

Transkribus

Data Export

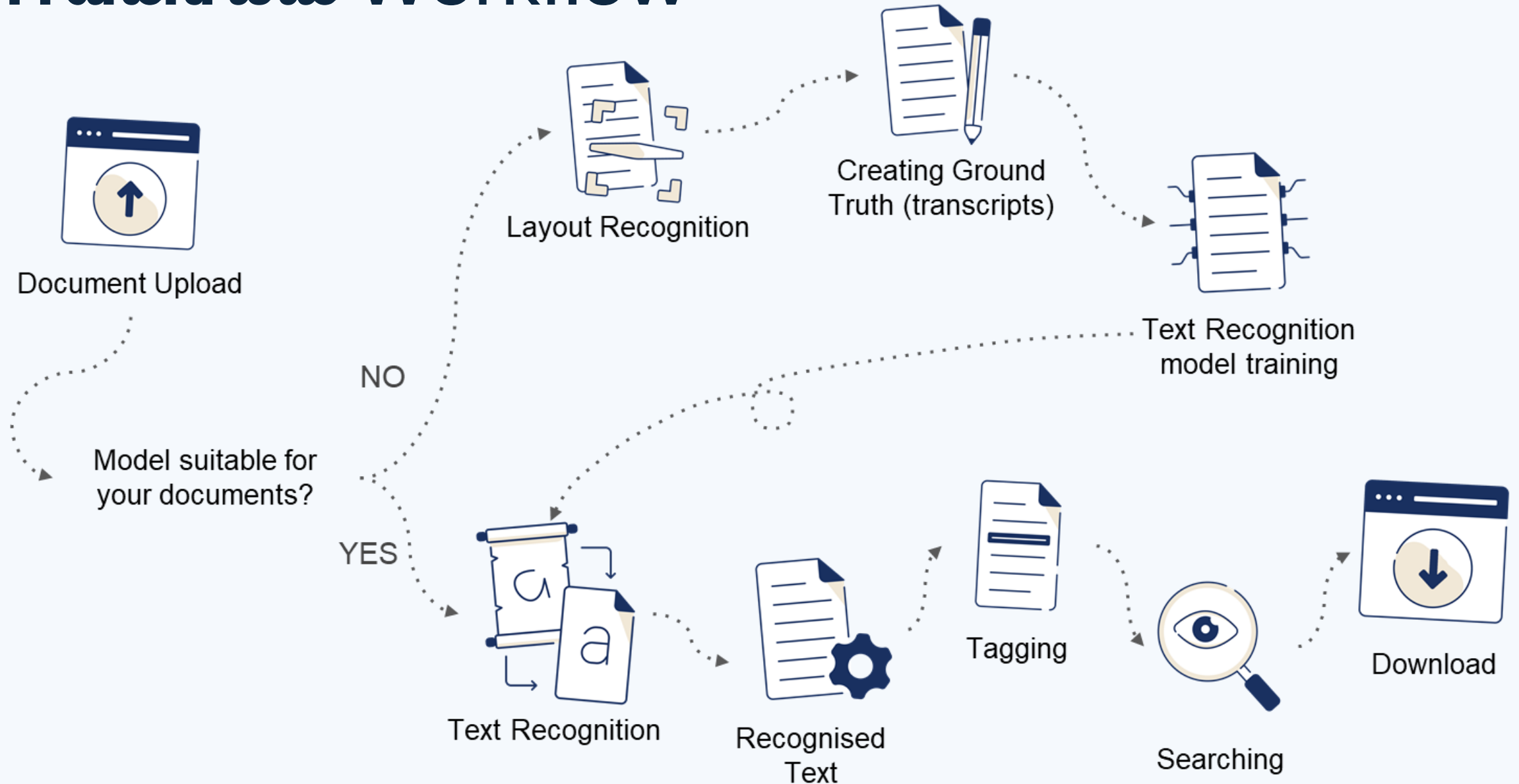
Michela Parma, JGU Mainz Universität

michela.parma@uni-mainz.de

Kassel, 29.01.2025

Exemplar Vatikanstadt 586 - Book 13

Transkribus Workflow



Transkribus Workflow



Transkribus Expert Client v1.20.1 (20_06_2022_09:54), Loaded doc: 01_Transkription_BAV_Pal_lat_586, ID: 1197208, Page 177, file: 176.jpg [Image Meta Info: (Resolution:1.0, w'h: 1845 * 2462)] [current line: w'h: 210 * 106]

Search current document...

