

READ

Recognition and Enrichment
of Archival Documents



Dokumente aus Transkribus exportieren

Version v1.9.1

Letztes Update dieses Guides: 28/11/2019

Dieser Guide erklärt, wie Sie Ihre Dokumente aus Transkribus exportieren und dabei verschiedene Dateiformate und weitere Optionen nutzen können.

Laden Sie den Transkribus Expert Client herunter, oder stellen Sie sicher, dass Sie die neueste Version verwenden:

- <https://transkribus.eu/>

Besuchen Sie das Transkribus Wiki für mehr Informationen und weitere Benutzeranleitungen:

- <https://transkribus.eu/wiki/>

Transkribus und die zugrundeliegende Technologie werden durch die folgenden Projekten und Plattformen verfügbar gemacht:

- <https://read.transkribus.eu/>
- <https://transcriptorium.eu/>
- <https://github.com/transkribus/>

Kontakt

- Das Transkribus Team: email@transkribus.eu

Inhalt

Einleitung.....	3
Exportfunktion.....	3
Serverexport.....	4
Clientexport.....	5
Speicherort.....	5
Zu exportierende Seiten.....	5
Export als Transkribus Document.....	6
ALTO.....	6
Dateiname.....	7
Export als PDF.....	7
Export als TEI.....	8
Export als DOCX.....	9
Export als einfaches TXT.....	9
Tag Export.....	9
Tabellenexport ins Excel.....	12
Mehr Optionen.....	12
Alle Formate exportieren.....	12
Versionsfilter verwenden.....	13
Do blackening.....	13
Titelseite erstellen.....	13
Danksagung.....	15



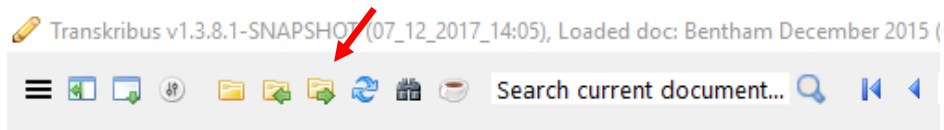
Das READ Projekt wird durch das Horizon 2020 Forschungs- und Innovationsprogramm im Rahmen des Fördervertrags Nr. 674943 finanziert.

Einleitung

- Wenn Sie mit Ihren Bildern und Transkriptionen außerhalb von Transkribus arbeiten möchten, können Sie Ihre Dokumente von der Plattform exportieren.
- Es sind verschiedene Exportformate und -features verfügbar.

