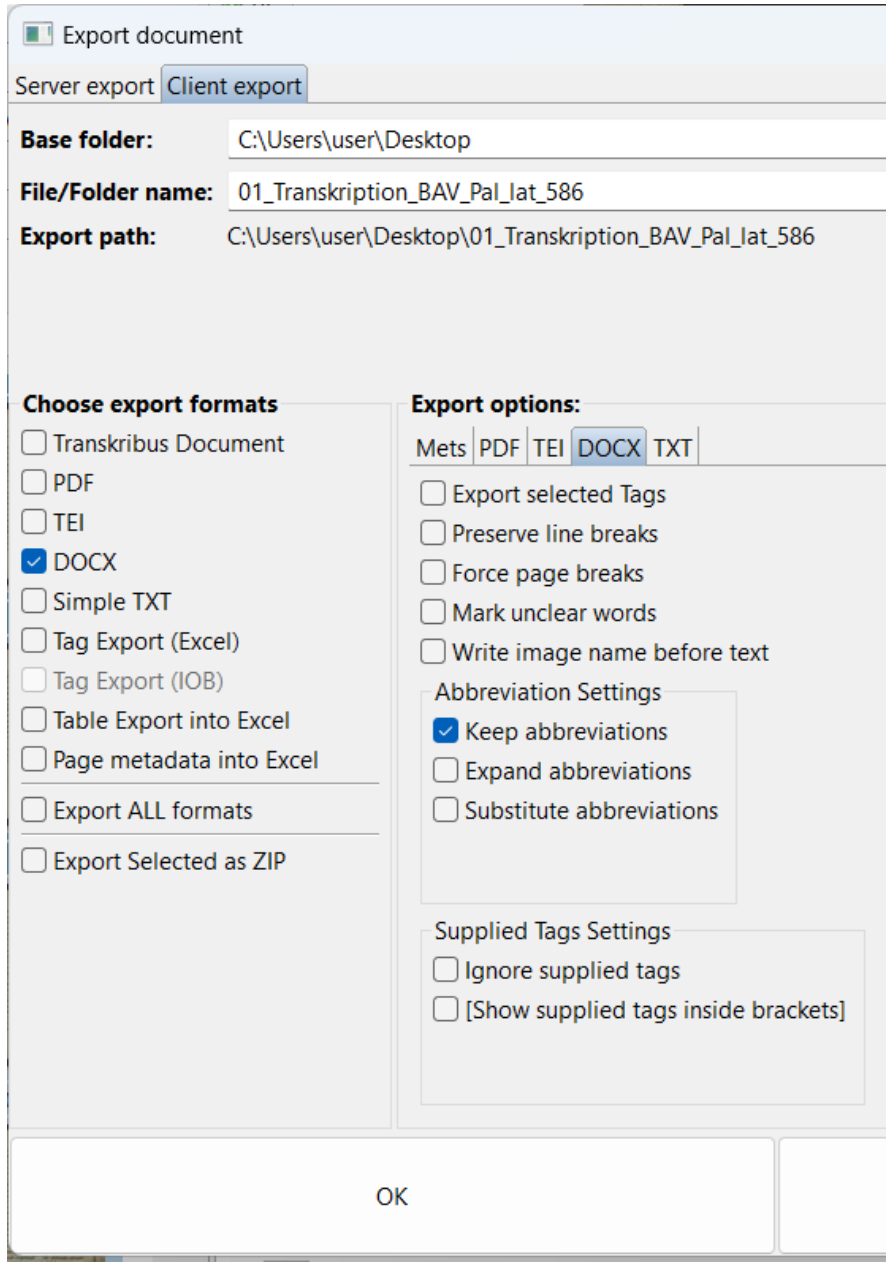
- Images plus text layer (invisible text layer for searching)
- Images only
- Extra Text pages
- Highlight tags
- Highlight articles
- PDF/A (long-term preservation)



TEI (Text Encoding Initiative)

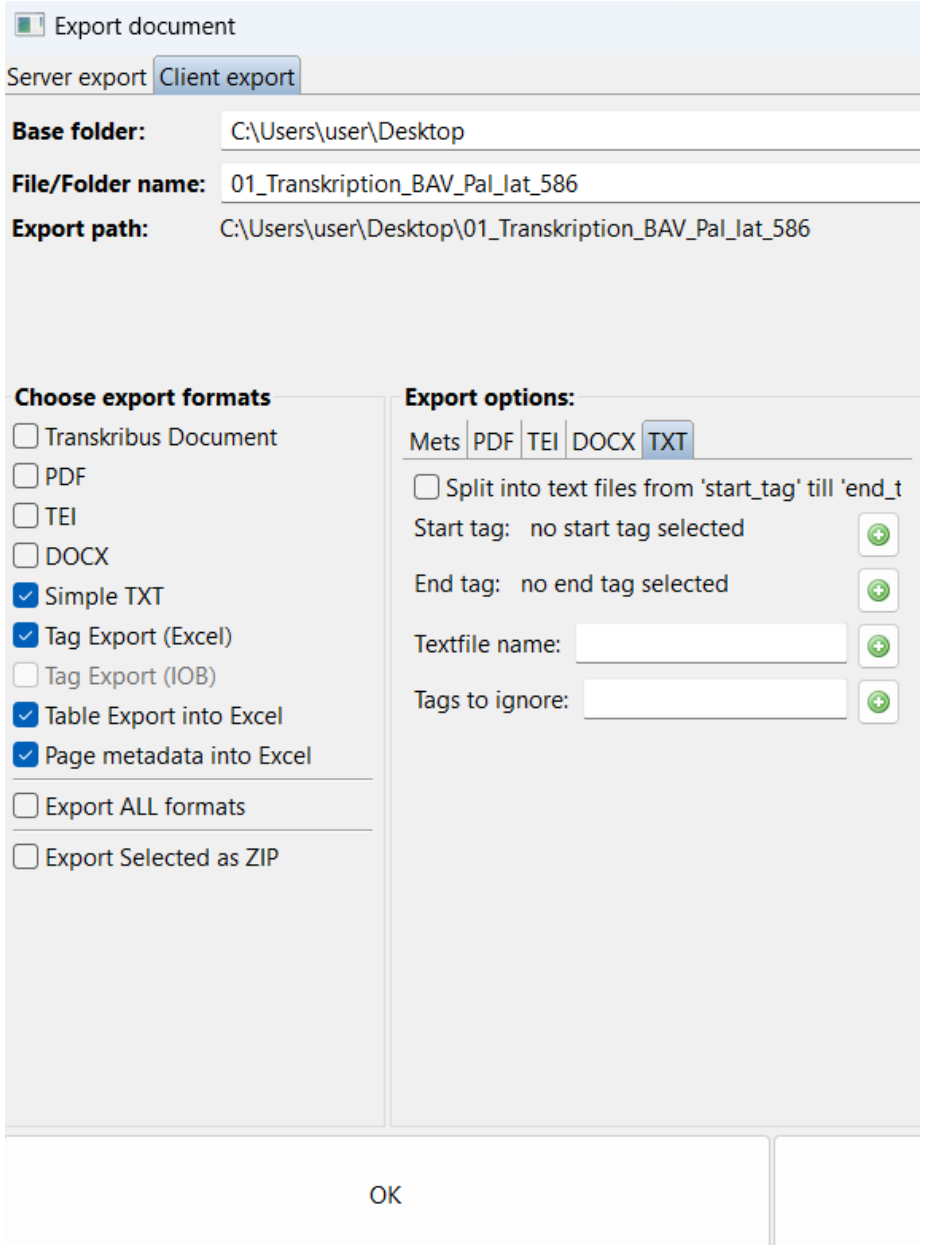
- No zones (text only, no layout information)
- Zone per region (e.g., paragraphs or sections).
- Zone per line
- Zone per word
- Use bounding box coordinates (coordinates for zones, lines, or words.)
- Image name as <pb> xml:id (page break identifier)

