

# Transkribus,

a research platform for the digitisation, transcription,  
recognition and searching of historical documents

Günter Mühlberger  
University of Innsbruck,  
Digitisation and Digital Preservation Group

**READ**

# Agenda

- Introduction
- Some highlights
  - ScanTent-DocScan
  - Workflow
  - Text Recognition
  - Keyword Spotting
- Future of Transkribus
  - Governance model
  - Business model

# Introduction

# READ

Recognition and Enrichment of Archival Documents

Home

News

Mission

Network

Research

Services

Publications

Contact



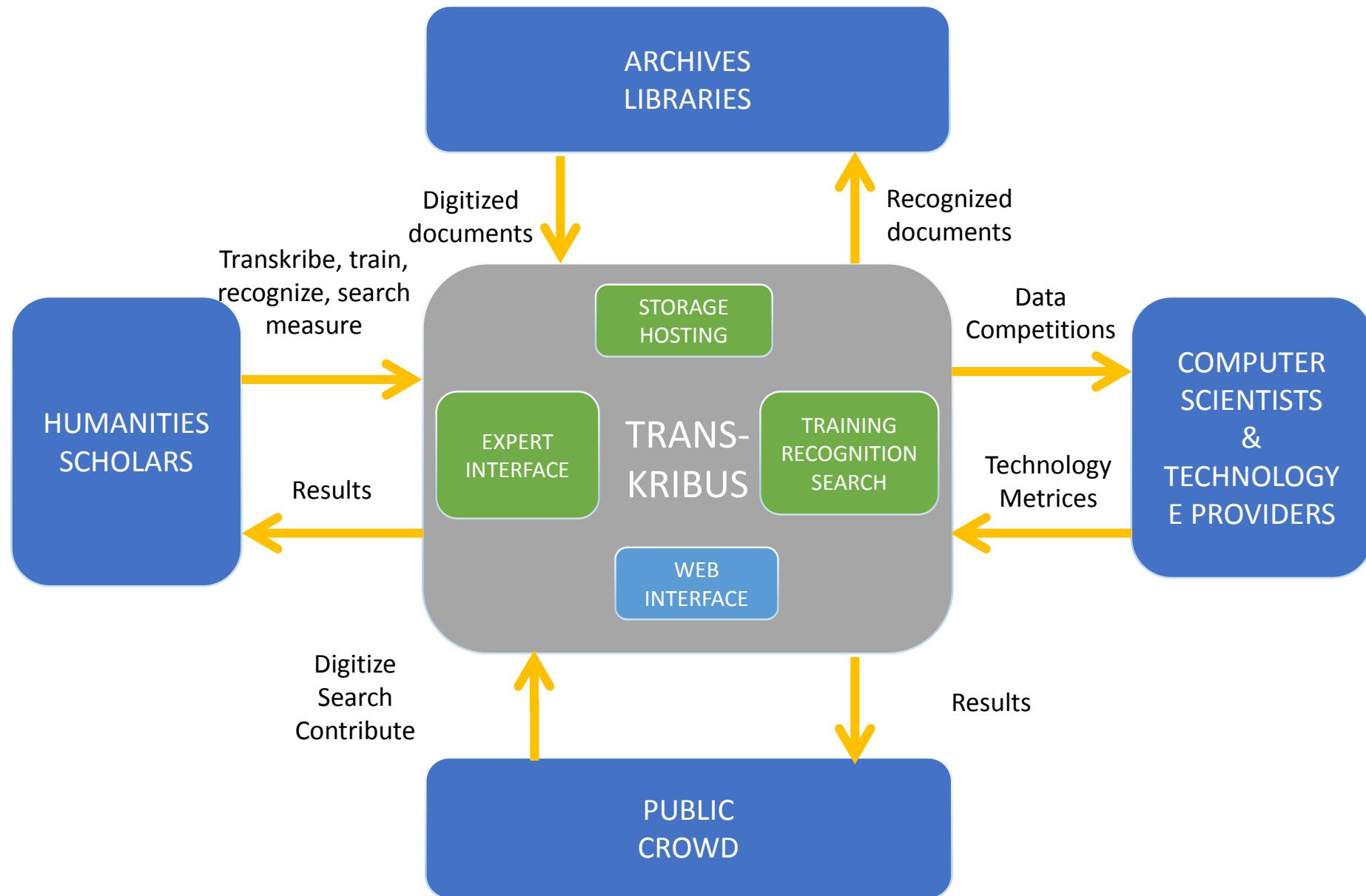
Das vorliegende Dokument ist vom Kläger bezogen und wurde am 19.01.2010 erfasst. Die Nummer der Urkunde ist  
Zurückgezahlt Nr. 108245 zu dem Konto Nummer Nr. 82785 (vgl. Blatt 156).  
Es handelt sich um eine Urkunde über die Abtretung eines Grundstückes (vgl. Blatt 156) aus dem Jahr.  
Die Urkunde ist von einem Notar ausgestellt am 18.09.1899 und wurde von dem Kläger am 18.09.1899  
Zurückgezahlt das Zurückgezahlte zu dem Käufer folgte. Am 21. Februar  
und hierzu bei einer zivilen Gerichtsinstanz in Berlin nimmt.