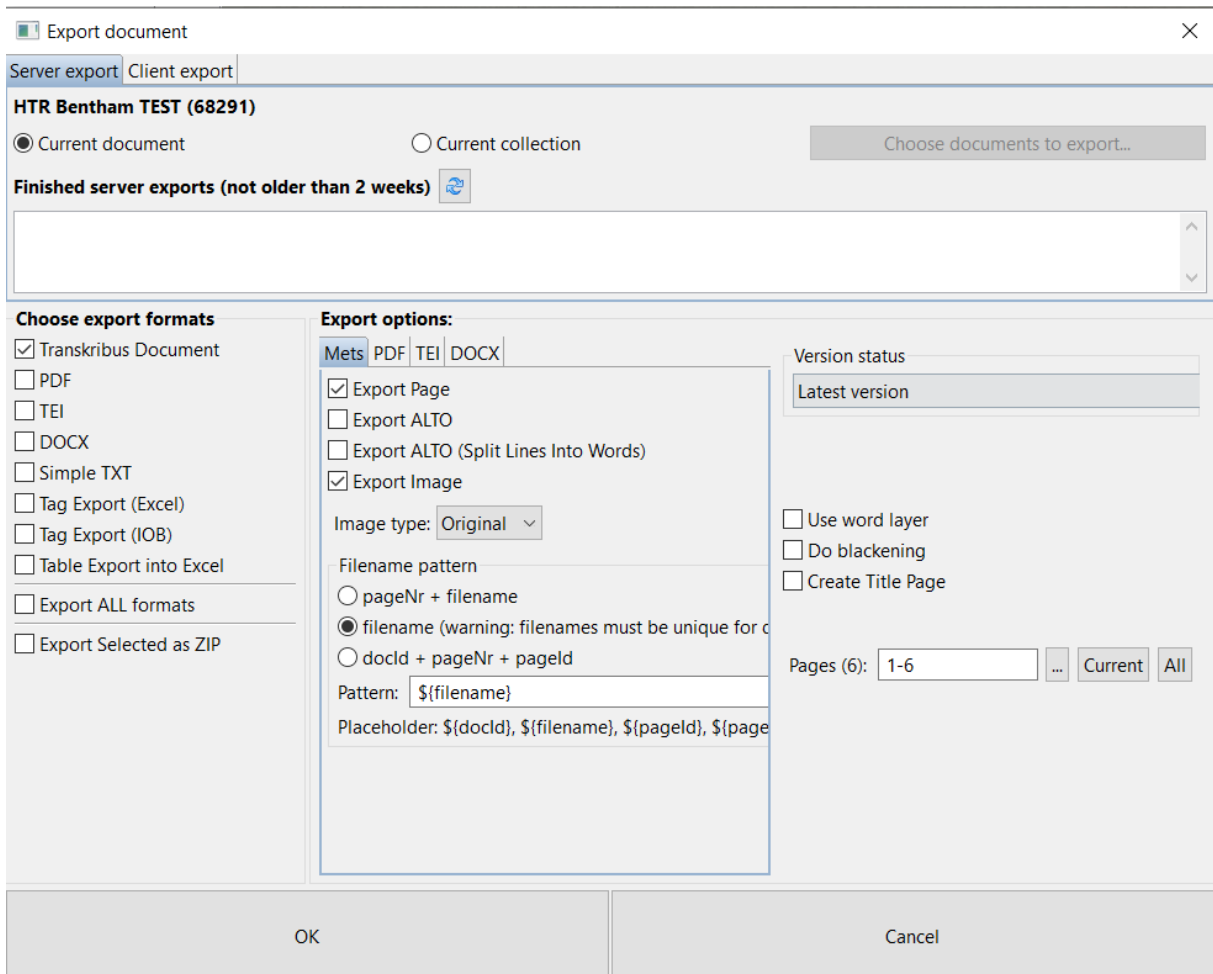
Exportfunktion

- Um die Exportfunktion zu verwenden, klicken Sie auf dieses Symbol:



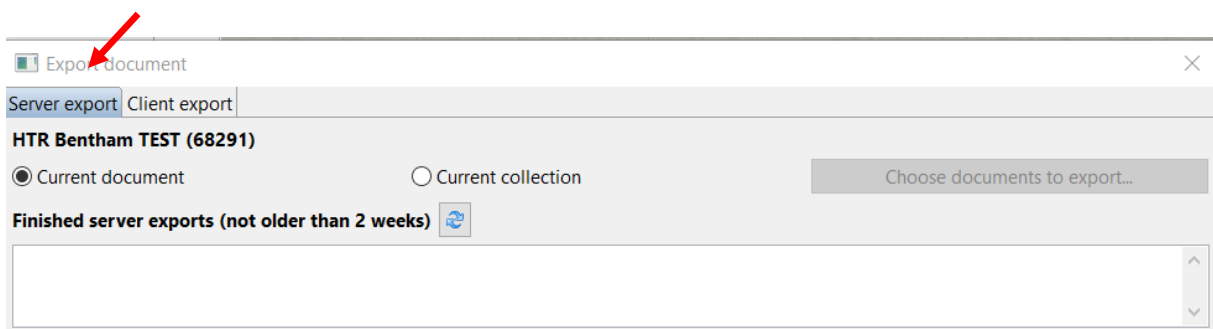
Darstellung 1 "Export document" Schaltfläche

- Das sich öffnende Fenster bietet zwei Exportmöglichkeiten: „Client export“ und „Server export“.
- Wenn Sie den Tab mit der „Server export“ Option auswählen, läuft der Export über den Transkribusserver und Sie erhalten einen Downloadlink für Ihre Dateien. Der Export hat keinerlei Auswirkungen auf Ihren Computer und wird nicht unterbrochen, wenn Sie Ihren PC ausschalten.
- Wenn Sie den „Client export“ Tab wählen, werden die Dateien direkt auf Ihren Computer gespeichert.



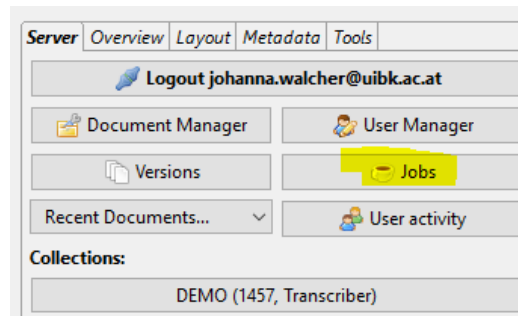
Darstellung 2 "Export document" Fenster

Serverexport



Darstellung 3 Serverexport

- Klicken Sie auf den „Server export“ Tab, wählen Sie Ihre gewünschten Optionen aus und klicken Sie auf OK.
- Sie können den Fortschritt Ihres Exports verfolgen, indem Sie auf die Schaltfläche „Jobs“ im „Server“ Tab klicken.