TEI https://tei-c.org/



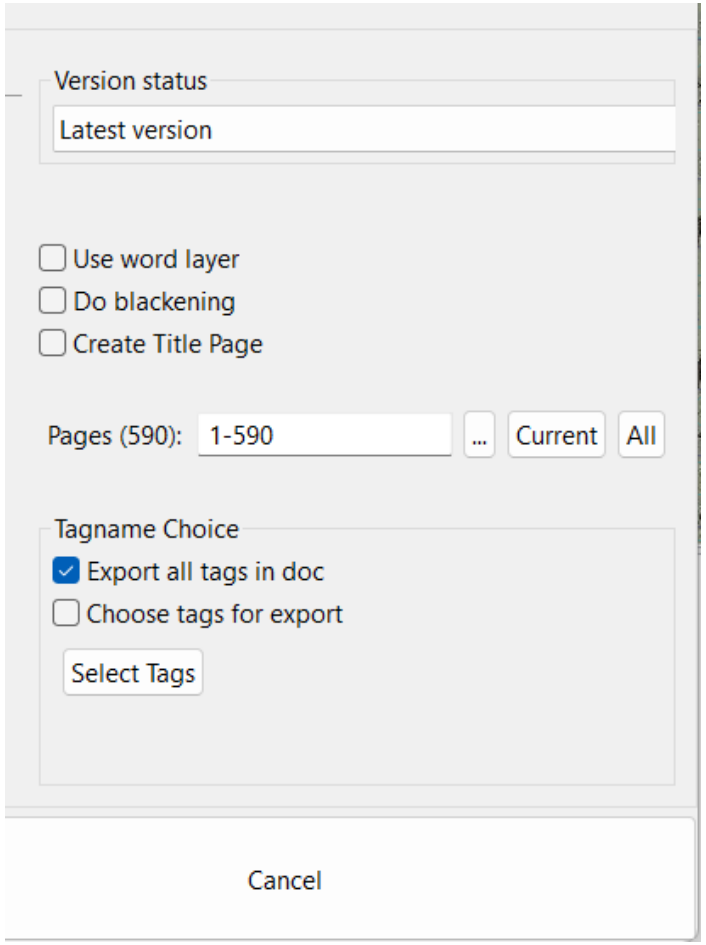
DOCX

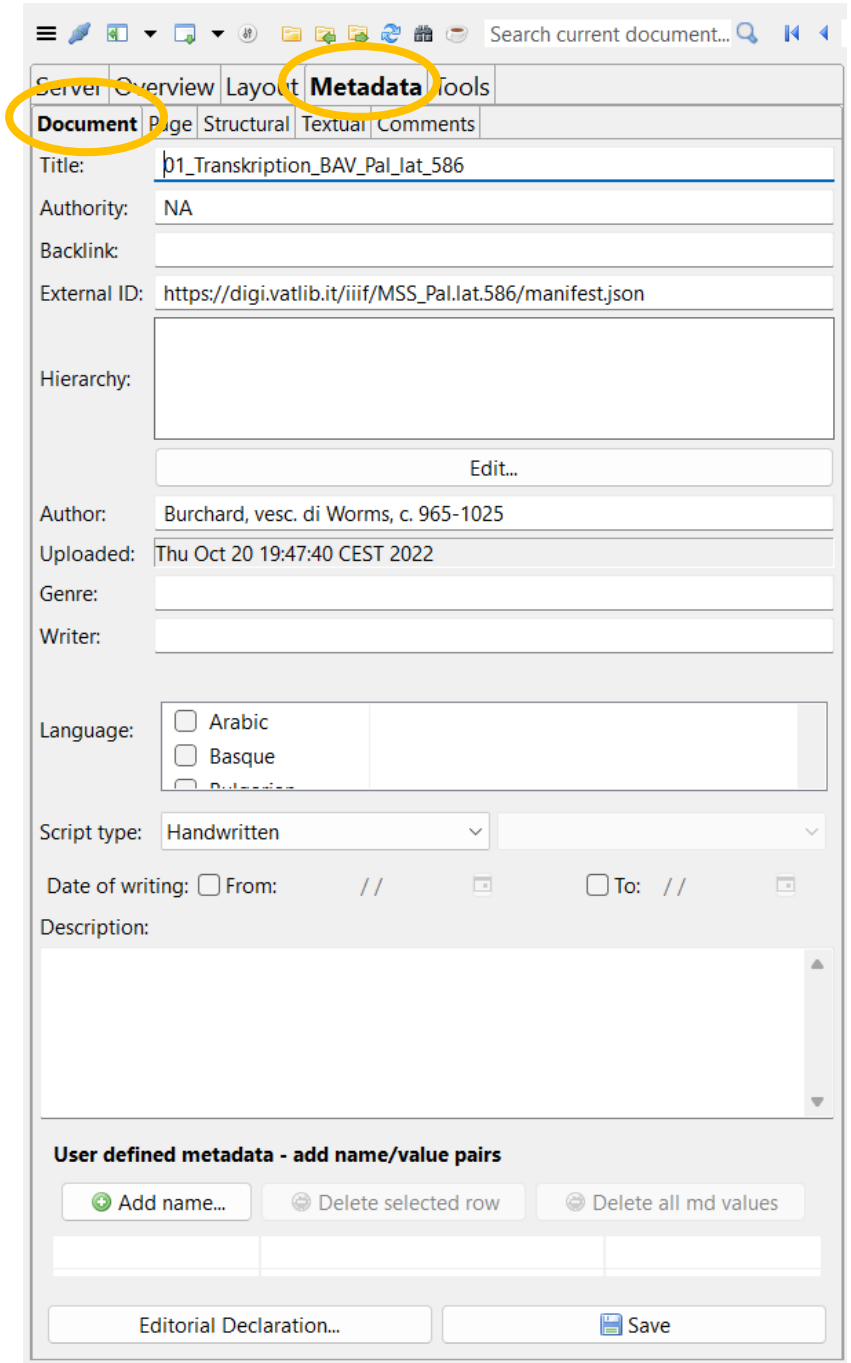
- Selected Tags
- Preserve line breaks
- Force page breaks
- Mark unclear words
- Write image name before text