177 /590

Server Overview Layout Metadata Tools

Document: 01_Transkription_BAV_Pal_lat_586, ID: 1197208
Collection: BDD, ID: 80437
Filename: 176.jpg
Image URL: https://files.transkribus.eu/Get?id=P...PYHSPNBVGSQWQWGCMBBF
Transcript URL: https://files.transkribus.eu/Get?id=OALYIZTAEGITETMUJAMBQBJH
Page/Transcript ID: 48539759/98889162

Thumbnail Overview:

Show Document Manager

105: Final 106: Final 107: Final 108: Final
169: Final 170: Final 171: Final 172: Final
173: Final 174: Final 175: Final 176: Final
177: Final 178: Final 179: Done 180: Done

Create thumbs for local doc Load thumbs

1-1 Cap libri xiii
2-1 ~i-INCIPIUNT CAPITULA
2-2 TERCII DECIMI LIBRI
2-3 DE IEIVNIO
2-4 ~1-De obseruatione q̄drages
2-5 ~2-De ieunio quatuor tēpoz
2-6 ~3-Quo tēpore ad missā flec
2-7 tamus genua dici debeat
2-8 ~4-De illis qui indicū ieiunū

INCIPIUNT CAPITULA
TERCII DECIMI LIBRI
DE IEIVNIO
De obseruatione q̄drages
De ieunio quatuor tēpoz
Quo tēpore ad missā flec
tamus genua dici debeat
De illis qui indicū ieiunū

la uirgia et debeant
ut in illis sacratiss dieb
uniorū ciugati ac iugib
se abstinere debeant
De his qui in clero sunt et
carnū osu habuerint
excofum
De uigiliis scōrum
De eadem re
De hoc si euenierit postilen

Export Document Button

<https://help.transkribus.org/downloading>

Search current document...

177 /590

Server **Overview** Layout Metadata **Export document**

TR

Export document

Server export Client export

Base folder: C:\Users\user\Desktop

File/Folder name: 01_Transkription_BAV_Pal_Iat_586

Export path: C:\Users\user\Desktop\01_Transkription_BAV_Pal_Iat_586

Choose export formats

☒ Transkribus Document

☐ PDF

☐ TEI

☐ DOCX

☐ Simple TXT

☐ Tag Export (Excel)

☐ Tag Export (IOB)

☐ Table Export into Excel

☐ Page metadata into Excel

☐ Export ALL formats

☐ Export Selected as ZIP

Export options:

Mets PDF TEI DOCX TXT

☐ Export Page

☐ Export ALTO

☐ Export ALTO (Split Lines Into Words)

☐ Export Image

☐ Export text files

☐ Export structural elements to Mets

Image type: Original

Filename pattern

☐ pageNr + filename

☒ filename (warning: filenames must be unique)

☐ docId + pageNr + pageId

Pattern: \${filename}

Placeholder: \${docId}, \${filename}, \${pageId}, \$

OK

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/>

Export document

Server export Client export

Base folder: C:\Users\user\Desktop

File/Folder name: 01_Transkription_BAV_Pal_lat_586

Export path: C:\Users\user\Desktop\01_Transkription_BAV_Pal_lat_586

Choose export formats

- ☐ Transkribus Document
- ☒ PDF
- ☐ TEI
- ☐ DOCX
- ☐ Simple TXT
- ☐ Tag Export (Excel)
- ☐ Tag Export (IOB)
- ☐ Table Export into Excel
- ☐ Page metadata into Excel
- ☐ Export ALL formats
- ☐ Export Selected as ZIP

Export options:

Mets PDF TEI DOCX TXT

- ☒ Images plus text layer
- ☐ Images only
- ☐ Extra text pages
- ☐ Highlight tags
- ☐ Highlight articles
- ☐ Export PDF/A

Select Font: FreeSerif

Image type: JPEG

OK

PDF

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

Export document

Server export Client export

Base folder: C:\Users\user\Desktop

File/Folder name: 01_Transkription_BAV_Pal_Iat_586

Export path: C:\Users\user\Desktop\01_Transkription_BAV_Pal_Iat_586

Choose export formats

☐ Transkribus Document

☐ PDF

☒ TEI

☐ DOCX

☐ Simple TXT

☐ Tag Export (Excel)