Darstellung 4 Unter "Jobs" den Fortschritt des Prozesses checken.

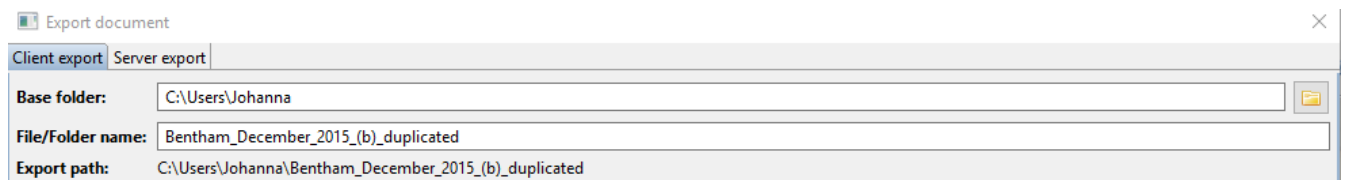
- Nachdem der Prozess abgeschlossen wurde, erhalten Sie eine E-Mail mit dem Downloadlink. Wenn Sie die Dateien jemand anderem zukommen lassen möchten, können Sie dazu die einfach diese Email weiterleiten.
- **Achtung:** Der Rest dieses Guides bezieht sich auf den Clientexport. Alle im Nachfolgenden behandelten Optionen sind aber auch für den Serverexport verfügbar

Clientexport

- Klicken Sie auf den „Client export“ Tab, um zu den Optionen zu gelangen.

Speicherort

Zuerst wählen Sie aus, wo Sie Ihre exportierten Dateien gerne speichern möchten. Geben Sie den Speicherort im Feld „Base folder“ im „Export document“ Fenster ein.



Darstellung 5 Speicherort angeben

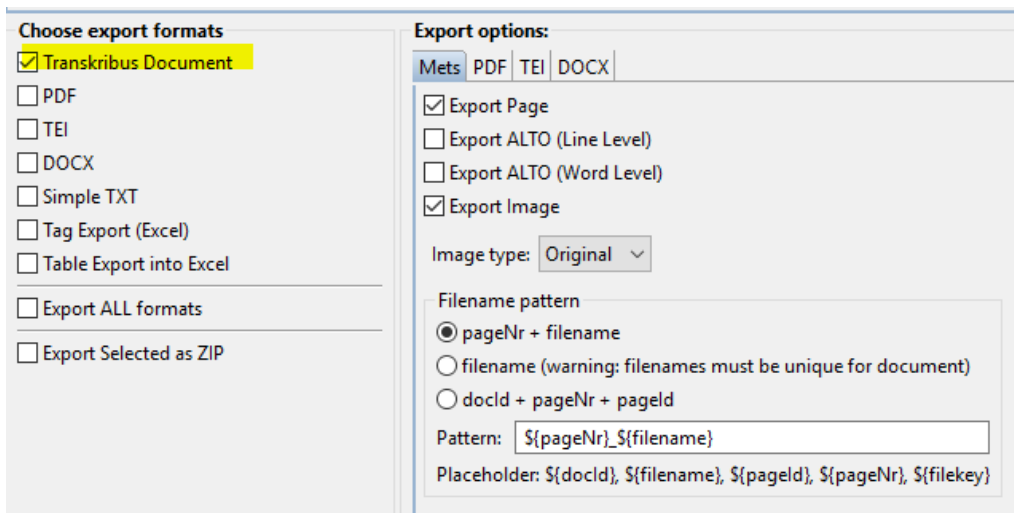
Zu exportierende Seiten

- Wählen Sie die zu exportierenden Seiten aus. Sie können alle Seiten Ihres Dokuments exportieren, oder nur einzelne.



Darstellung 6 Wählen Sie die zu exportierenden Seiten aus

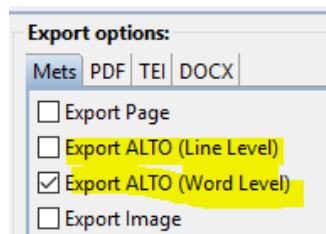
Export als Transkribus Document



Darstellung 7 Export als Transkribus Document

- Wenn Sie Ihre Transkription als Transkribus Document exportieren, erhalten Sie eine METS (Metadata Encoding and Transmission Standard) Datei mit links zu PAGE; XML, ALTO und/oder Bilddateien, abhängig von der gewählten Option. Eine METS Datei ist wie ein Container, der alle Hintergrundinformationen einer Datei enthält. Genauere Informationen über METS finden Sie unter: <http://www.loc.gov/standards/mets/>

