



**Kanton Zürich
Direktion der Justiz und des Innern
Staatsarchiv**

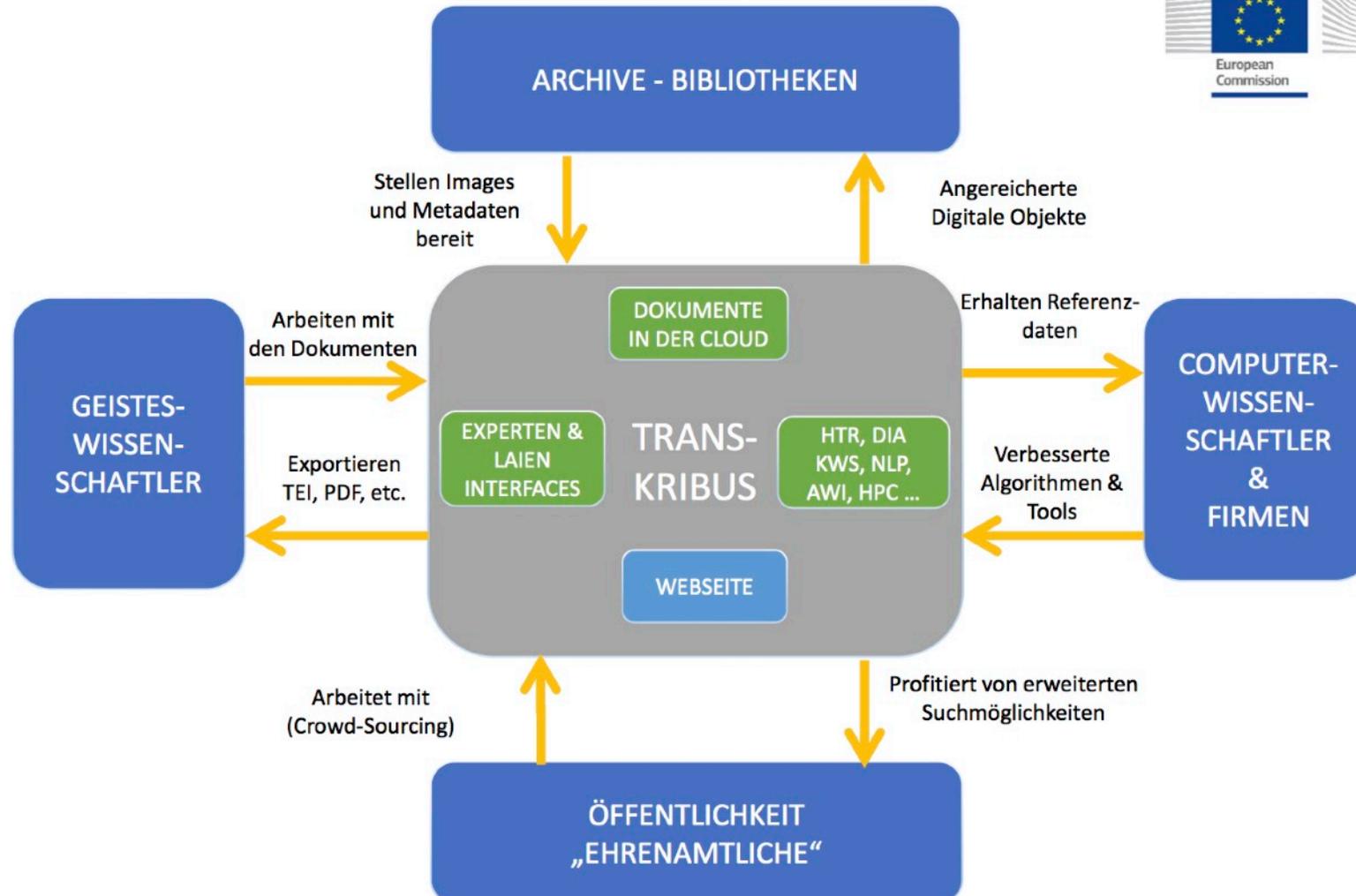
READ – Transkribus

Tobias Hodel



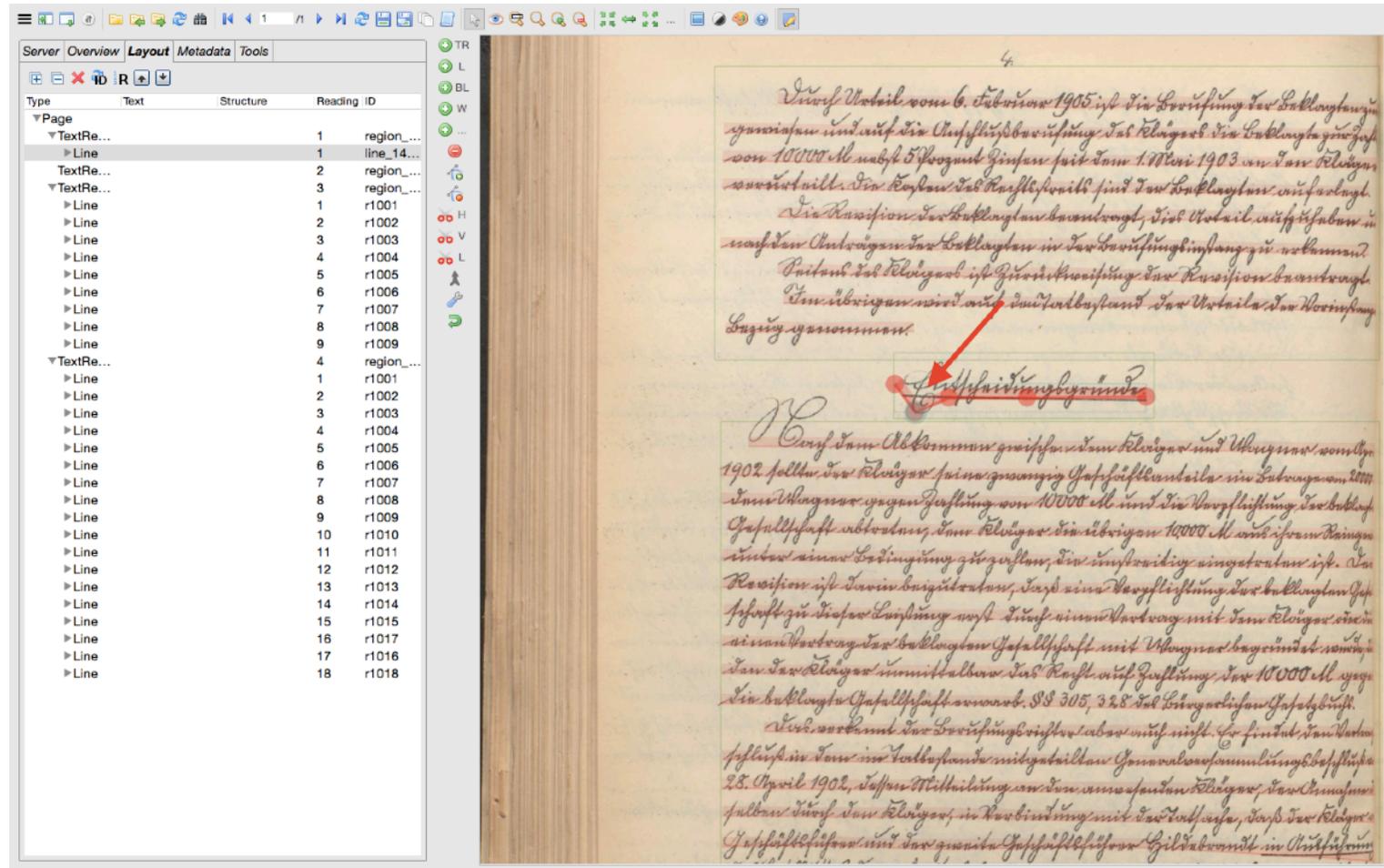
Nutzung von Transkribus (Expert Interface)

- Layout Analyse/Annotation
- Transkription/Handschriftenerkennung/OCR
- Tagging (Struktur, Textkritisch, Entitäten)
- Suchmöglichkeiten (Volltext, Tags, Metadaten, Keyword Spotting)
- Export (PDF, DOCX, TEI XML, PAGE XML, ALTO XML)



Erkennung und/oder Annotation von Layout und Text

Transkribus als Expert Interface



The screenshot displays a software interface for document transcription. On the left, a table lists detected text regions with their types, structures, and reading IDs. The main area shows a scanned page of handwritten text in German, with a red box highlighting a specific line and a red arrow pointing to it from the table.

Type	Text	Structure	Reading ID
Page			
TextRe...		1	region_...
Line		1	line_14...
TextRe...		2	region_...
TextRe...		3	region_...
Line		1	r1001
Line		2	r1002
Line		3	r1003
Line		4	r1004
Line		5	r1005
Line		6	r1006
Line		7	r1007
Line		8	r1008
Line		9	r1009
TextRe...		4	region_...
Line		1	r1001
Line		2	r1002
Line		3	r1003
Line		4	r1004
Line		5	r1005
Line		6	r1006
Line		7	r1007
Line		8	r1008
Line		9	r1009
Line		10	r1010
Line		11	r1011
Line		12	r1012
Line		13	r1013
Line		14	r1014
Line		15	r1015
Line		16	r1017
Line		17	r1016
Line		18	r1018