☐ Tag Export (IOB)

☐ Table Export into Excel

☐ Page metadata into Excel

☐ Export ALL formats

☐ Export Selected as ZIP

Export options:

Mets PDF **TEI** DOCX TXT

☐ Export predefined tags/attributes only (for v:

Zones

☐ No zones

☐ Zone per region

☐ Zone per line

☐ Zone per word

☐ Use bounding box coordinates

☐ Image name as <pb> xml:id

Line breaks

☒ Line tags (<l>...</l>)

☐ Line breaks (<lb>)

OK

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/>

Export document

Server export Client export

Base folder: C:\Users\user\Desktop

File/Folder name: 01_Transkription_BAV_Pal_lat_586

Export path: C:\Users\user\Desktop\01_Transkription_BAV_Pal_lat_586

Choose export formats

- ☐ Transkribus Document
- ☐ PDF
- ☐ TEI
- ☒ DOCX
- ☐ Simple TXT
- ☐ Tag Export (Excel)
- ☐ Tag Export (IOB)
- ☐ Table Export into Excel
- ☐ Page metadata into Excel
- ☐ Export ALL formats
- ☐ Export Selected as ZIP

Export options:

Mets PDF TEI **DOCX** TXT

- ☐ Export selected Tags
- ☐ Preserve line breaks
- ☐ Force page breaks
- ☐ Mark unclear words
- ☐ Write image name before text

Abbreviation Settings

- ☒ Keep abbreviations
- ☐ Expand abbreviations
- ☐ Substitute abbreviations

Supplied Tags Settings

- ☐ Ignore supplied tags
- ☐ [Show supplied tags inside brackets]

OK

DOCX

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

Export document

Server export Client export

Base folder: C:\Users\user\Desktop

File/Folder name: 01_Transkription_BAV_Pal_Iat_586

Export path: C:\Users\user\Desktop\01_Transkription_BAV_Pal_Iat_586

Choose export formats

- ☐ Transkribus Document
- ☐ PDF
- ☐ TEI
- ☐ DOCX
- ☒ Simple TXT
- ☒ Tag Export (Excel)
- ☐ Tag Export (IOB)
- ☒ Table Export into Excel
- ☒ Page metadata into Excel
- ☐ Export ALL formats
- ☐ Export Selected as ZIP

Export options:

Mets PDF TEI DOCX **TXT**

☐ Split into text files from 'start_tag' till 'end_t'

Start tag: no start tag selected +

End tag: no end tag selected +

Textfile name: +

Tags to ignore: +

OK

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

Server Overview Layout **Metadata** Tools

Document Page Structural Textual Comments

Title: D1_Transkription_BAV_Pal_lat_586

Authority: NA

Backlink:

External ID: https://digi.vatlib.it/iiiif/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 ☐ Portuguese

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

The screenshot displays the Transkribus web application interface. At the top, the Transkribus logo is on the left, and navigation links for Desk, Models, Sites, and Jobs are on the right. Below this is a dark blue navigation bar with Home, Collections, and Tags. A search bar on the right of this bar says "Search text across all collections" and "Global Text Search".

The main content area shows a breadcrumb trail: Home > Collections > BDD > 01_Transkription_BAV_Pal_lat_586. To the right of the breadcrumb are buttons for "Import Text" and "Add pages". Below the breadcrumb is a toolbar with buttons for "7 Selected", "Recognise", "Train Model", and "Export" (which is circled in yellow). To the right of the toolbar are buttons for "Pages", "Filter (0)", and a view toggle.

The main area displays a grid of 36 manuscript page thumbnails, numbered 1 to 36. Pages 1-8 are orange, and pages 9-36 are yellow. Pages 9, 10, 11, 13, 14, 15, 16, and 17 have a blue checkmark in the bottom left corner, indicating they are selected.

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

Transkribus Update: Web App

Start export



Export options

Standard Export



Select formats to be included in your export

- ☐ Images
- ☐ Docx files
- ☐ Transkribus PDF
- ☐ Text Files (TXT)
- ☐ Page XML
- ☐ Export structural elements to Mets ⓘ

Selection summary

Page Title

Search

8.jpg

9.jpg

10.jpg

11.jpg

12.jpg

13.jpg

...

Your export will include the pages: 9-16.