ALTO



Darstellung 8 Export als ALTO document

- ALTO ist ein spezielles Ausgabeformat, das ermöglicht das exportierte Dokument in anderen Programmen, die dieses Format unterstützen, weiterzuverwenden. Es ist dem XML Format ähnlich und funktioniert z.B. für OCR-Software und wird oft in Kombination mit METS verwendet. METS beschreibt in diesem Fall das ganze digitalisierte Objekt und erstellt Referenzen zwischen den ALTO-Dateien wie z.B. die Angabe der Lesereihenfolge. Genauere Informationen zu ALTO finden Sie unter: <http://www.loc.gov/standards/alto/>
- Mit der „Word Level“-Option unterteilt Transkribus Zeilen in Worte. Das Programm analysiert hierfür die Leerräume zwischen den Wörtern, auch wenn vorher keine Segmentierung durchgeführt wurde.

Dateiname

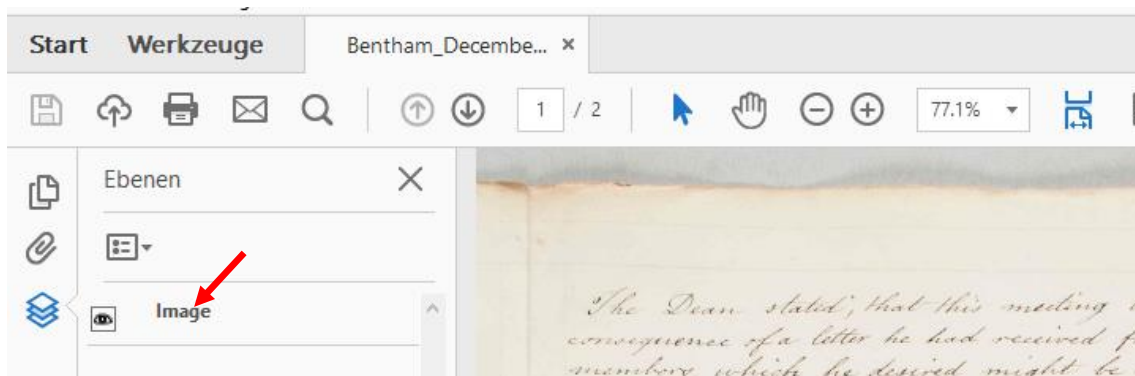
Darstellung 9 Dateiname bearbeiten

- Unter "Filename pattern" können Sie auswählen, wie der Dateiname zusammengestellt werden soll.
- **Achtung:** Die zweite Option "filename" ist die Standardoption. Mit dieser Option hat die Exportdatei denselben Namen wie die Datei, die Sie hochgeladen haben. Das ist wichtig, wenn Sie lokale Transkripte den Bildern in Transkribus zuordnen möchten. Wenn Sie also ein Dokument exportieren dann ändern Sie den Dateinamen lokal und laden die Datei dann wieder auf Transkribus hoch. Das Programm braucht zwei gleiche Dateinamen, um die Dateien richtig zu erkennen und zuzuordnen.

Export als PDF

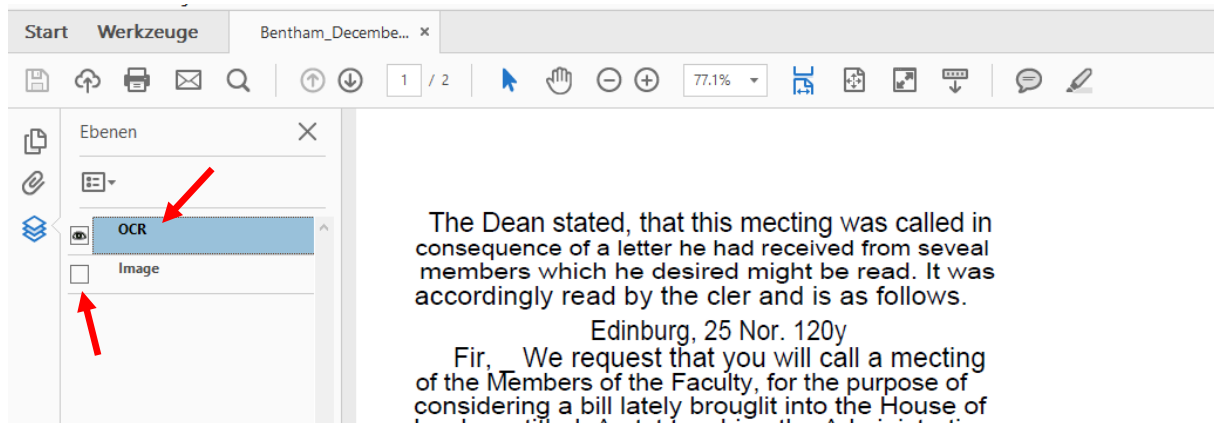
Darstellung 10 Export als PDF

- Wenn Sie eine PDF-Datei exportieren möchten, können Sie zwischen „Images only“ und „Images plus text layer“ auswählen.
 - o **“Images only”**: Sie erhalten eine PDF-Datei, die das Dokument als Bild enthält. Der transkribierte Text ist nicht sichtbar.