# Consortia

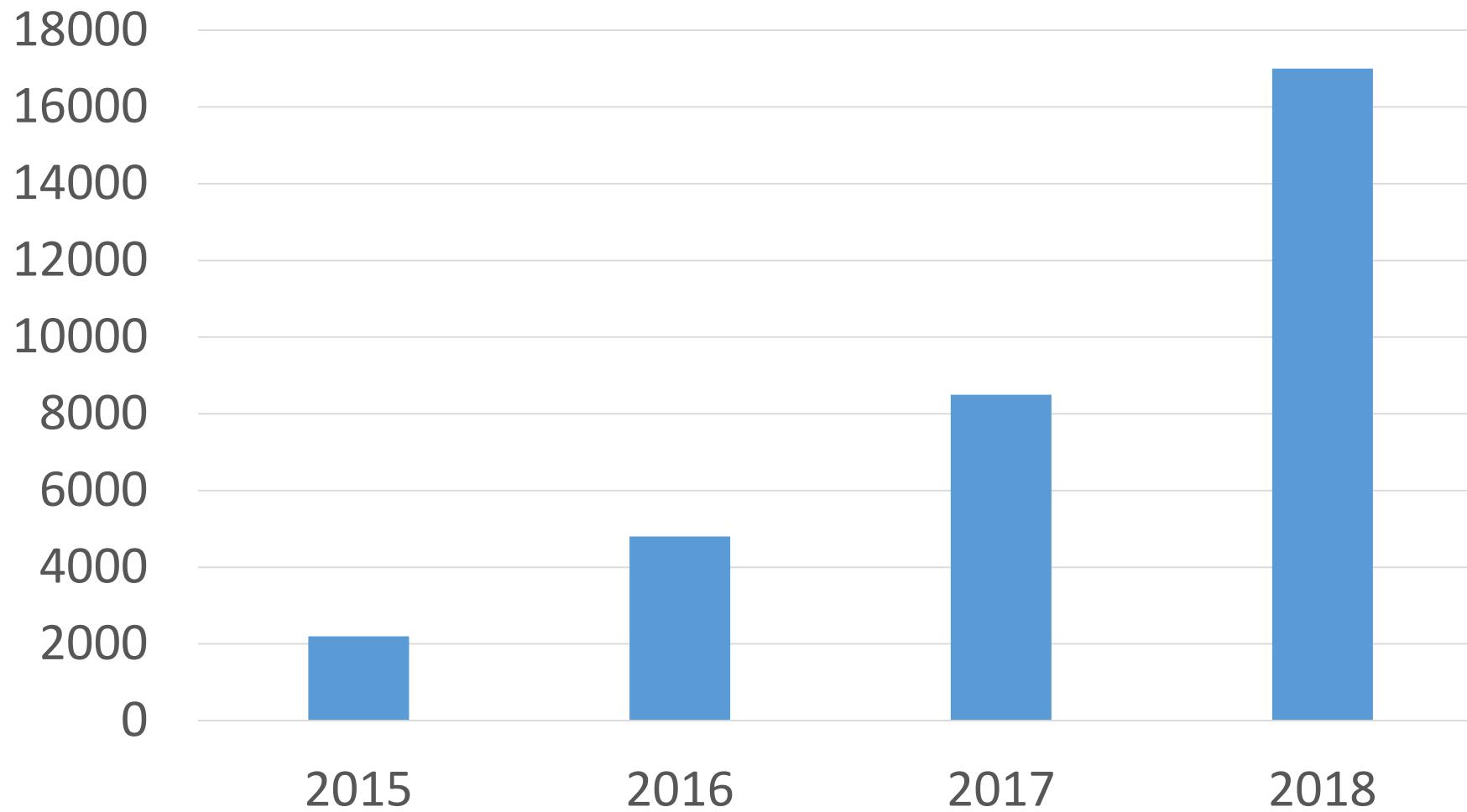
- University Innsbruck (Coordinator / Austria)
- Technical University Valencia (Spain)
- University College London (UK)
- National Center for Scientific Research “Demokritos” (Greece)
- University of London Computer Centre (UK)
- Technical University Vienna – Computer Vision Lab (Austria)
- University Rostock (Germany)
- University Leipzig (Germany)
- Naver Labs (France)
- Technical University Lausanne (Schwitzerland)
- National Archive Finland (Finland)
- Statearchive Zurich (Switzerland)
- Passau Diözesanarchiv (Germany)
- Democritus University Thrace (Greece)

# READ - Transkribus

- Facts and figures
  - 8,2 mill. EUR funding
  - Duration: 1.1.2016 – 30.6.2019 (31.12.2019)
  - 14 partners, coordinated by University of Innsbruck
- Research infrastructure
  - Enable digitisation, transcription, recognition and searching of historical documents
  - Cover the complete workflow so that users in the best case need not to use any other tool for receiving the best text or the best search options currently possible
- Set up a “culture of cooperation”
  - To get the best results archives/libraries, humanities scholars, volunteers and computer scientists need to collaborate
  - Transparency – make results fully available to other researchers
  - Sharing – enable users to share results among each other