Start export >

Transkribus Update: Web App

Start export

Export options

Standard Export

Standard Export

PDF Document

Docx Document

Spreadsheet (Excel / CSV)

Alto XML

PAGE XML

TEI Export

☐ Export structural elements to mets

Selection summary

Page Title

Search

8.jpg

9.jpg

10.jpg

11.jpg

12.jpg

13.jpg

...

Your export will include the pages: 9-16.

Start export >


Transkribus Update: Plans and Pricing


Enjoy Transkribus Scholar

Get enhanced Export Options with Transkribus Scholar

With the Scholar plan, gain access to premium export features designed for advanced research needs. Export your documents in a variety of formats, tailored for seamless integration with your workflow.

- 🔗 Spreadsheet export
- 🔗 TEI export
- 🔗 ALTO XML export

Start trialLearn more 

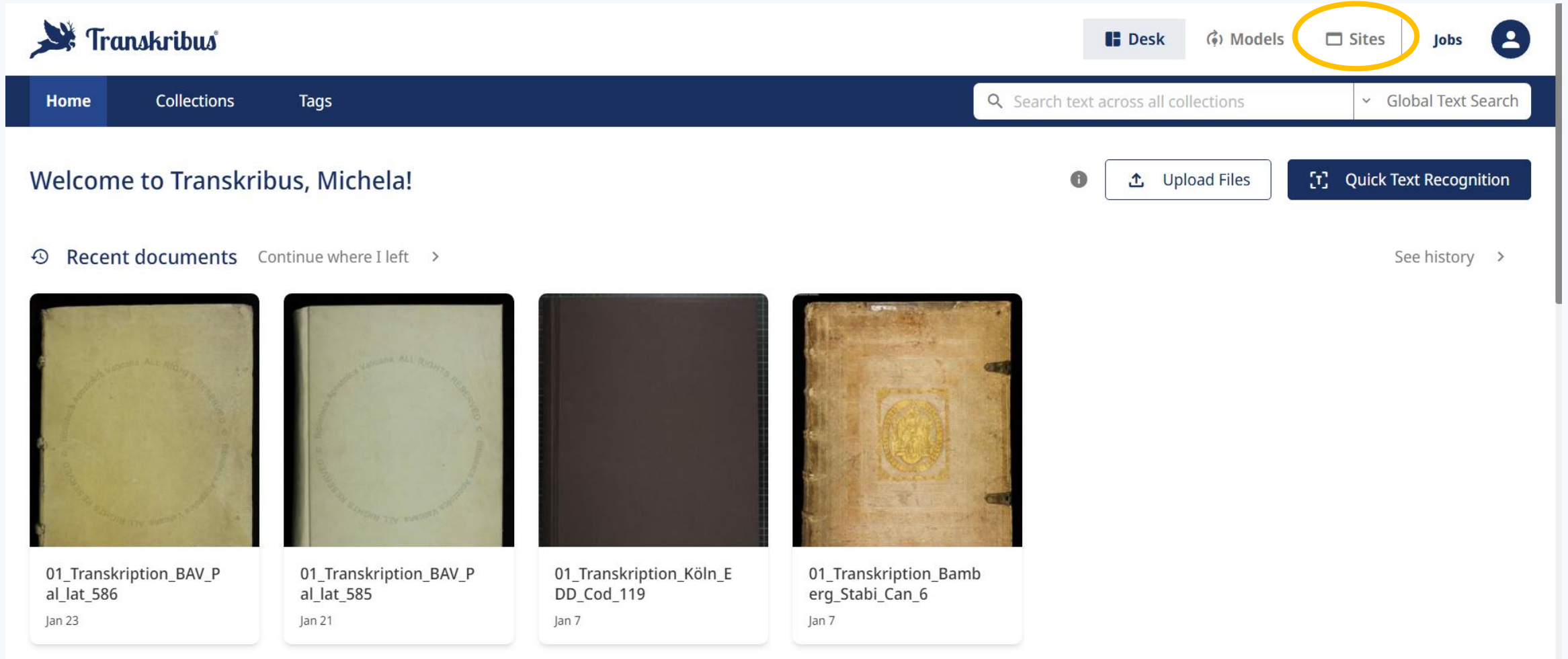


Free	Scholar	Team	Organisation
0 €	19.99 €	69 €	-
Get started with Transkribus 50 free credits per month	Tailored for Professionals /month incl. 20% VAT* Credits available per month	Perfect for small teams /month incl. 20% VAT* Credits available per month	For Research & Cultural Institutions Tailored to your needs
<ul style="list-style-type: none">✓ AI Text Recognition✓ Custom AI Model Training✓ Powerful Document Editor	<ul style="list-style-type: none">✓ Super Models✓ Advanced AI Tools✓ Priority Job Processing	<ul style="list-style-type: none">✓ 5 User Seats✓ Shareable Credits✓ Advanced AI Tools	<ul style="list-style-type: none">✓ User Seats & Management✓ API Access✓ Success Team