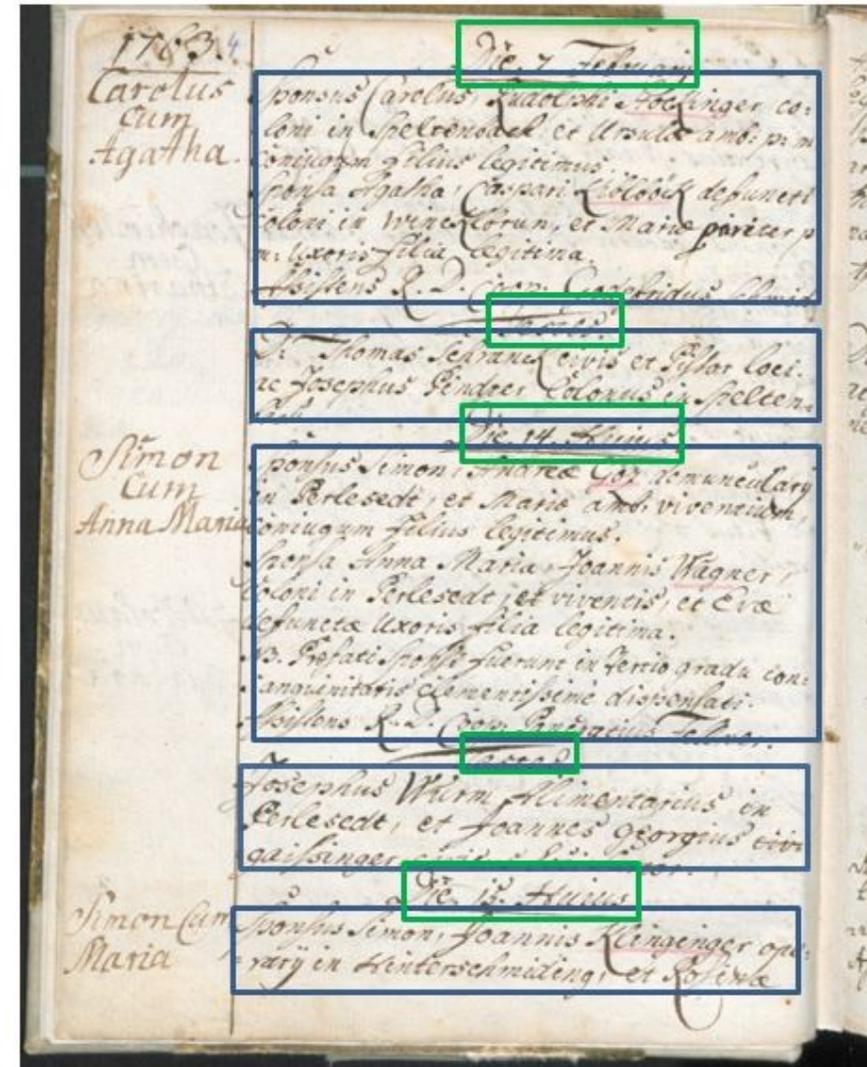
The main window shows a scanned page of handwritten text in German. A red box highlights a line of text, and a red arrow points to it from the table. The text on the page is:

Ursch. Urteil vom 6. Februar 1905 ist die Bestimmung der Beklagten zu
 zahlen und die Bestimmung der Beklagten zu zahlen
 von 10000 M. nach 5 Prozent Zinsen seit dem 1. Juni 1903 an den Kläger
 zurück zu zahlen. Die Kosten des Verfahrens sind der Beklagten anzuliegen.
 Die Revision der Beklagten beruht auf dem Urteil, welches in
 der Sache der Beklagten in der Bestimmungsbeklagten
 nicht der Beklagten ist. Die Revision beruht auf dem Urteil der
 Beklagten und ist im Urteil der Beklagten zu zahlen.
 Die Revision beruht auf dem Urteil der Beklagten und ist im Urteil der
 Beklagten zu zahlen.

The highlighted line in the table is: **Bestimmungsbeklagten**. A red box highlights this line in the table, and a red arrow points to it from the table.