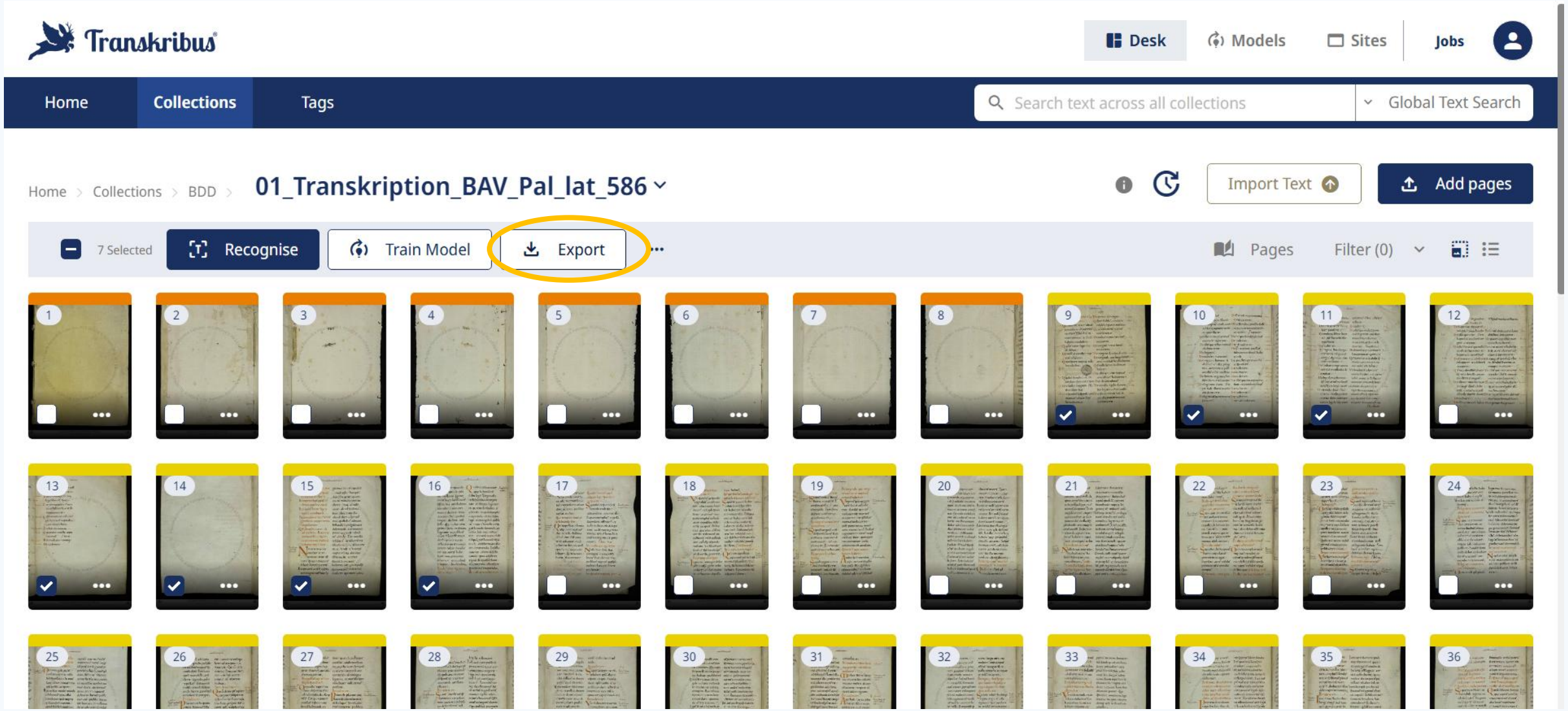
OTHER FORMATS

- Simple TXT
- Tag Export (Excel)
- Table Export into Excel
- Page metadata into Excel


Version status
Latest version
Use word layer
Do blackening
Create Title Page
Pages (590):
1-590
...
Current
All
Tagname Choice
Export all tags in doc
Choose tags for export
Select Tags
Cancel