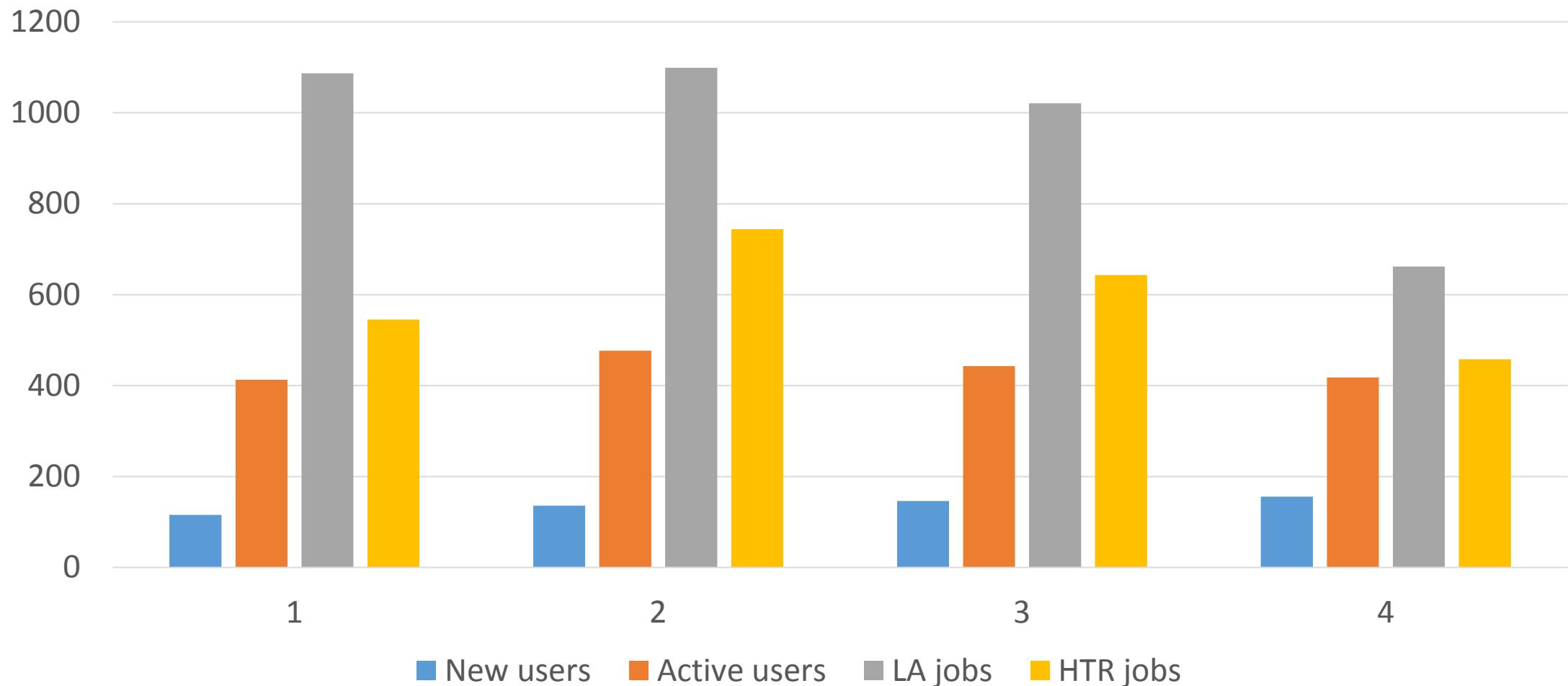


## Registered users in Transkribus

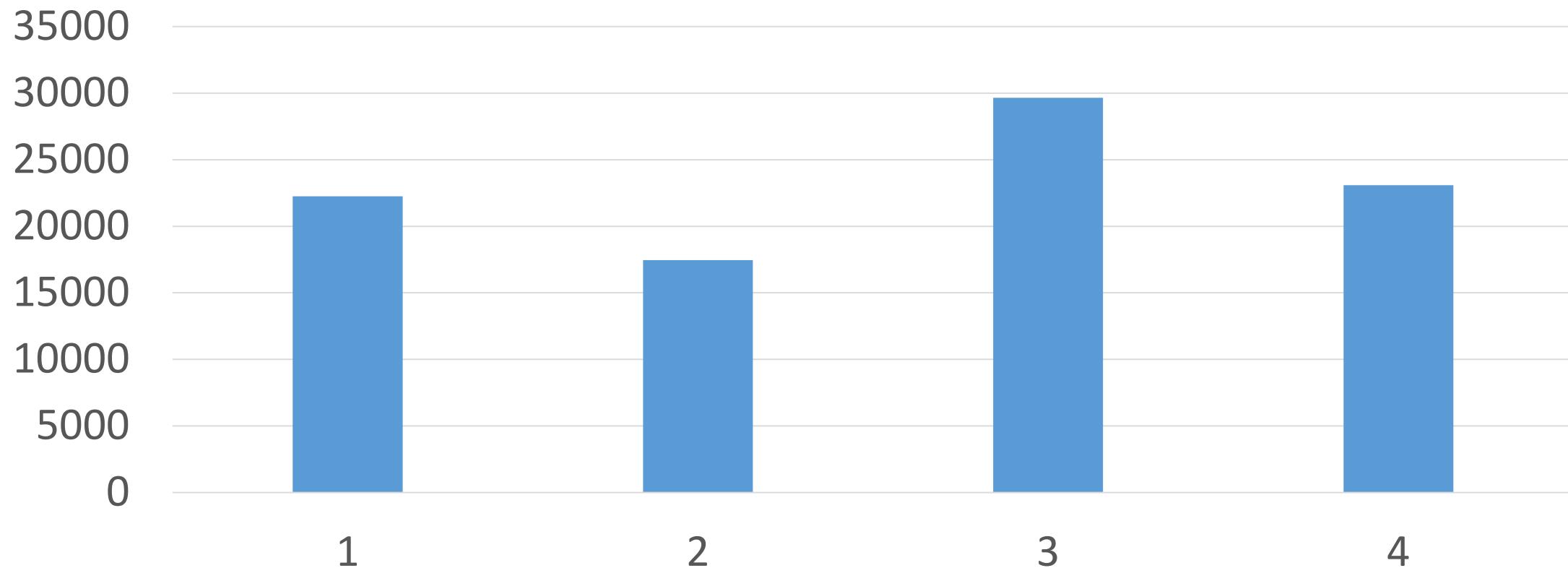


# Activities in September 2018

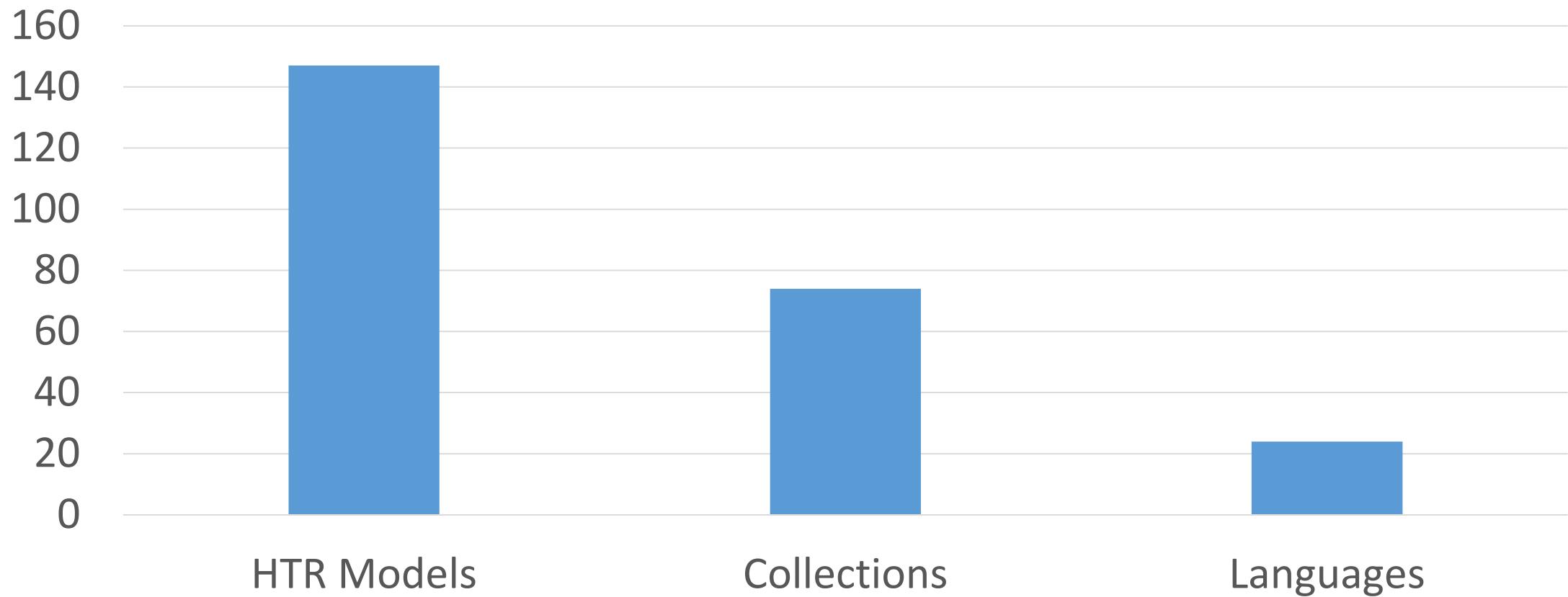
## September 2018 per week



Images uploaded by users in September 2018 / per  
week / 92.448 in total



## HTR models trained by users in September 2018



# Data

- Software and algorithms will change – data remain!
- Ground truth data are highly valuable
- Users created or provided for their models **6525 pages of ground truth**
- 1 page = 1 hour of work = 6525 hours = **3,8 person years**
- If we count one hour with 20 EUR we come to **130.500 EUR** of value created by users just for September 2018

# MoU Partners

- **Australia**
- [Geopast](#)
- [National Library of Australia](#)
- **Austria**
- [University Vienna, German Studies Department](#)
- [Technical University Graz – University Archive](#)
- **Canada**
- [DEEDS Project – University of Toronto](#)
- **Denmark**
- [Aarhus City Archives](#)
- **Finland**
- [Finnish Literature Society](#)
- [The Society of Swedish Literature in Finland](#)
- [Tampere University Folklife Archives](#)

- **France**
- [ALMAnaCH – Inria Paris Research Centre](#)