Darstellung 11 Exportierte PDF-Datei

- **“Images plus text layer”**: Sie sehen zwei Ebenen in der exportierten PDF-Datei: „OCR“ (transkribierter Text) und „Image“ (Bild des Dokuments).

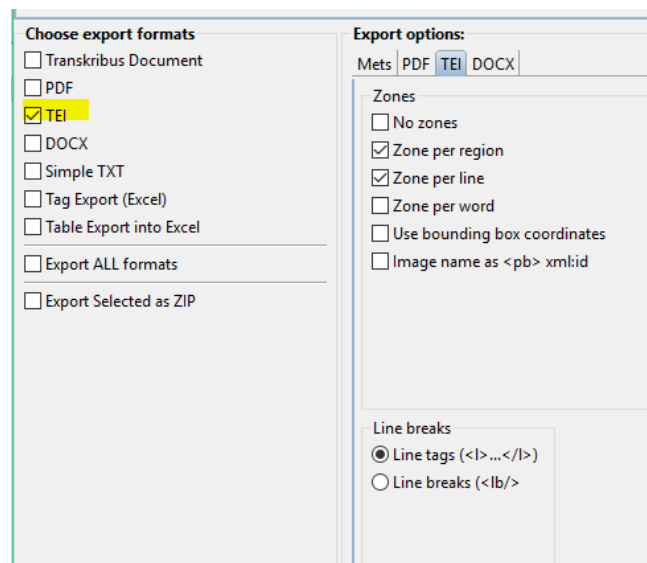


Darstellung 12 Ebenen in exportierter PDF-Datei

- **“Extra text pages”**: wenn Sie in der exportierten PDF-Datei hinunterscrollen, erscheint der transkribierte Text nach dem Bild.
- **“Highlight tags”**: diese Option hebt alle Tags in Ihrer Transkription hervor. Die Tags werden in denselben Farben, wie in Transkribus angezeigt. Außerdem sehen Sie am Ende des Dokuments eine Legende, die die Bedeutung der verschiedenen Farben erklärt.

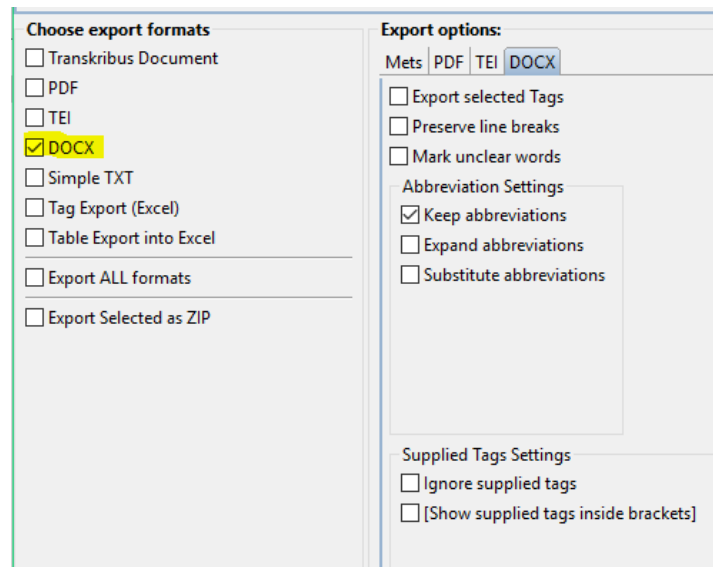
Export als TEI

- Diese Option ist für jene, die mit dem **Text Encoding Initiative (TEI)** Format arbeiten. Transkribus ermöglicht es Ihnen, die benötigten **Zonen** auszuwählen.
- Außerdem können Sie zwischen **line tags and line breaks** auswählen.
- Die Text Encoding Initiative ist eine textgerichtete “Community of practice” im akademischen Feld der Digital Humanities, die seit den 1980ern aktiv ist. Weitere Informationen unter: <http://www.tei-c.org/index.xml>



Darstellung 13 Export als TEI

Export als DOCX

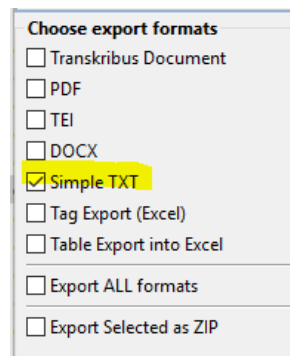


Darstellung 14 Export als DOCX

- Mit dieser Option erhalten Sie Ihre Transkriptionen als Worddatei.
- Sie können Optionen für Zeilenumbrüche, Abkürzungen etc. individuell anpassen.