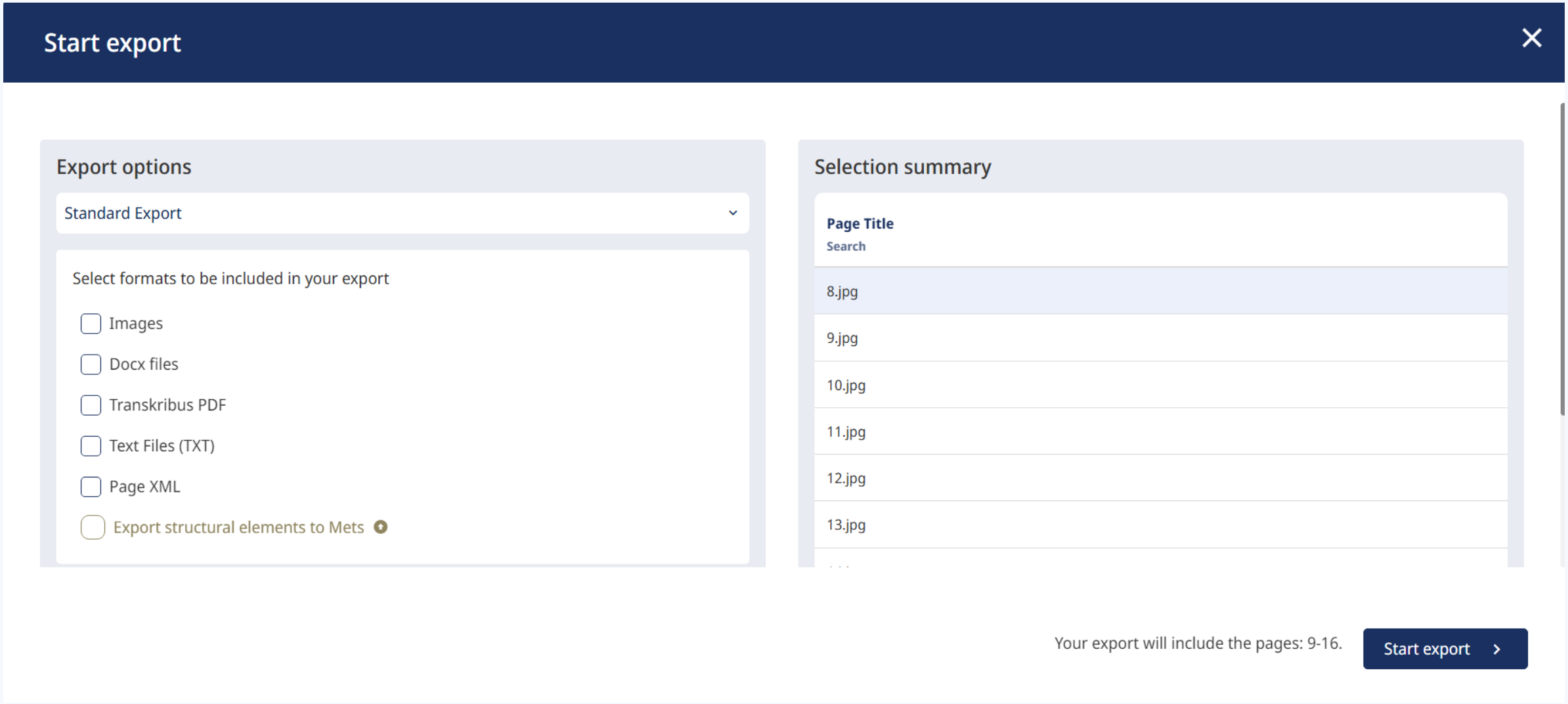

Search current document...
Server
Overview
Layout
Metadata
Tools
Document
Page
Structural
Textual
Comments
Title:
01_Transkription_BAV_Pal_lat_586
Authority:
NA
Backlink:
External ID:
https://digi.vatlib.it/iiif/MSS_Pal.lat.586/manifest.json
Hierarchy:
Edit...
Author:
Burchard, vesc. di Worms, c. 965-1025
Uploaded:
Thu Oct 20 19:47:40 CEST 2022
Genre:
Writer:
Language:
Arabic
Basque
Script type:
Handwritten
Date of writing:
From:
To:
Description:
User defined metadata - add name/value pairs
Add name...
Delete selected row
Delete all md values
Editorial Declaration...
Save