- [National Library of France](#)
- **Germany**
- [Deutsche Zentralbibliothek für Medizin](#)
- [Ernst-Moritz-Arndt-University of Greifswald](#)
- [Fernuniversität Hagen – Department for History and Biography](#)
- [Friedrich-Alexander-Universität Erlangen-Nürnberg – Pattern Recognition Lab](#)
- [Germanisches Nationalmuseum](#)
- [Gottfried Wilhelm Leibniz Bibliothek](#)
- [Heidelberg University, Department of History](#)
- [The Hessian State Archives](#)
- [Institute for Documentology and Scholarly Editing](#)
- [Museum for Musical Instruments of Leipzig University](#)
- [PLANET](#)

- [University and Research Library Erfurt/Gotha](#)
- [Hochschule Hanover – University of Applied Sciences and Arts](#)
- [Österreichischer Bibelübersetzer – Bavarian Academy of Science at the University of Augsburg](#)
- [University Library Tübingen](#)
- [University of Rostock – Edited Letters – Barlach 2020](#)
- [University of Rostock Library](#)
- **Greece**
- [National Bank of Greece Cultural Foundation](#)
- **Italy**
- [Archivo Storico Ricordi](#)
- [BOhisto – The Civic Archives of Bozen-Bolzano](#)
- [Marciana National Library](#)
- **Luxembourg**
- [Centre virtuel de la connaissance sur l'Europe – Digital Humanities Lab](#)