<button>Start for free</button>	<button>Choose</button>	<button>Choose</button>	<button>Get in Touch</button>

<https://www.transkribus.org/plans>

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

Transkribus Sites




The screenshot shows the Transkribus website interface. At the top, the Transkribus logo is on the left, and navigation links for 'Desk', 'Models', 'Sites' (highlighted with a yellow circle), and 'Jobs' are on the right. Below this is a dark blue header with 'Home', 'Collections', and 'Tags' on the left, and a search bar with 'Search text across all collections' and a 'Global Text Search' dropdown on the right. The main content area starts with a welcome message 'Welcome to Transkribus, Michela!'. To the right of this are buttons for 'Upload Files' and 'Quick Text Recognition'. Below the welcome message is a section for 'Recent documents' with a 'Continue where I left' link and a 'See history' link. Four document thumbnails are displayed in a row, each with a title and a date:

- 01_Transkription_BAV_Pal_lat_586 (Jan 23)
- 01_Transkription_BAV_Pal_lat_585 (Jan 21)
- 01_Transkription_Köln_EDD_Cod_119 (Jan 7)
- 01_Transkription_Bamberger_Stabilien_Can_6 (Jan 7)

<https://www.transkribus.org/sites>

Publish documents with Transkribus Sites <https://www.youtube.com/watch?v=LIaPHRWN6Uo>


Transkribus Sites: Vereinsarchiv Tiroler Landesmuseum Ferdinandeum

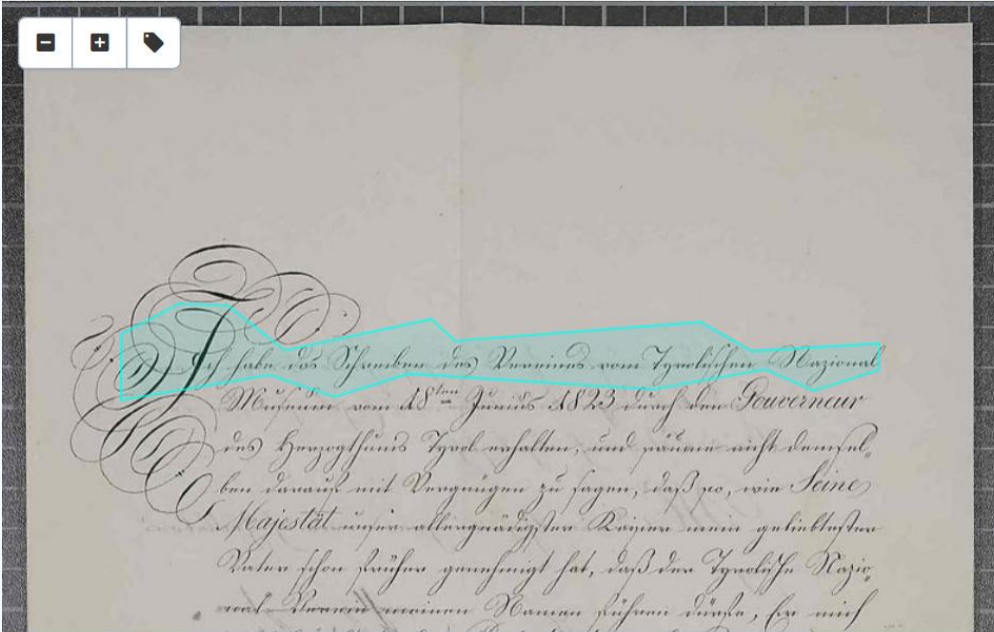
Vereinsarchiv Tiroler Landesmuseum Ferdinandeum

StartseiteEntdeckenSucheInhaltDEU

[< Zurück zum Dokument](#)