Transkribus Update: Web App

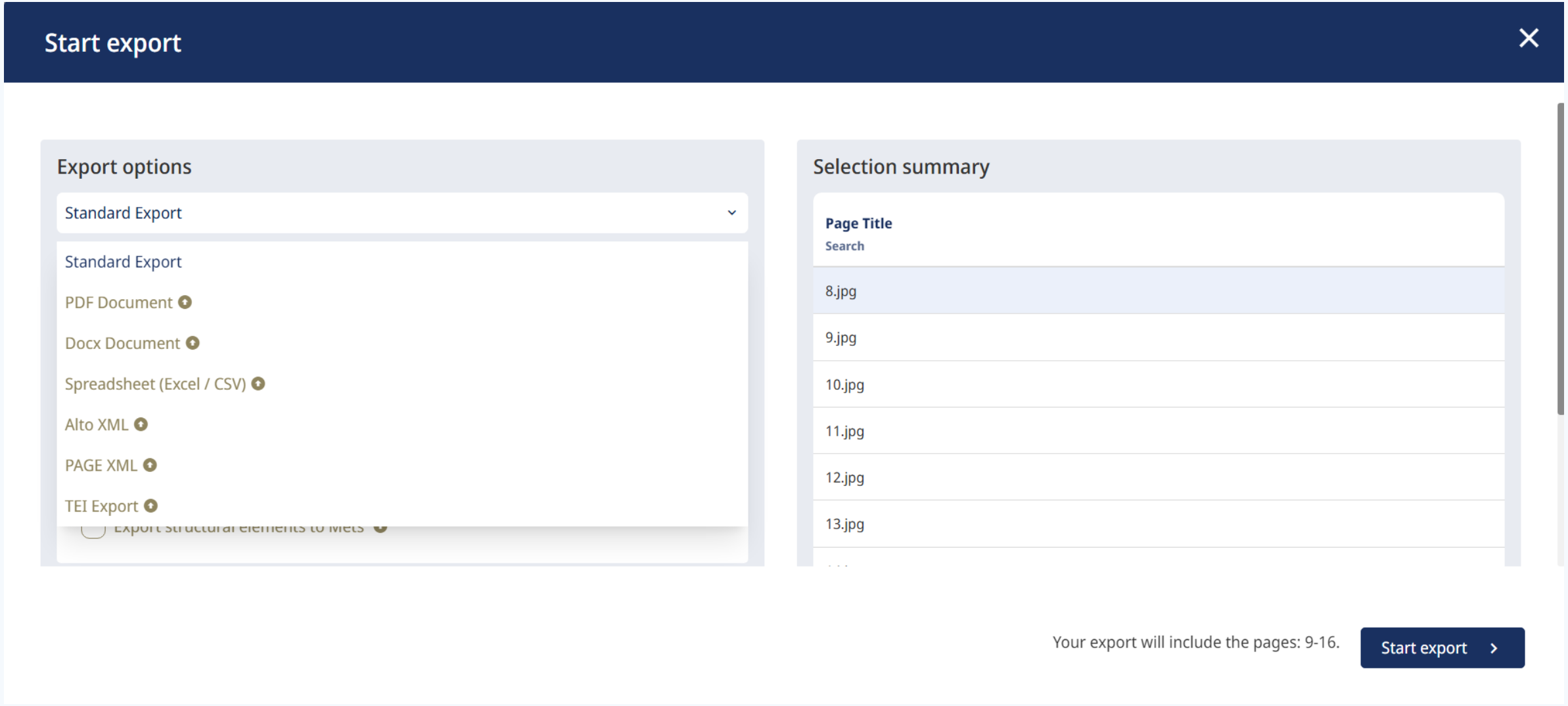


https://readcoop.eu/introducing-the-new-transkribus-web-app

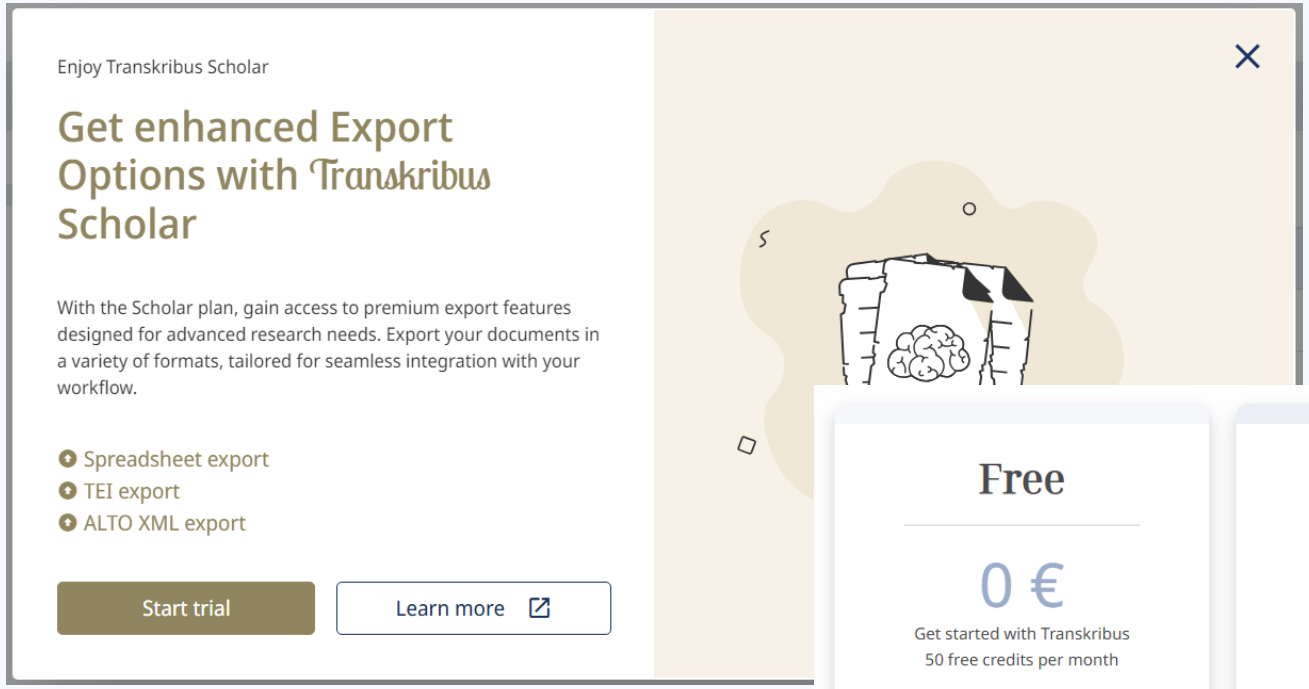
Transkribus Update: Web App

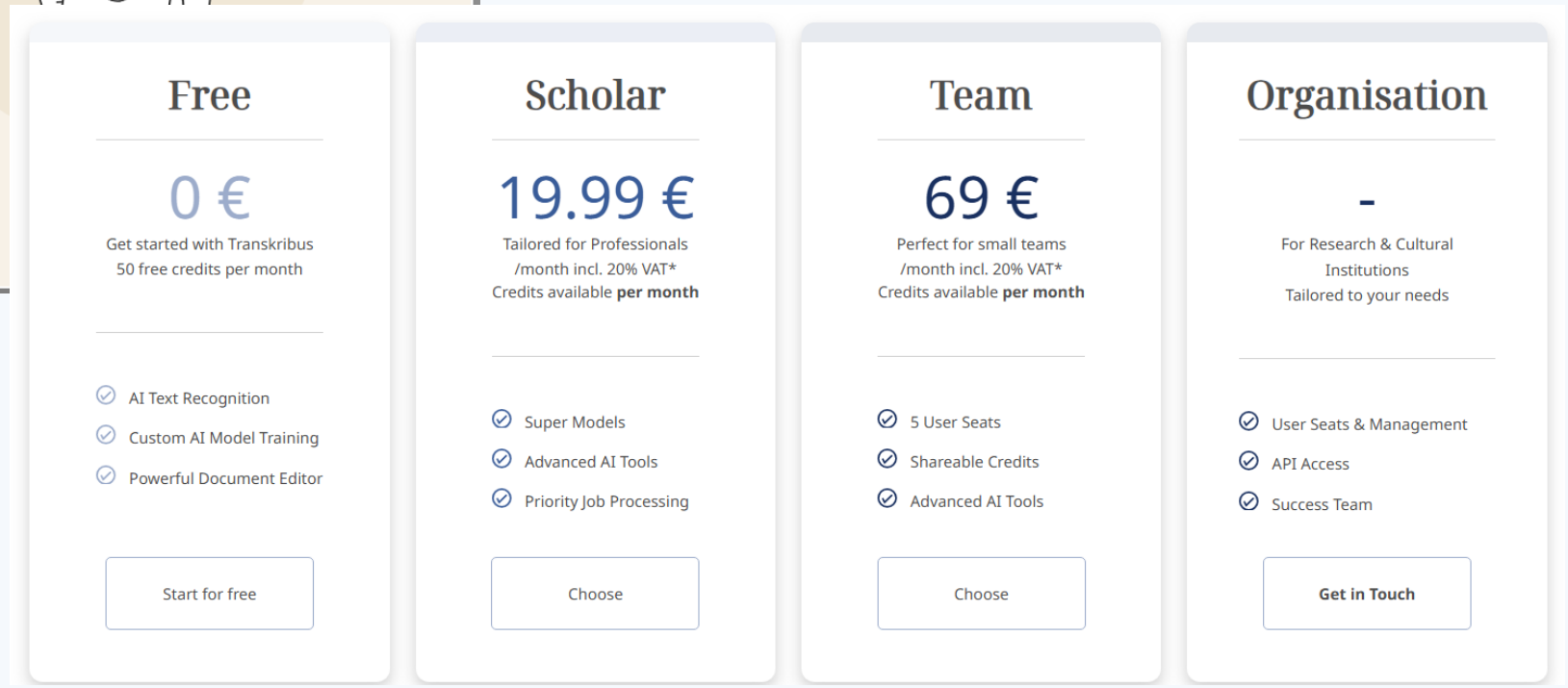


Transkribus Update: Web App



Transkribus Update: Plans and Pricing

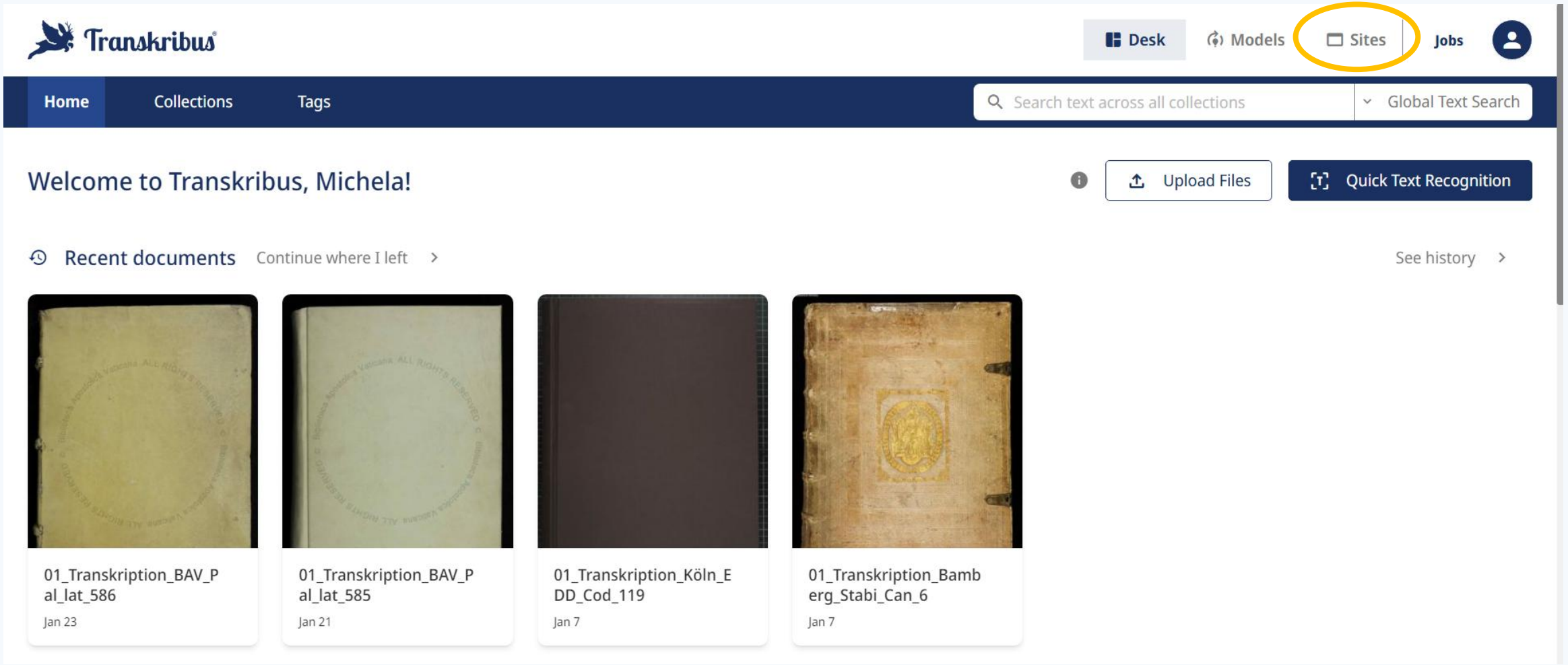