- **Netherlands**
- [Center for Global Heritage and Development \(Leiden University, Erasmus University Rotterdam, Delft University of Technology\)](#)
- [Cultural Heritage Agency](#)
- [Huygens Institute for the History of the Netherlands](#)
- [International Institut of Social History](#)
- **Norway**
- [Department of Cultural Studies and Oriental Languages, University of Oslo](#)
- [The Munch Museum](#)
- [National Library of Norway](#)
- **Serbia**
- [University Library Belgrade](#)
- **Spain**
- [National Library of Spain](#)
- [PROLOPE](#)
- [University of Girona – Center de Recerca d'Historia Rural](#)

- **Switzerland**
- [Alfred Escher Foundation](#)
- [Rechtsquellenstiftung des Schweizerischen Juristenvereins](#)
- [University Library Basel](#)
- [University of Bern, Digital Edition of Joseph Dietrich's Diary](#)
- [University of Zurich, History Department](#)
- **United Kingdom**
- [The British Library](#)
- [The Honourable Society of the Middle Temple – Library and Archive](#)
- [The Linnean Society](#)
- [London Metropolitan Archives](#)
- [The University of Aberdeen](#)
- **United States**
- [Northern Arizona University](#)
- [Western Michigan University – Kalamazoo](#)

Some highlights

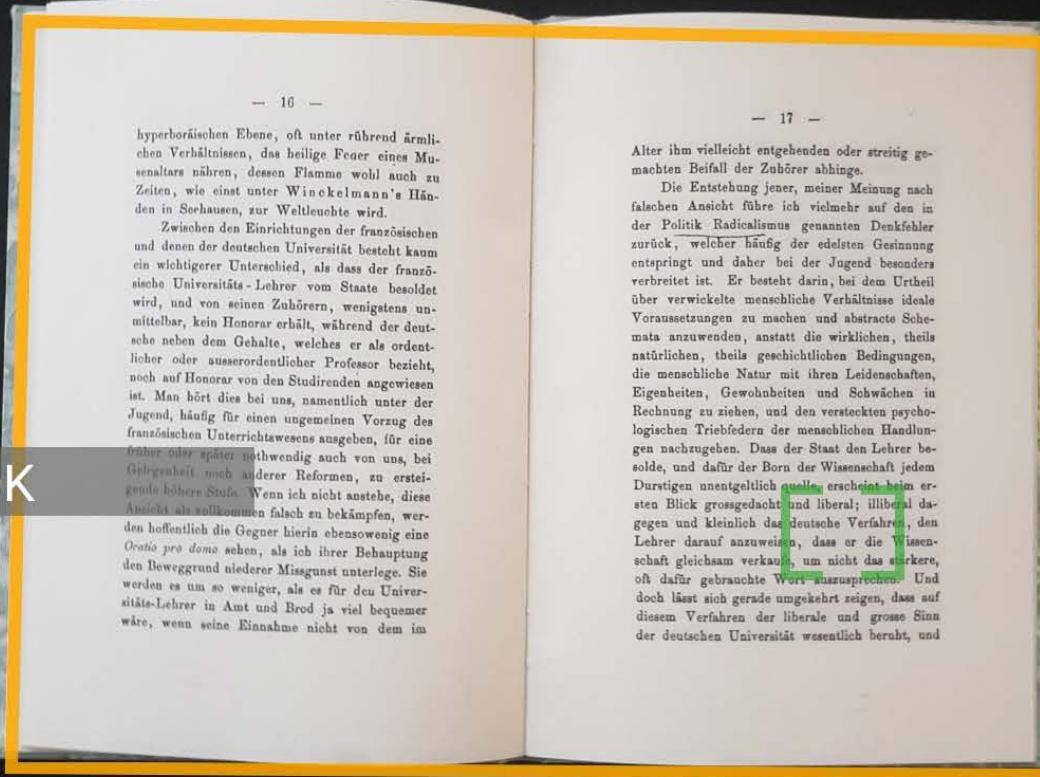
# Mission achieved...

- Make the whole process available to non-technical users
- Allow them to
  - Scan documents
  - Upload images to the platform
  - Segment and transcribe them
  - Train an HTR model
  - Measure the results
  - Apply the model to other documents
  - Carry out Keyword Spotting on the recognized documents
  - Export the data
- All this can also be performed via the API

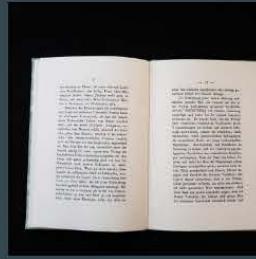




binder



OK



Single

- 8) Die vorliegende Urkunde ist gültig.  
 9) Sie liegt in der Verwaltung des Staatsbahns, kann jedoch bei  
 einer Untersuchung des polizeilichen oder anderen Rechtes auf  
 sie zur Verhinderung von Kosten, ferner bei Anwendung der  
 auf der Gebrauchserlaubnis liegenden Strafe, bis zu 1000 Kronen.  
 10) Sie ist allein für diejenigen Verträge ohne aufzufordernde Zeug-  
 stellungserklärung, welche nach dem Gütervertrag sowie nach den  
 Güterverträgen besonderen Spezialformularn geschlossen sind, in  
 Abhängigkeit von jenseitig zugesetzten Pfänden am Objekt der Güter-  
 verträge in Innsbruck und Pfändung zugleich.  
 11) Das Urteil wird in einem Exemplare, welches an den  
 im K.K. Staatsbahnen-Direktion Innsbruck vorliegt, in  
 jederzeitlich und auf Wunsch Josef Schmidmaier ein  
 abzugeben. Darauf Urkunde.