Export als einfaches TXT

- Wenn Sie normalerweise nicht mit Microsoft Word arbeiten, können Sie Ihre Transkription als einfache TXT-Datei exportieren..



Darstellung 15 Export als TXT

Tag Export

Wenn Sie auch die Tags, die Sie ihrer Transkription zugewiesen haben, exportieren möchten, bietet Transkribus drei Optionen:

- 1. Excel-Datei:** Wählen Sie diese Option um eine Excel-Datei mit Reitern für jede Tag-Kategorie und einem Reiter mit einer Übersicht über alle Tags zu erhalten.

Choose export formats

- Transkribus Document
- PDF
- TEI
- DOCX
- Simple TXT
- Tag Export (Excel)
- Table Export into Excel

- Export ALL formats
- Export Selected as ZIP

Darstellung 16 Tag Export Excel

- 2. PDF-Datei:** Wählen Sie diese Option um die Tags in der exportierten PDF-Datei hervorzuheben.

Choose export formats

- Transkribus Document
- PDF
- TEI
- DOCX
- Simple TXT
- Tag Export (Excel)
- Table Export into Excel

- Export ALL formats
- Export Selected as ZIP

Export options:

Mets PDF TEI DOCX

- Images plus text layer
- Images only
- Extra text pages
- Highlight tags

Darstellung 17 Tag Export PDF

- 3. DOCX-Datei:** Wählen Sie diese Option um Ihre Tags in der exportierten DOCX-Datei sichtbar zu machen.

Choose export formats

- Transkribus Document
- PDF
- TEI
- DOCX
- Simple TXT
- Tag Export (Excel)
- Table Export into Excel

- Export ALL formats
- Export Selected as ZIP

Export options:

Mets PDF TEI DOCX

- Export selected Tags
- Preserve line breaks
- Mark unclear words

Abbreviation Settings

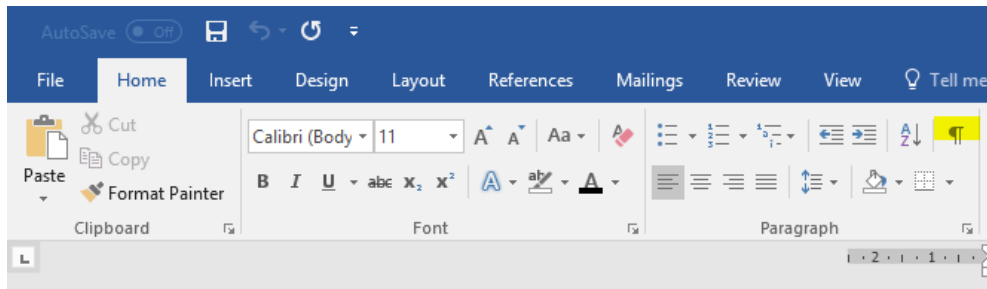
- Keep abbreviations
- Expand abbreviations
- Substitute abbreviations

Supplied Tags Settings

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

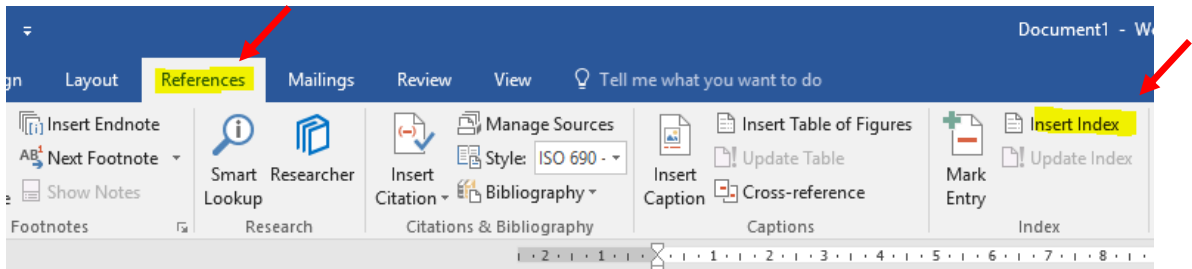
Darstellung 18 Tag Export DOCX

- Nach dem Export Word Dokument öffnen und folgende Aktion ausführen:
 - o Klicken Sie auf „Absatzmarken anzeigen“ im „Start“ Tab von Word.