https://www.transkribus.org/plans

Getting started with Transkribus Web App
https://youtube.com/playlist?list=PL7UbQtd4qlhIMP1KfdjGW3C-KXTxw4KYb&si=udvwEPE7XXwU6VJq

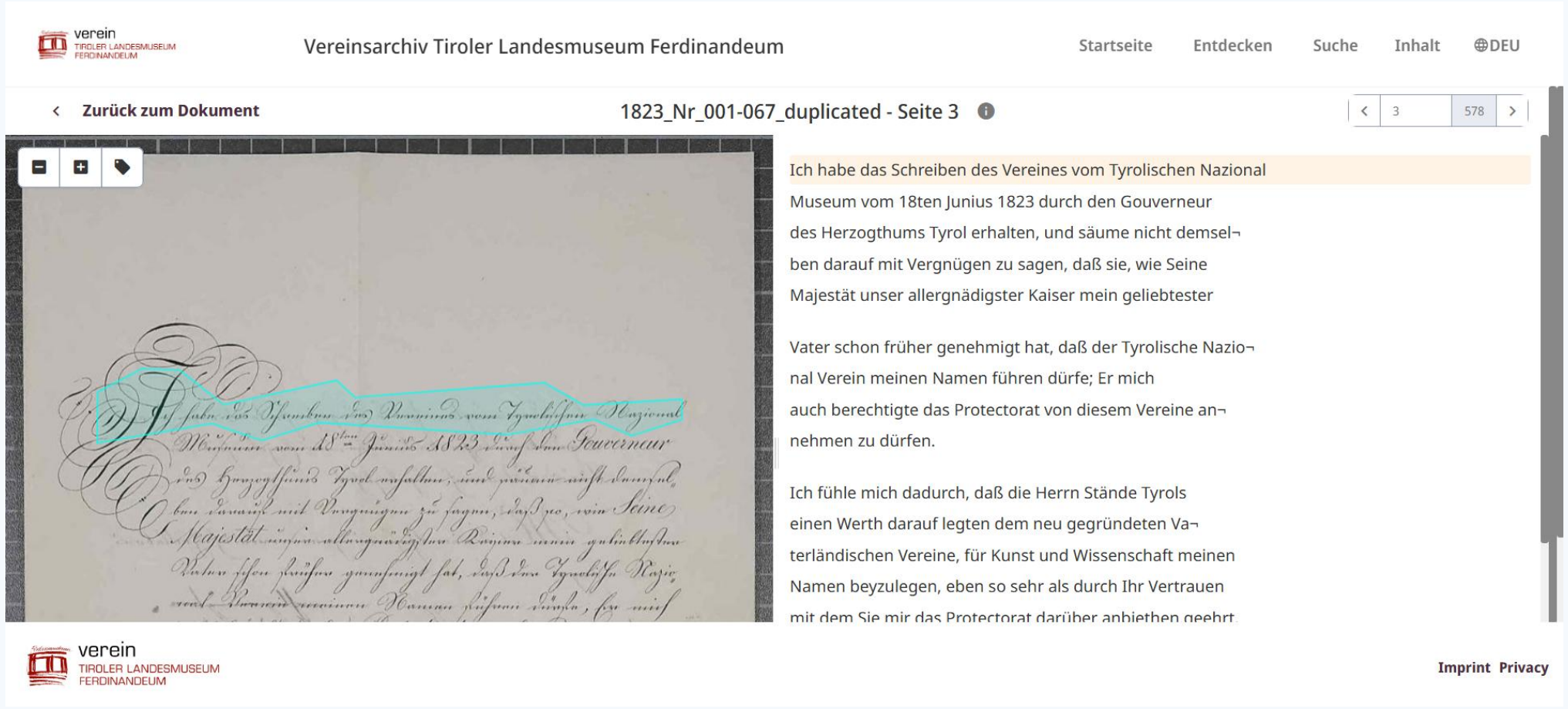
Transkribus Sites



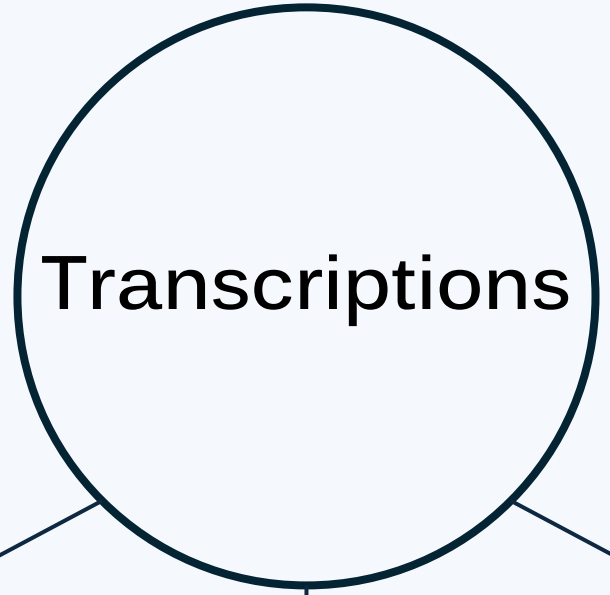
https://www.transkribus.org/sites

Publish documents with Transkribus Sites https://www.youtube.com/watch?v=LlApHRWN6Uo