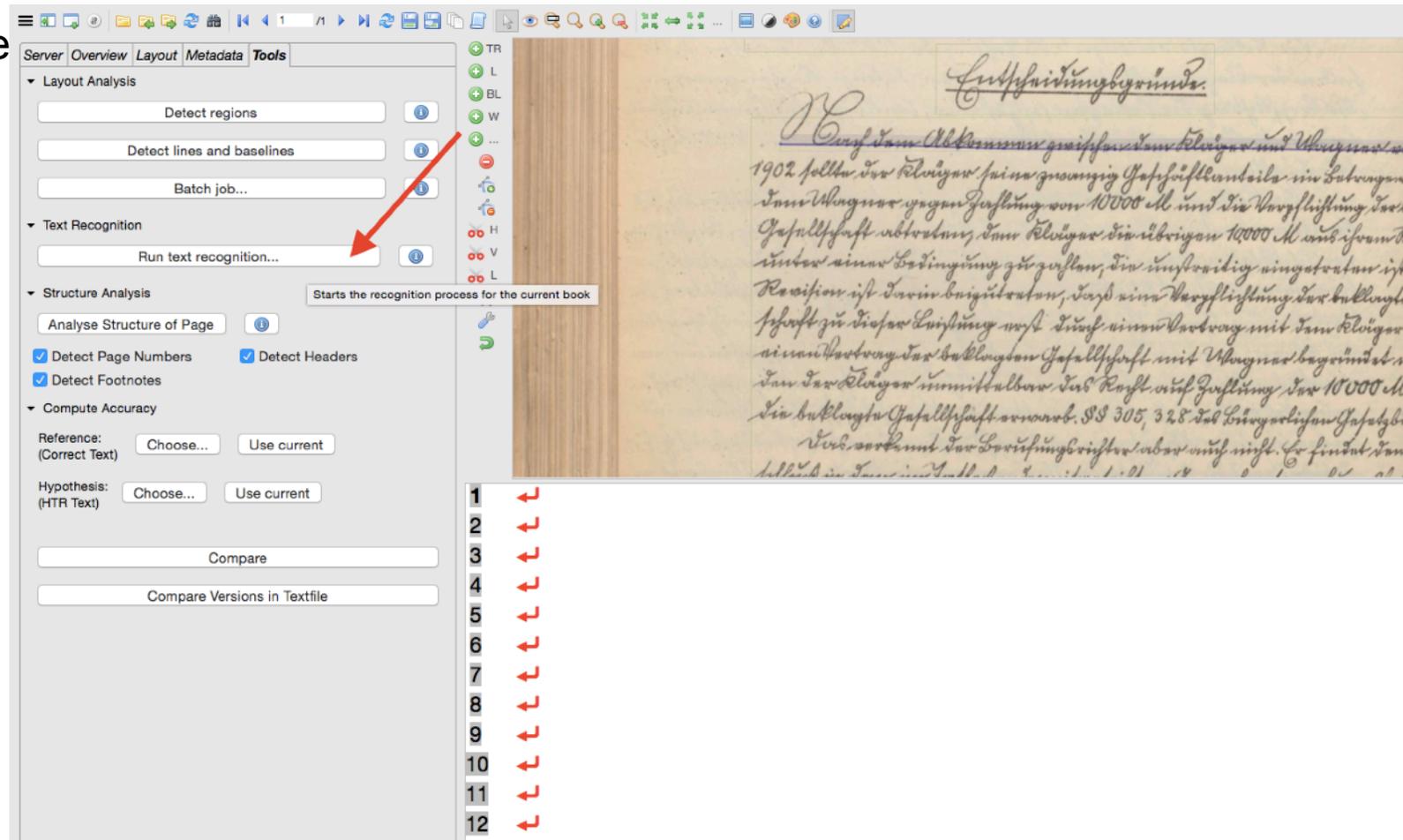
Erkennung und/oder Annotation von Layout und Text

Document Understanding



Erkennung und/oder Annotation von Layout und Text

Transkribus als Expert Interface



The screenshot displays the Transkribus Expert Interface. The left sidebar contains several tool categories:

- Layout Analysis:** Includes buttons for "Detect regions", "Detect lines and baselines", and "Batch job...".
- Text Recognition:** Features a "Run text recognition..." button, which is highlighted with a red arrow. A tooltip below it reads "Starts the recognition process for the current book".
- Structure Analysis:** Includes "Analyse Structure of Page" and checkboxes for "Detect Page Numbers", "Detect Footnotes", and "Detect Headers".
- Compute Accuracy:** Includes fields for "Reference: (Correct Text)" and "Hypothesis: (HTR Text)", each with "Choose..." and "Use current" buttons, and "Compare" and "Compare Versions in Textfile" buttons.

The main window shows a scanned page of a handwritten document with the title "Leibensversicherungsgesetz". The text is in German and discusses insurance regulations. A vertical toolbar on the right side of the document view includes icons for TR, L, BL, W, H, V, and L. At the bottom, a list of page numbers from 1 to 12 is visible, with red arrows pointing to the right next to each number.