1823_Nr_001-067_duplicated - Seite 3 ⓘ





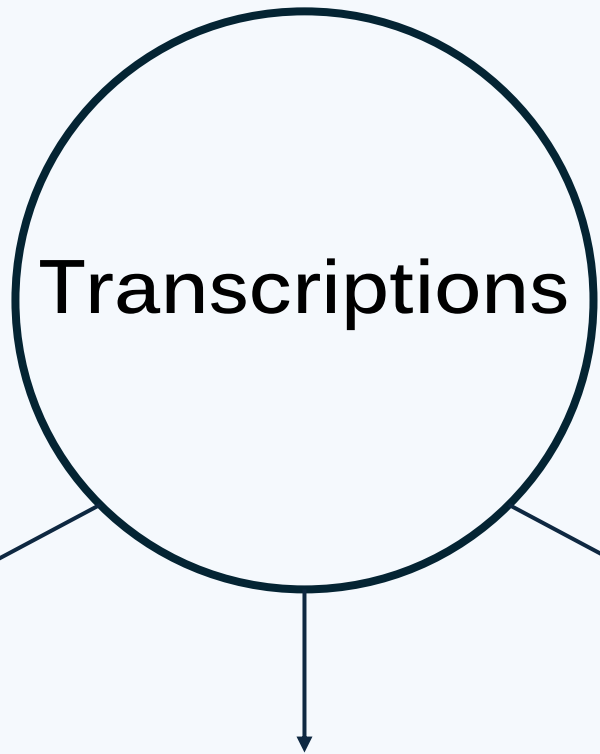
Ich habe das Schreiben des Vereines vom Tyrolischen Nazional Museum vom 18ten Junius 1823 durch den Gouverneur des Herzogthums Tyrol erhalten, und säume nicht demselben darauf mit Vergnügen zu sagen, daß sie, wie Seine Majestät unser allernädigster Kaiser mein geliebtester Vater schon früher genehmigt hat, daß der Tyrolische Nazional Verein meinen Namen führen dürfe; Er mich auch berechtigte das Protectorat von diesem Vereine annehmen zu dürfen.

Ich fühle mich dadurch, daß die Herrn Stände Tyrols einen Werth darauf legten dem neu gegründeten Vaterländischen Vereine, für Kunst und Wissenschaft meinen Namen beyzulegen, eben so sehr als durch Ihr Vertrauen mit dem Sie mir das Protectorat darüber anbiethen geehrt.

verein
TIROLER LANDESMUSEUM
FERDINANDEUM

Imprint Privacy

<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.

Export document

Server export Client export

Base folder: C:\Users\user\Desktop

File/Folder name: 01_Transkription_BAV_Pal_Iat_585

Export path: C:\Users\user\Desktop\01_Transkription_BAV_Pal_Iat_585

Choose export formats

- ☒ Transkribus Document
- ☒ PDF
- ☒ TEI
- ☒ DOCX
- ☒ Simple TXT
- ☒ Tag Export (Excel)
- ☐ Tag Export (IOB)
- ☒ Table Export into Excel
- ☒ Page metadata into Excel
- ☐ Export ALL formats
- ☐ Export Selected as ZIP

Export options:

Mets PDF TEI DOCX TXT

- ☒ Export Page
- ☐ Export ALTO
- ☐ Export ALTO (Split Lines Into Words)
- ☒ Export Image
- ☐ Export text files
- ☐ Export structural elements to Mets

Image type: Original

Filename pattern

- ☐ pageNr + filename
- ☒ filename (warning: filenames must be unique 1
- ☐ docId + pageNr + pageId

Pattern: \${filename}

Placeholder: \${docId}, \${filename}, \${pageId}, \${p

Version status

Latest version

Pages (686): 1-686 ... Current All

Tagname Choice

- ☒ Export all tags in doc
- ☐ Choose tags for export

Select Tags

OK Cancel

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”.



Add Texts

Cap libri xiii

~i~INCIPIUNT CAPITULA
TERCII DECIMI LIBRI
DE IEIVNIO
~1~De observatione g drage

Open

Upload

✓ Reveal

Voyant Tools is a web-based reading and analysis environment for digital texts.

Announcing the [Voyant Consortium](#). Register to be a [member here](#).

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

Voyant Tools, Stéfan Sinclair & Geoffrey Rockwell (© 2025) Privacy v. 2025-01-26 | Voyant Consortium