Transkribus Sites: Vereinsarchiv Tiroler Landesmuseum Ferdinandeum



https://app.transkribus.org/sites/ferdinandeum-archiv



Textual Analysis and Research Applications

- **Text search and linguistic analysis** (concordance analysis, pattern recognition)
- **Philological and historical research** (textual variation, annotation for digital edition, comparative analysis)
- **Text mining and computational analysis** (pattern detection, sentiment analysis, semantic studies)

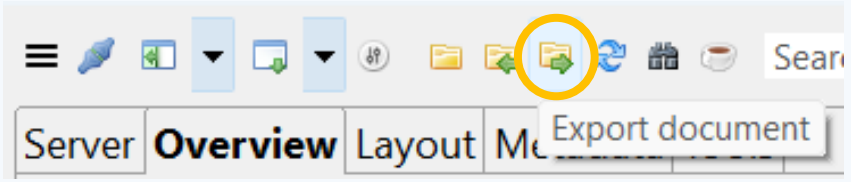
Data visualisation and Presentation

- **Geospatial Analysis:** (QGIS).
- **Creating Visuals from Textual Data** (word clouds, frequency graphs)
- **Graphs and network analysis** (connections between people, events, or concepts)

Machine Learning and AI Applications

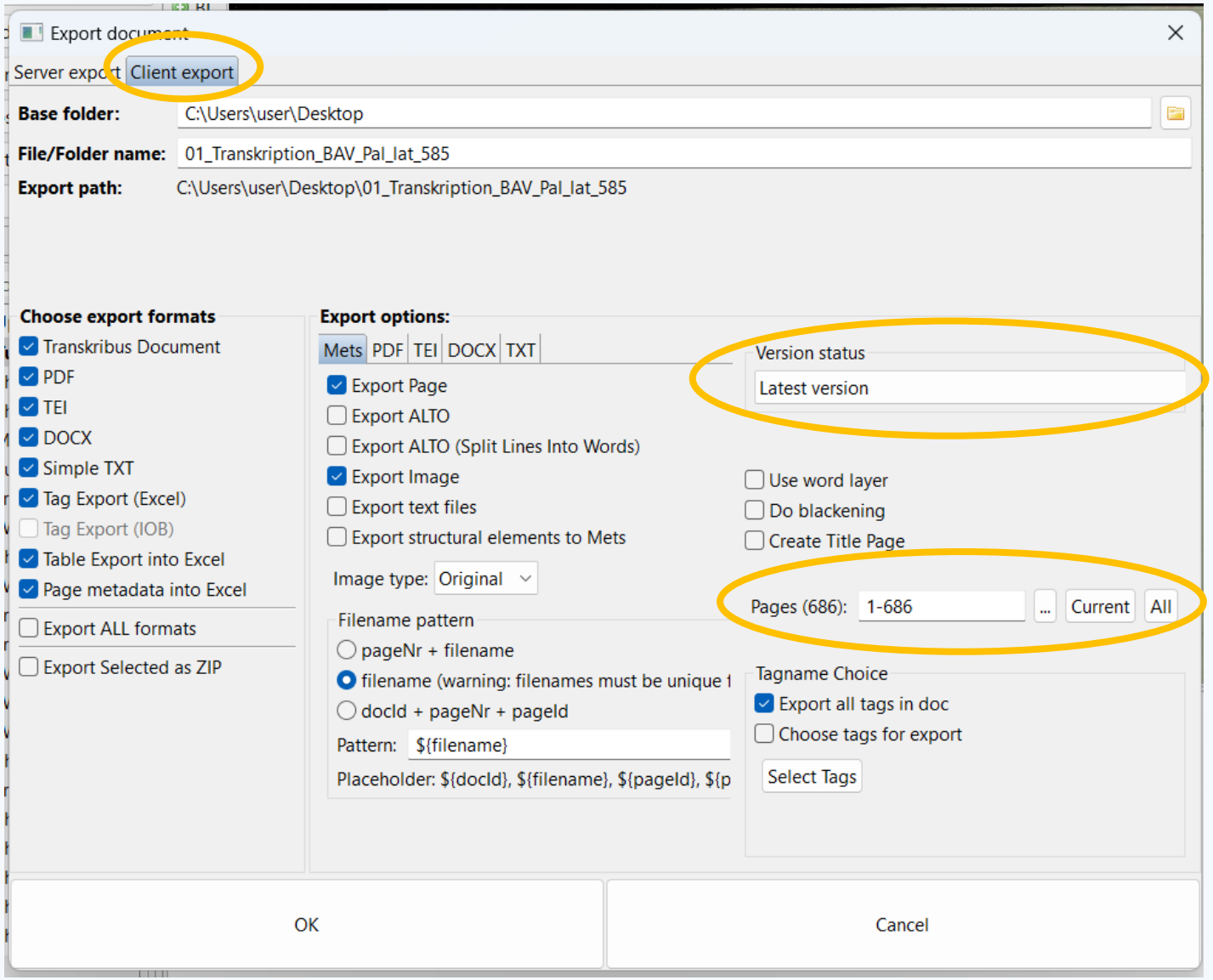
- **Natural Language Processing** (NLP) (linguistic and semantic analysis)
- **Training recognition models or improving recognition**
- **Predictive analysis** (identify patterns or make predictions)

LAB session



Step 1: Export your transcription

- **PAGE XML**: For advanced analysis, semantic annotation, and ground-truth data creation.
- **ALTO XML**: For layout-specific applications or digital repositories.
- **PDF**: To create a readable, searchable document.
- **TEI**: For structured, scholarly analysis.
- **DOCX**: For collaborative editing or sharing.
- **Simple TXT**: For text editing.
- **Tag Export into Excel**: For analysing or editing annotated tags in tabular form.
- **Table into Excel**: For extracting and working with structured table data.
- **Metadata into Excel**: For reviewing, managing, or sharing document metadata in a spreadsheet format.