Texterkennung: Fehlerraten (Character Error Rate)

Schreiber, 150 Seiten Trainingsmaterial:

10%

Schreiber, 450 Seiten Trainingsmaterial:

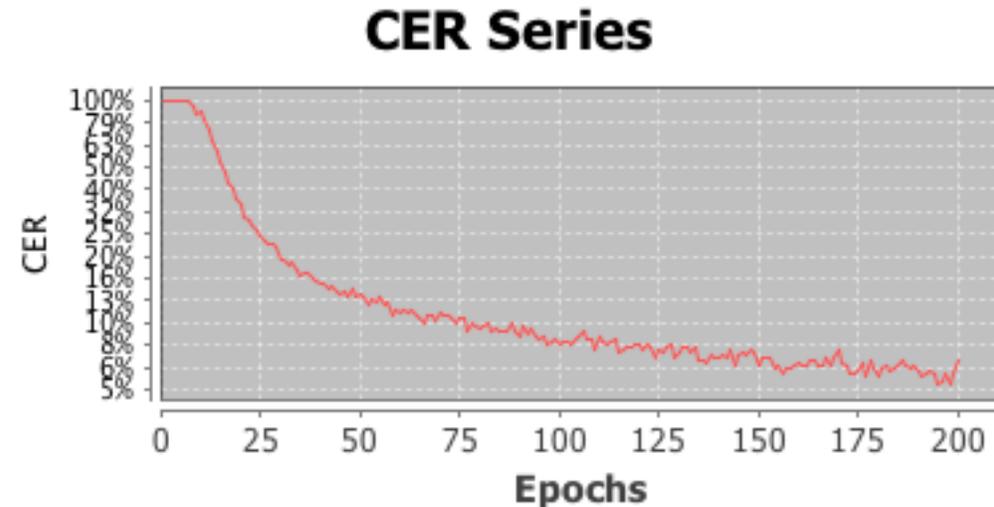
4,4%

Gleicher Schreiber, 10 Jahre später,
ohne Trainingsmaterial:

9,2%

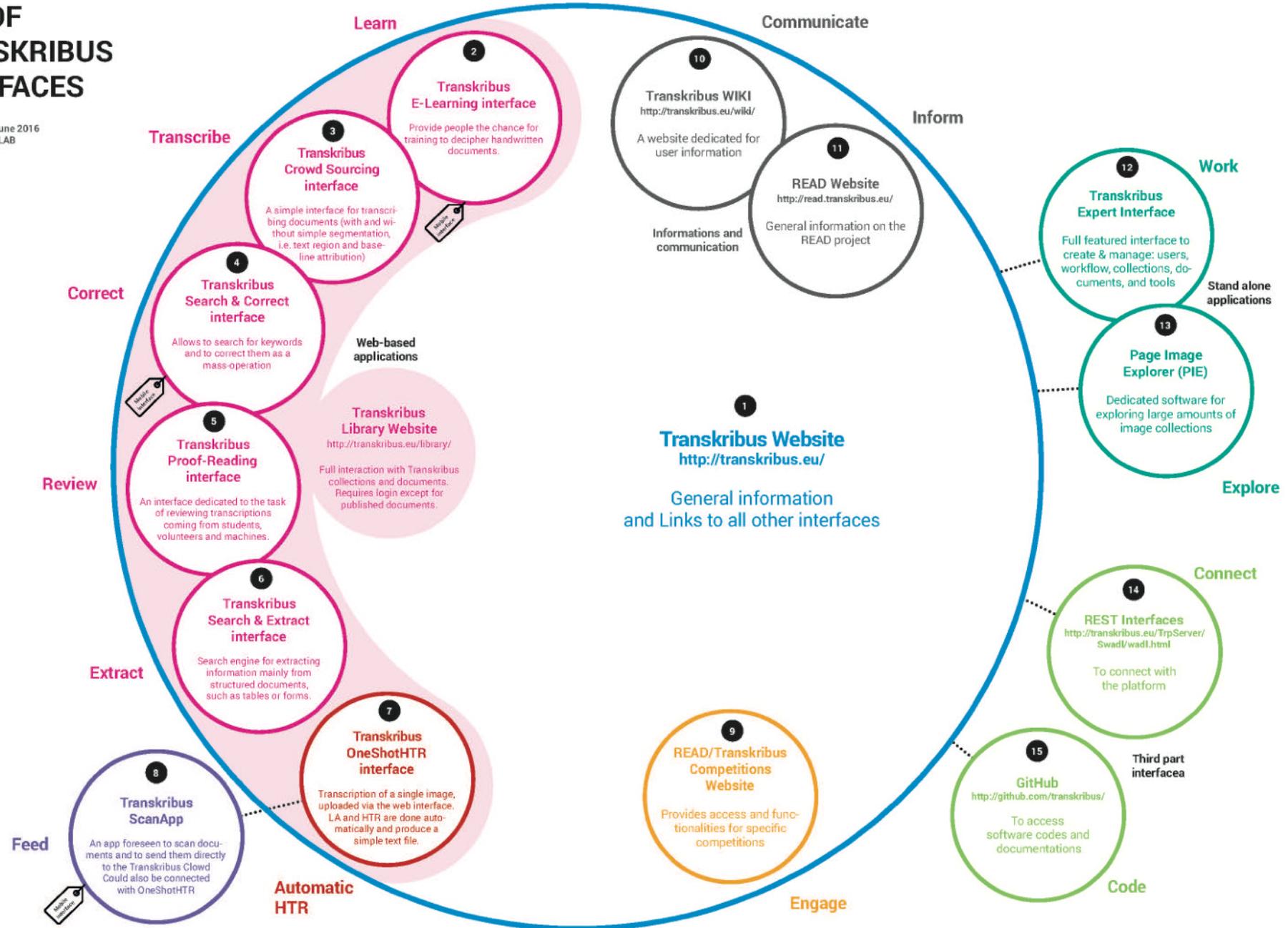
Schreiber, 1132 Seiten Trainingsmaterial:

3%



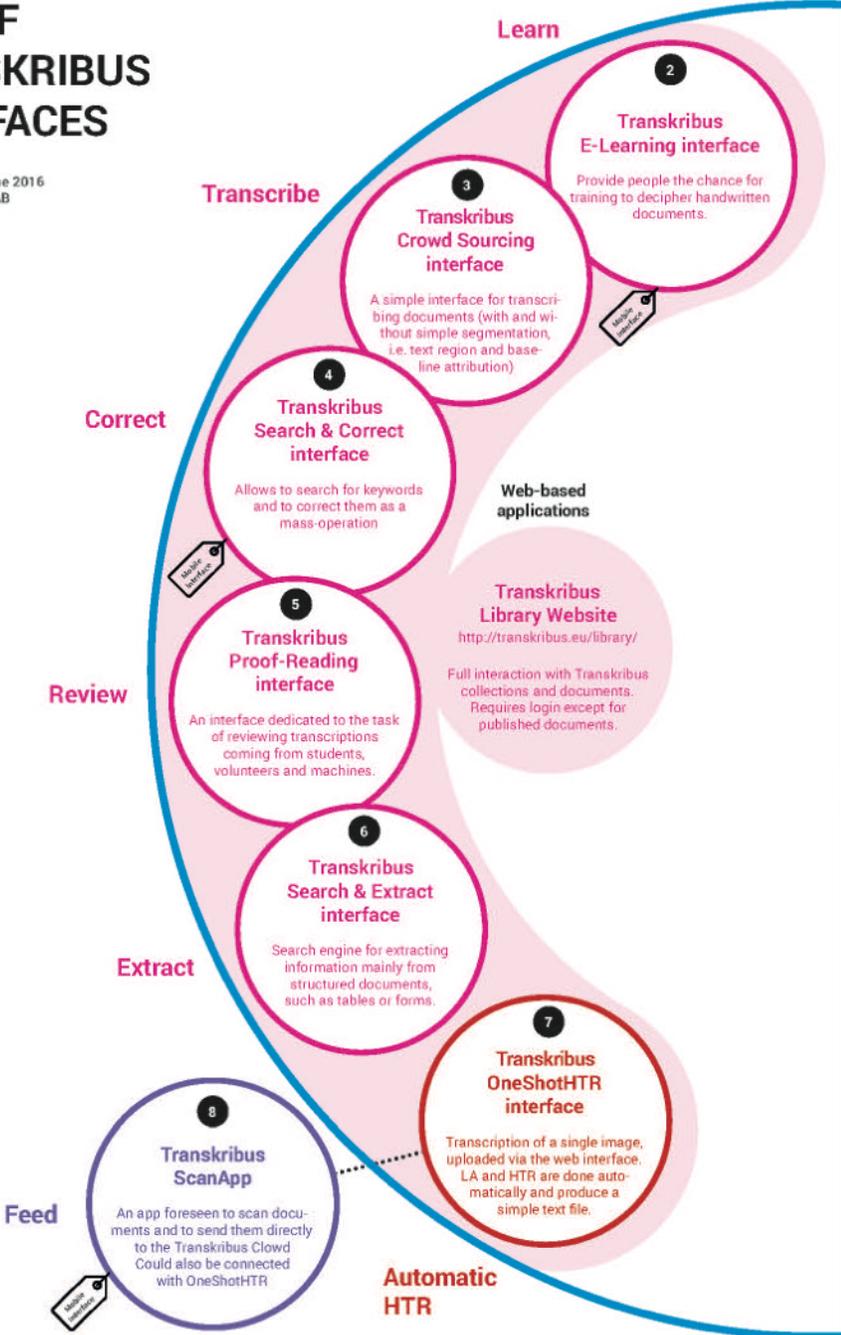
MAP OF TRANSKRIBUS INTERFACES

Version 1.0
PROJECT READ - June 2016
L. Bolli - EPFL / DHLAB



MAP OF TRANSKRIBUS INTERFACES

Version 1.0
PROJECT READ - June 2016
L. Bolli - EPFL / DHLAB



(Web-)Interfaces

- Transkription/Crowdsourcing
- Korrektur
- Suchen/Extrahieren
- E-Learning
- ScanApp