Kitzbühel, am 5. August 1910

Innsbruck, am 27. Oktober 1910  
 K.K. Staatsbahnen-Direktion Innsbruck Josef Schmidmaier  
 In K.K. Staatsbahnen-Direktion Innsbruck  
 L. S. Wagner <sup>pp</sup>  
 72 h vgt.  
 Laut Befreiungsurkunde Register Nr. 235 ist hier vom  
 Aufschmattet, bzw. Zentralamt Pfändung  
 der Güterverträge innerhalb —  
 von der Gerichts-Rampe der K.K. Bezirk  
 gerichtetes Kitzbühel.  
 Am 5. August 1910

L. S. Wagner <sup>pp</sup>

Der mit — Urkunden und 72h  
 Legalisierungstempel versehenen Urkunde  
 gleich  
 K. k. Bezirksgericht Kitzbühel  
 am 17. Dezember 1910

- 8) Die vorliegende Urkunde ist gültig.  
 9) Sie liegt in der Verwaltung des Staatsbahns, kann jedoch bei  
 einer Untersuchung des polizeilichen oder anderen Rechtes auf  
 sie zur Verhinderung von Kosten, ferner bei Anwendung der  
 auf der Gebrauchserlaubnis liegenden Strafe, bis zu 1000 Kronen.  
 10) Sie ist allein für diejenigen Verträge ohne aufzufordernde Zeug-  
 stellungserklärung, welche nach dem Gütervertrag sowie nach den  
 Güterverträgen besonderen Spezialformularn geschlossen sind, in  
 Abhängigkeit von jenseitig zugesetzten Pfänden am Objekt der Güter-  
 verträge in Innsbruck und Pfändung zugleich.  
 11) Das Urteil wird in einem Exemplare, welches an den  
 im K.K. Staatsbahnen-Direktion Innsbruck vorliegt, in  
 jederzeitlich und auf Wunsch Josef Schmidmaier ein  
 abzugeben. Darauf Urkunde.

Kitzbühel, am 5. August 1910

Innsbruck, am 27. Oktober 1910  
 K.K. Staatsbahnen-Direktion Innsbruck Josef Schmidmaier  
 In K.K. Staatsbahnen-Direktion Innsbruck  
 L. S. Wagner <sup>pp</sup>  
 72 h vgt.  
 Laut Befreiungsurkunde Register Nr. 235 ist hier vom  
 Aufschmattet, bzw. Zentralamt Pfändung  
 der Güterverträge innerhalb —  
 von der Gerichts-Rampe der K.K. Bezirk  
 gerichtetes Kitzbühel.  
 Am 5. August 1910

L. S. Wagner <sup>pp</sup>

Der mit — Urkunden und 72h  
 Legalisierungstempel versehenen Urkunde  
 gleich  
 K. k. Bezirksgericht Kitzbühel  
 am 17. Dezember 1910

# **CONGO BELGE**

Province de Saguenay-Lac-Saint-Jean  
District de la Matapedia

## OBSERVATIONS MÉTÉOROLOGIQUES

du mois de Janvier 1935 (faites à 6 heures.)

Terrain de Gagnéville  
Date du Jardin Botanique d'États  
Observateur J. Denis.

Date	BAROMÈTRE			THERM. S.A.		PSYCHROMÈTRE			ALTIMÉTRE		GIROUETTE		NURGES		EAU	
	TEMP.	HAUTEUR LIE	HAUTEUR OAS	MINIMA	EVAP.	SEC	HUMIDE	HUMIDITÉ	O	•	DIRECTIONS	FORCES	NEB.	FORMES	TOMBÉE	
1	22.8	736.0	734.2	325	1.0	22.9	22.4	14	22.9	22.1	Valise	0	3	ac	-	Brouillard léger
2	22.5	736.9	734.1	325	1.1	22.3	22.0	14	22.5	22.6	✓	1	6	cic	-	Bancs de Brouillard
3	22.1	736.3	734.3	325	1.1	22.1	22.7	14	22.1	22.2	x	1	5	ac	-	
4	22.0	736.9	734.2	325	0.9	22.0	22.4	15	22.5	22.6	6	1	3	cic	-	
5	22.0	736.0	734.3	325	0.1	22.0	22.6	16	22.4	22.5	8	1	10	cic	-	
6	22.2	736.0	734.3	325	1.0	22.3	22.9	16	22.3	22.4	Valise	0	7	cic	6.0	quelques gouttes lourdes
7	22.2	736.9	734.5	325	0.4	22.2	22.1	15	22.1	22.3	2	2	10	ac	-	
8	22.1	736.9	734.5	325	0.8	22.5	20.0	15	22.2	22.1	2	2	3	cic	-	
9	22.0	736.0	734.3	325	0.2	22.2	21.8	16	22.1	22.2	Valise	0	1	cic	-	
10	22.1	736.9	734.2	325	0.9	22.3	21.4	16	22.1	22.1	8	1	6	cic	18.0	
11	22.1	736.1	734.6	325	0.4	22.1	20.1	16	22.3	22.4	2	1	8	cic	-	Temps de 22 h 10 à 22 h
12	22.1	736.1	734.1	325	0.1	22.1	21.4	15	22.1	22.3	16	1	3	cic	-	
13	22.1	736.9	734.2	325	0.1	22.2	21.1	15	22.5	22.6	c	1	6	cic	-	Bancs de Brouillard
14	22.1	736.0	734.2	325	0.1	22.1	22.7	15	22.1	22.3	Valise	0	7	cic	5.1	température
15	22.0	736.4	734.7	325	0.3	21.5	22.8	15	20.7	20.8	2	3	10	n	1.4	plus de 4 h 30 à 1 h température
16	22.6	736.0	734.3	325	1.0	22.1	22.0	15	22.8	22.9	6	3	3	cic	-	
17	22.6	737.0	734.3	325	0.8	22.1	22.3	15	22.1	22.3	2	1	2	cic	-	
18	22.7	736.9	734.2	325	1.2	22.0	22.5	15	22.4	22.5	2	2	6	cic	26.0	grande de Brouillard
19	22.0	736.7	734.8	325	0.5	22.1	22.9	15	20.6	20.8	2	2	10	n	-	plus de 2 h 30 à 1 h température
20	22.1	736.0	734.3	325	0.9	22.2	21.5	16	22.3	22.5	8C	1	8	cic	-	
21	22.0	736.6	734.4	325	0.2	22.0	22.4	16	22.2	22.1	16	1	1	cic	-	
22	22.1	736.9	734.1	325	1.2	22.0	22.5	16	22.5	22.4	2	1	7	cic	-	Bancs de Brouillard
23	22.1	736.9	734.1	325	0.8	22.0	22.7	17	22.5	22.4	2	1	2	cic	0.8	Brouillard léger
24	22.0	736.9	734.1	325	0.1	22.1	22.0	16	22.1	22.1	16	1	1	cic	-	
25	22.1	736.7	734.0	325	0.6	22.5	21.9	15	22.5	22.6	16	1	2	cic	-	
26	22.1	736.0	734.4	325	0.9	22.3	21.1	15	22.1	22.0	2	1	7	cic	-	Bancs de Brouillard
27	22.1	736.9	734.2	325	0.8	22.6	21.1	15	22.1	22.2	NNW	1	7	cic	-	
28	22.1	736.9	734.2	325	0.2	22.4	22.9	16	22.7	22.6	Valise	0	7	ac	0.6	Bancs de Brouillard
29	22.5	737.0	734.2	325	0.5	22.9	21.1	15	22.1	22.3	NNW	1	1	cic	-	
30	22.1	736.0	734.7	325	0.3	22.2	21.9	15	22.5	22.6	8	1	6	cic	1.4	température le matin
31	22.1	736.0	734.1	325	0.1	22.6	21.1	15	22.1	22.5	56	1	2	n	-	plus température
MEILLEURS 22.1		736.1	734.1	325	325	22.3	22.1	15	22.6	22.6	2	1	7	cic	7.1	