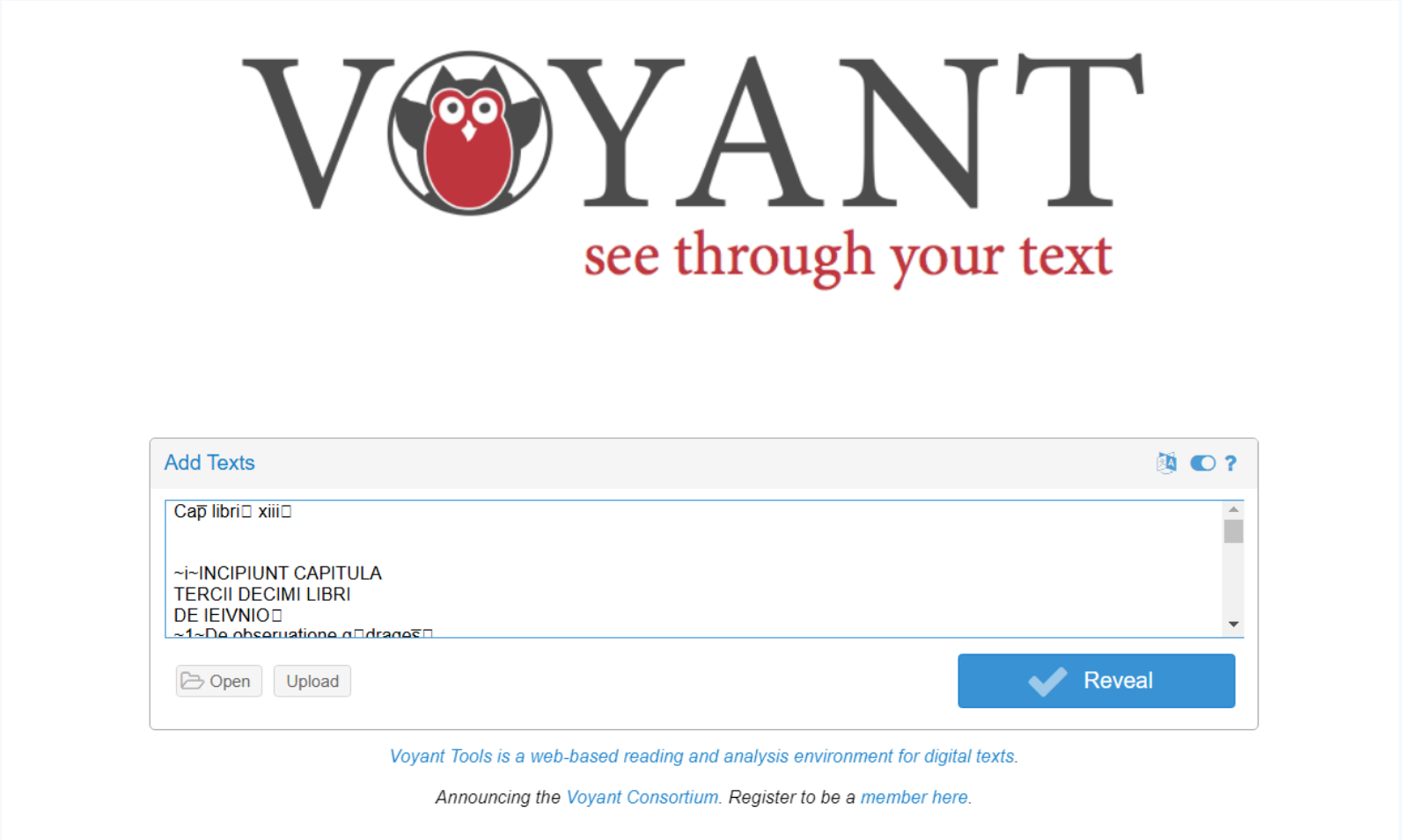


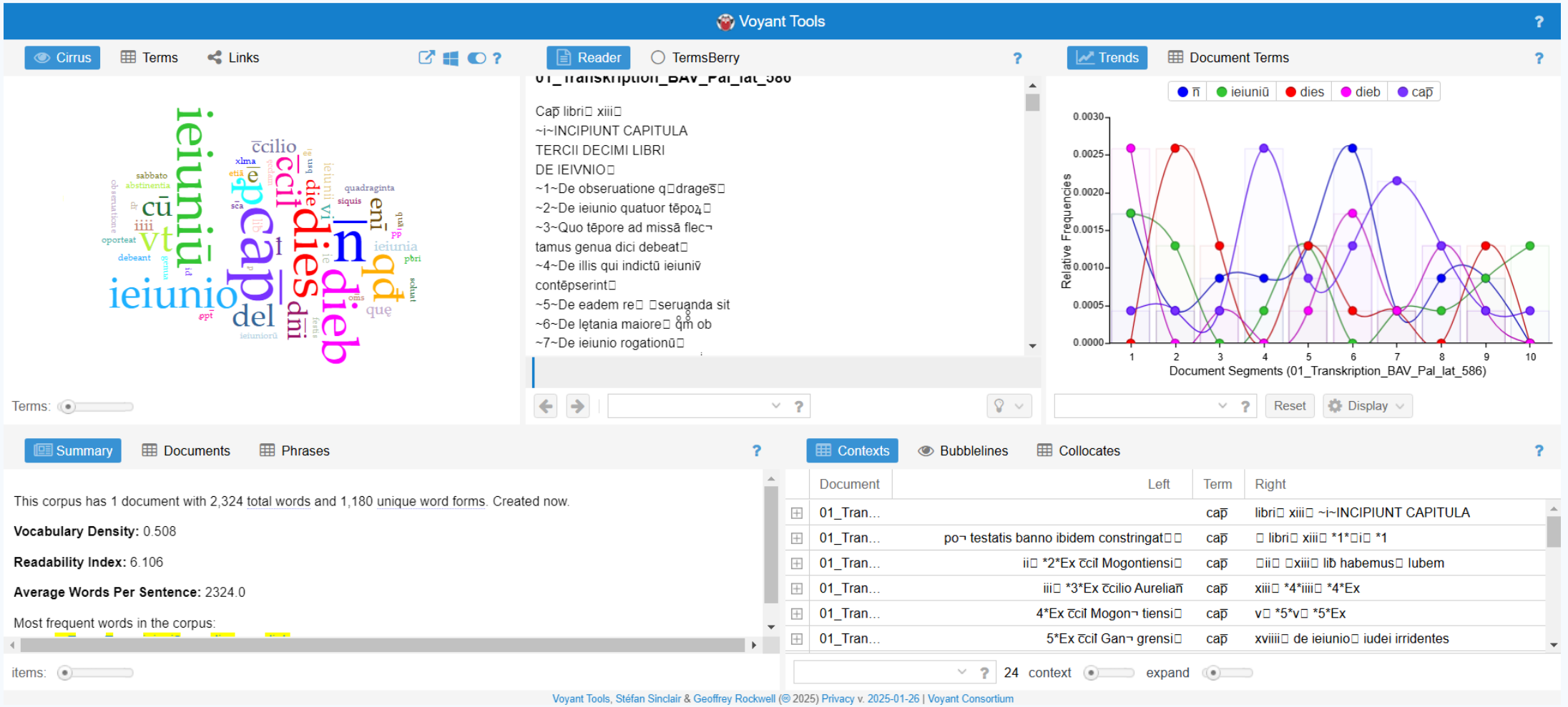
LAB session: optional

Step 2: Export your transcription in simple TXT and open Voyant Tools

https://voyant-tools.org/

Paste the text in the box or upload the TXT document. Click “Reveal”.



LAB session: optional

Step 3: Present what you found out about the text you transcribed.