READ

Transkribus Base

Collections

Search

Ingest documents

Profile

Logout

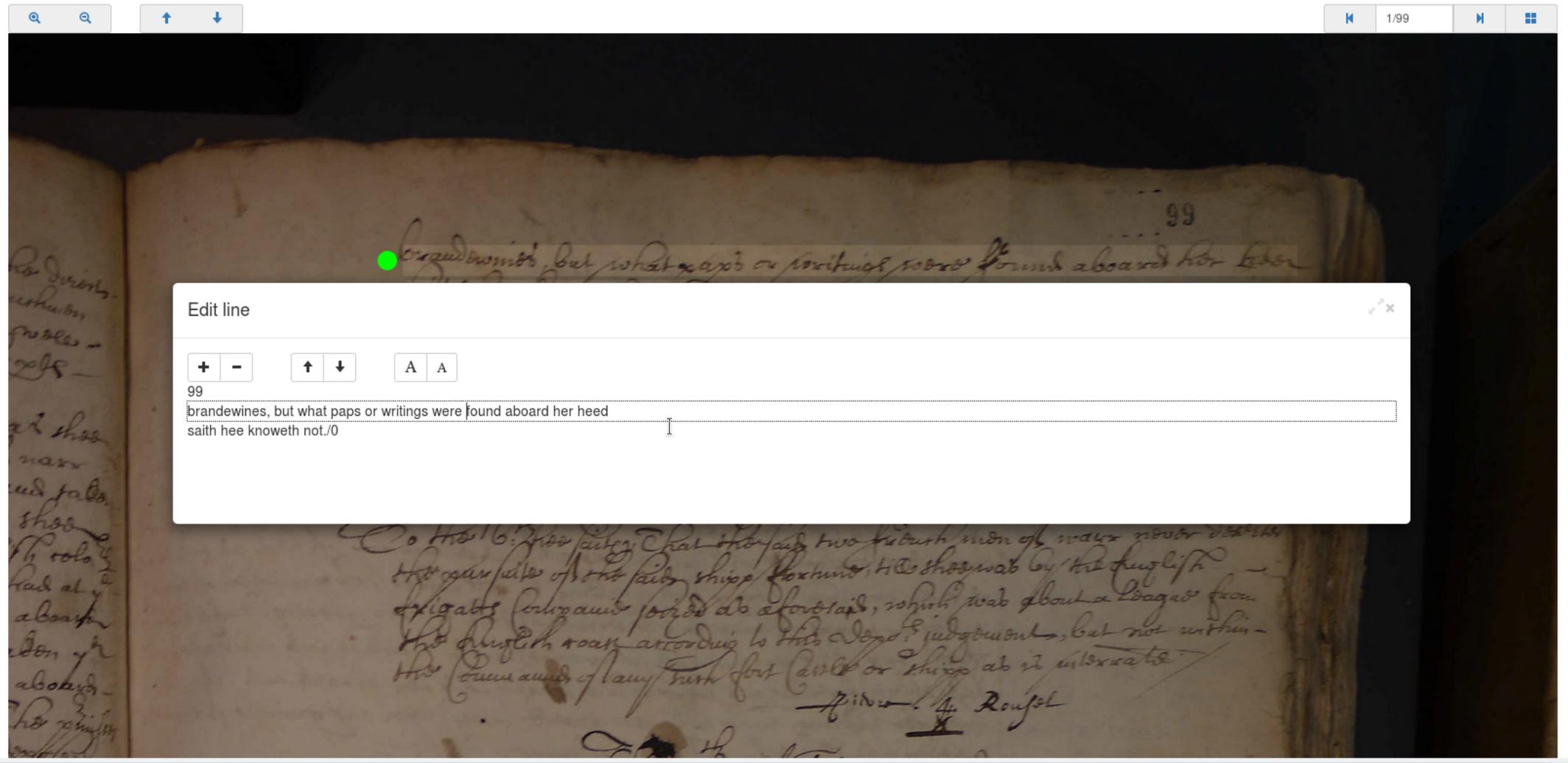
About

User Guide

English



Correction Interface, Line by Line (working title)



Edit line ✕

99

brandewines, but what paps or writings were found aboard her heed
saith hee knoweth not./0