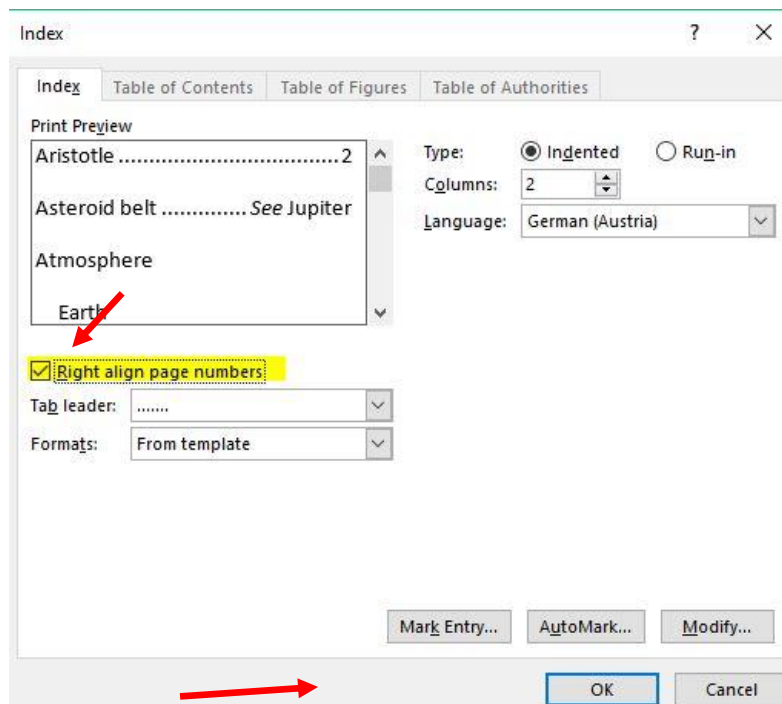
Darstellung 19 Absatzmarken anzeigen Schaltfläche in Word

- Dann gehen Sie zum Verweise „Tab“ und klicken auf „Inhaltsverzeichnis einfügen“



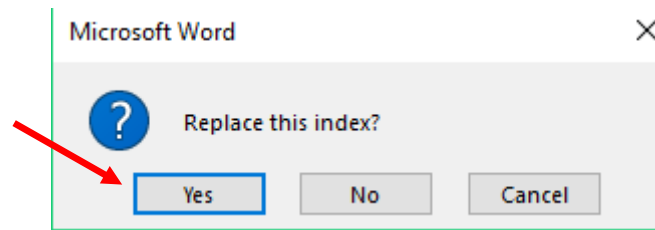
Darstellung 20 Inhaltsverzeichnis in Word einfügen

- Es wird das folgende Fenster in Word angezeigt:



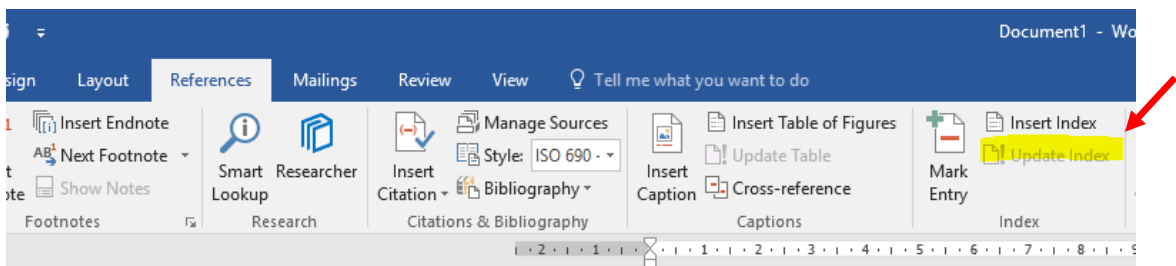
Darstellung 21 Inhaltsverzeichniseinstellungen in Word

- Wählen Sie "Seitenzahlen rechtsbündig" aus und drücken Sie „OK“
- Drücken Sie „Ja“ im Bestätigungsfenster



Darstellung 22 Bestätigungsfenster in Word

- Sie sollten jetzt eine Übersicht Ihrer Tags am Dokumentende sehen
- Wenn die Übersicht nicht erscheint, klicken Sie auf „Inhaltsverzeichnis aktualisieren“. Das sollte das Problem lösen.