MEYERSON 221 336.1 3041  
MAX ET MIN RESOLVE 28 T' 52°

七

三

**FREQUENCE DES VENTS**      **NOMBRES DE GROUPE**

~~1. NAME THE FIVE SENSES WHICH WE USE~~

TOTAL SW TONSEE 74 -

### Nombre de jours de pluie

DE BROWILLARD 1  
DE TONNERRE 2  
DE GRÈLE

Transkribus v1.4.1.13-SNAPSHOT (21\_06\_2018\_13:40), Loaded doc: HTR Bentham TEST, ID: 3158, Page 1, file: 116\_630\_001.jpg [Image Meta Info: (Resolution:300.0, w\*h: 2731 \* 4096)]

Server Overview Layout Metadata Tools

Layout Analysis

Method: CITLab Advanced Configure...

Current page  Pages (6): 1-6 ...

Find Text Regions  Use only unsegmented pages for LA

Find Lines in Text Regions

Run

Text Recognition

Method: HTR (CITLab)

Models... Train... Run...

Compute Accuracy

Reference: (Correct Text) 11.06.17 15:27:34 - guenter - In Progress - CITLab HTR: English Writing M1

Hypothesis: (HTR Text) 11.06.17 15:27:34 - guenter - In Progress - CITLab HTR: English Writing M1

Compare Compare Versions in Textfile

Other Tools

Current page  Pages (6): 1-6 ...

Add Baselines to Polygons Add Polygons to Baselines

more than six hours from morning to night, without  
allowing them leisure for any other employment."\*

"A page to "If the inhabitants were not unwilling  
to my observation and habitude of life," says the worthy Angle-  
rate,) I would not hesitate to contract for the profits of  
the prisoners! Blue, if they were allowed to work in  
their respective callings, and in such their labour as  
proper for prisoners,) and to pay the County £60 per  
annum."

Alas, though true for the moment, an observation  
suggested by the hypothetical offer, supported by Mr. H., it  
is now, my Lord, to the mind part of my test. God  
was good as it might be for the Committee to be forced  
to make over connections with these impudent persons.  
as it is, they will not all of them at least suffer them-  
selves to be thus paid. Submitting to the tax too ex-  
cessively and actually imposed upon them, paid by the  
Grace, and now by Your Lordship, for the maintenance  
of prisoners whom Parliament has ordered elsewhere, and  
concerning to this tax, perhaps for want of want of having  
so distinct a value of the noble friends to which they  
are indebted for it as this association will now help  
them to, they demand compensation to the so much  
higher as well as by directly imposed tax, also intended  
to be in store for them - the tax of adding them, & in  
considering with Blackbury associations. The tax for  
annum.

\* A specimen of the expense of the office establishment upon  
the state of the economic situation especially when in the gally-  
your road.