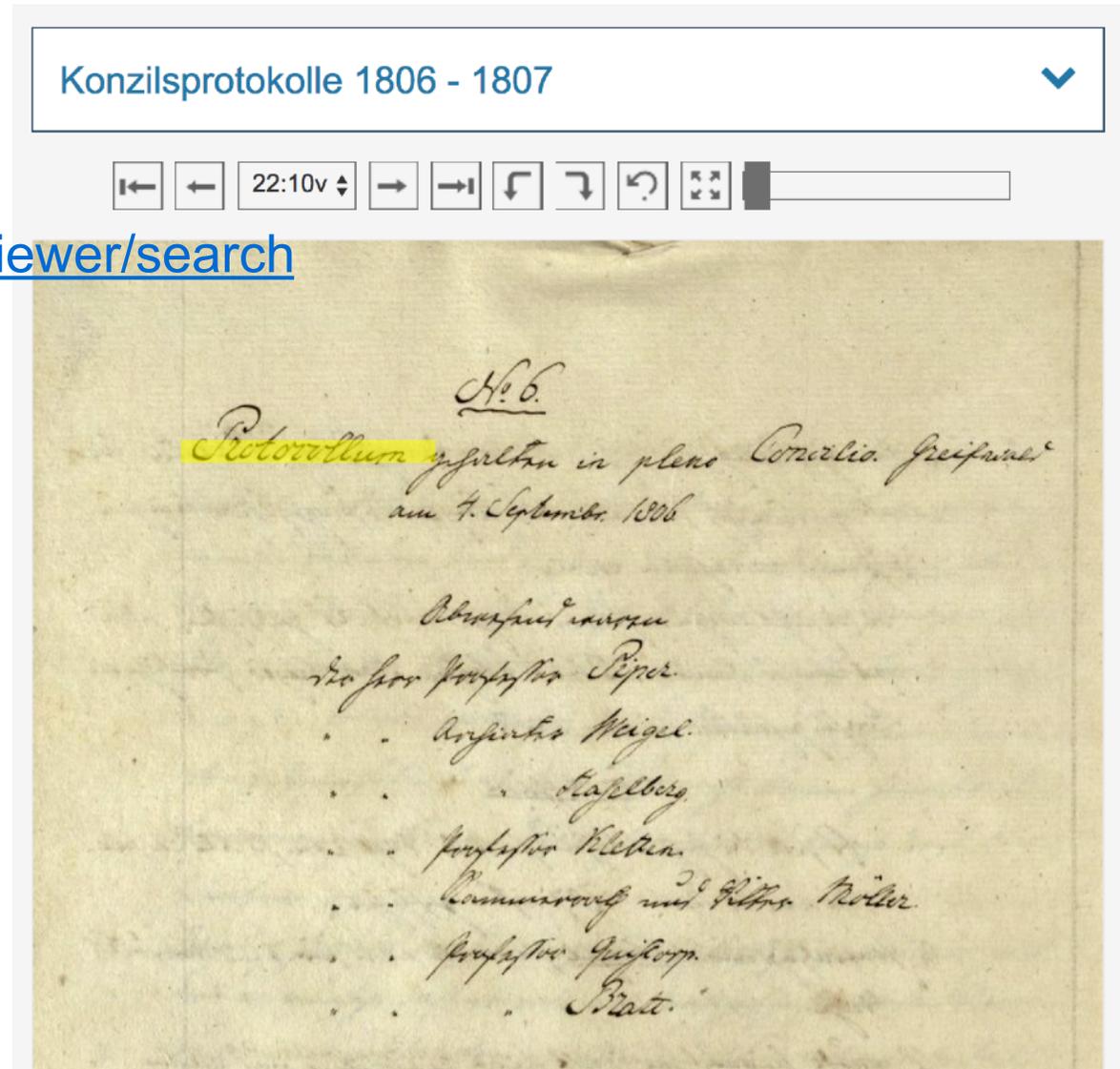


Keyword Spotting

Umsetzung (Beispiel von Intranda/goobi)

<http://ub-goobi-pr2.ub.uni-greifswald.de/viewer/search>

(Stichwort: «protocollum»)



Projekte mit Memorandum of Understanding



Projektpartner READ

- University of Innsbruck (co-ordinator / Austria)
- Universitat Politecnica de Valencia (Spain)
- University College London (United Kingdom)
- National Center for Scientific Research “Demokritos” (Greece)
- Democritus University of Thrace (Greece)
- University of London Computer Centre (United Kingdom)
- Vienna University of Technology (Austria)
- University of Rostock (Germany)
- Leipzig University (Germany)
- Xerox Research Centre Europe (France)
- Ecole Polytechnique Federale de Lausanne (Switzerland)
- National Archives Finland (Finland)
- Passau Diocesan Archives (Germany)

Transkribus: transkribus.eu / email@transkribus.eu / transkribus.eu/wiki/

READ: read.transkribus.eu (auch für News)

Staatsarchiv Zürich: tobias.hodel@ji.zh.ch

Vielen Dank für die Aufmerksamkeit