Darstellung 23 "Inhaltsverzeichnis aktualisieren" in Word

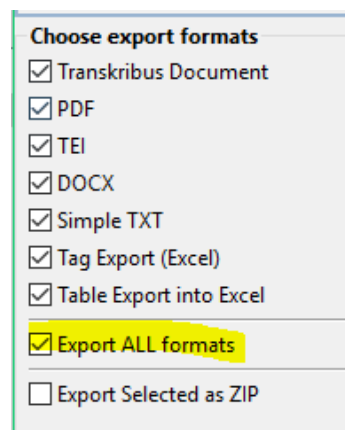
Tabellenexport ins Excel

Mehr Informationen zum Export von Tabellen finden Sie in der Anleitung [Tabellen in Transkribus](#).

Mehr Optionen

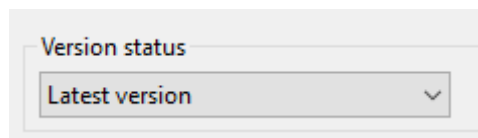
Alle Formate exportieren

- Mit dieser Option wird die Transkription in allen verfügbaren Formaten gespeichert.
- Wenn Sie "**Export ALL formats**" auswählen, werden alle Exportformate automatisch ausgewählt.



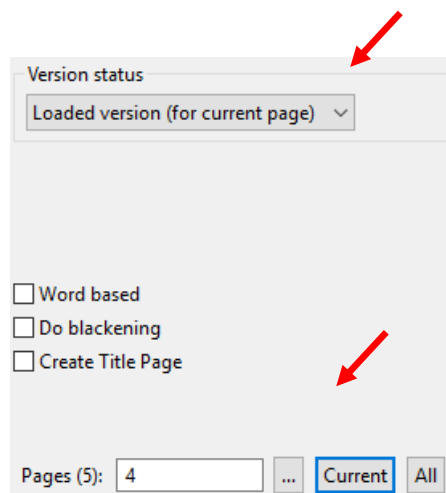
Darstellung 24 Alle Formate exportieren

Versionsfilter verwenden



Darstellung 25 Version auswählen

- Diese Option ermöglicht es, bestimmte Versionen des Dokuments zu exportieren.
- Wenn Sie beispielsweise „Ground Truth“ auswählen, exportiert Transkribus nur die Seiten Ihres Dokuments, die Sie als „Ground Truth“ markiert haben.
- Für den Export greift das Programm auf frühere Versionen Ihres Dokuments zurück. Falls Sie alle Seiten „In Progress“ exportieren möchten, wird das Programm alle Seiten exportieren die einmal als „In Progress“ markiert waren, selbst wenn der Status zum heutigen Zeitpunkt ein anderer ist.
- Das Programm wird die aktuellste „In Progress“-Version Ihres Dokuments exportieren.
- Wenn Sie eine spezifische frühere „In Progress“-Version einer Seite exportieren möchten, öffnen Sie die Seite in Transkribus. Dann öffnen Sie das Exportfenster und wählen Sie „Loaded version for current page“. Bei den „Pages“-Optionen wählen Sie „Current“ bevor Sie bestätigen.



Darstellung 26 Aktuelle Seite exportieren

Do blackening

- Wenn Sie vertrauliche Informationen Ihrer Transkription geschwärzt haben, können diese Worte oder Phrasen auch in den Exportdateien so gezeigt werden.
- Hierfür wählen Sie „Do blackening“ in den Exportoptionen aus.
- **Achtung:** diese Option funktioniert nur für Word, PDF und METS Dateien.

Titelseite erstellen

- Transkribus bietet die Möglichkeit eine Titelseite zu erstellen. Diese Seite wird auf Basis der Informationen, die Sie im Reiter „Document“ im Reiter „Metadata“ eingefügt haben, erstellt.
- Im Reiter „Document“ können Sie Informationen zum Titel, Autor, Sprache und dem Datum Ihres Dokuments hinzufügen. Außerdem können Sie eine Editorial Declaration erstellen, aus der hervorgeht, wie Ihr Dokument transkribiert wurde. Mehr Informationen zur Editorial Declaration finden Sie in der Anleitung [Transkribieren mit Transkribus](#).

Darstellung 27 Informationen für die Titelseite

- Diese Informationen werden auf Ihrer Titelseite wie folgt dargestellt.

Darstellung 28 Titelseite

Danksagung

Wir möchten den vielen Nutzern danken, die mit ihrem Feedback zur Verbesserung der Transkribussoftware beigetragen haben.

Transkribus wird der Öffentlichkeit im Rahmen des H2020e Infrastrukturprojekts READ (Recognition and Enrichment of Archival Documents) zugänglich gemacht. Dieses Projekt wird von der Europäischen Kommission im Rahmen des Fördervertrags Nr. 674943 finanziert