Transkribus v1.4.1.13-SNAPSHOT (21\_06\_2018\_13:40), Loaded doc: HTR Bentham TEST, ID: 3158, Page 1, file: 116\_630\_001.jpg [Image Meta Info: (Resolution:300.0, w\*h: 2731 \* 4096) ] [ current line: w\*h: 1733 \* 130 ]

Server Overview Layout Metadata Tools

Logout guenter

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

Collections:

DEMO (1457, Owner)

1-4 / 4

Title	Pages	Uploader	Uploaded
396 Bentham December 2015 (b)_duplicated	29	guenter	Thu Dec
19 Copy of Bentham May 2015	100	guenter	Tue Oct
43 Copy of HTR Bentham TEST	6	guenter	Thu Jan
58 HTR Bentham TEST	6	guenter	Thu No

37

116b  
630

37

"more than two turnheys from morning to night, without  
allowing them leisure for any other employment."\*

N<sup>o</sup> 8. Page 26. "If the undertaking were not unsuitable  
to my character and habits of life" (says the worthy Magis-  
trate) "I would not hesitate to contract for the profits of  
the prisoners' labour, (if they were allowed to work in  
their respective callings, and in such other labour as is

1 "more than two turnheys from morning to night, without ↵  
2 "allowing them leisure for any other employment.: ↵  
3 No 8. Page 26. "If the undertaking were not unsuitable ↵  
4 "to my character and habits of life" (says the worthy Magis- ↵  
5 -trate) "I would not hesitate to contract for the profits of ↵  
6 the prisoners' labour, (if they were allowed to work in ↵  
7 "their respective callings, and in such other labour as is ↵  
8 "proper for prisoners,) and to pay the County £400 per ↵

### HTR Training

<b>Model Name:</b> <input type="text" value="Test"/>	<b>Language:</b> <input type="text" value="Some language"/>
<b>Description:</b> <div style="border: 1px solid #ccc; padding: 5px; height: 150px; overflow-y: auto;"> Some description </div>	
<b>CITlab RNN</b> <input checked="" type="checkbox"/> <b>CITlab T2I</b> <input type="checkbox"/> <b>Nr. of Epochs:</b> <input type="text" value="20"/> <b>Learning Rate:</b> <input type="text" value="1e-3"/> <b>Noise:</b> <input type="text" value="both"/> <b>Train Size per Epoch:</b> <input type="text" value="10000"/> <b>Base Model:</b> <input type="button" value="Choose..."/> <input type="button" value="Reset to defaults"/>	
<input type="button" value="OK"/> <input type="button" value="Cancel"/>	

**Training Set**

- ▷  52494 - TEST\_CITlab\_NZZ 0.1\_duplicated (5 pages)
- ▷  51955 - TEST\_CITlab\_Tantis Letters Verda M1\_duplicated (6 pages)
  - Page 1 (33 lines)
  - Page 2 (37 lines)
  - Page 3 (28 lines)
  - Page 4 (28 lines)
  - Page 5 (25 lines)
  - Page 6 (29 lines)
- ▷  50833 - TEST\_CITlab\_est\_ouest\_M1\_duplicated (2 pages)
- ▷  50713 - TEST\_CITlab\_est uest\_duplicated (2 pages)
- ▷  50670 - TEST\_CITlab\_Holzmeister0.2t\_duplicated (4 pages)
- ▷  48783 - TEST\_CITlab\_WrDiarium\_M6\_sw3\_duplicated (5 pages)
- ▷  45504 - TEST\_CITlab\_Sigenot - M1\_duplicated (2 pages)
- ▷  44573 - TEST\_CITlab\_Selden M1\_duplicated (2 pages)
- ▷  44504 - HS\_116\_Test (92 pages)
- ▷  44378 - TEST\_CITlab\_Agriculture M3\_duplicated (2 pages)
- ▷  44376 - TEST\_CITlab\_Observations M2\_duplicated (2 pages)
- ▷  44325 - TEST\_CITlab\_Okopenko M1\_duplicated (61 pages)
- ▷  43488 - TEST\_CITlab\_Boston 19\_duplicated (1 pages)
- ▷  43419 - t2i\_example\_page (1 pages)
- ▷  43389 - TEST\_CITlab\_PROR 11 from photo images\_duplicated (3 pages)

**Overview**

Use Groundtruth versions  Use initial('New') versions

**Training Set**

ID	Title	Pages
51955	TEST_CITlab_Tantis Letters Verda M1_duplic...	1

**Test Set**

ID	Title	Pages
51955	TEST_CITlab_Tantis Letters Verda M1_duplic...	3

**Text Recognition Configuration**

**CITlab RNN HTR**

Name	Language	Curator	ID
Graphic_Novels_M3	English	tobias.hodel@...	4419
prob11_mfilm	English	mark.bell@nat...	4418
Marotte_M2	French	tobias.hodel@...	4417
PROB11_mfilm12	English	mark.bell@nat...	4416
Late Bentham 3	English	louise.seaward...	4396
Suppes	german	transkribus@d...	4376
PROB_mfilm	English	mark.bell@nat...	4356
Senckenberg_M2	German	tobias.hodel@...	4337
FinnishWarJournals M3	Finnish	lauri.hirvonen...	4336
Uqar	french	guenter.hackl...	4317
Froestl_M1	german	guenter.hackl...	4316
test	eng	dusankatusca...	4297
Noscemus_test_2	Latin	stefan.zatham...	4296
Graphic_Novels_M1	English	tobias.hodel@...	4279
Atelinerperunkirjat M2	Swedish	lauri.hirvonen...	4278
<b>De Ruyter M1</b>	<b>dutch</b>	<b>guenter.hackl...</b>	<b>4277</b>
Gusbeth	german	guenter.hackl...	4276
Edelfelt M10	Swedish	elisabeth.stub...	4256
Graphic_Novels_M1	English	tobias.hodel@...	4237
FinnishWarJournals M2	Finnish	lauri.hirvonen...	4236
My Mission Journal_2nd model	English	russell.ault7@...	4216
Itinera_Nova_SAL8121_M1	dutch	tobias.hodel@...	4196
new model	English	jscrane@gmai...	4177
DusanEX	slovak	dusankatusca...	4176
Bassermann_0.4	German	karcher@uni...	4157
RDehmelM1	German	Tamara.Terbul...	4156
new model	English	jscrane@gmai...	4138
William_Hey_M2	English	louise.seaward...	4137
foo	English	jscrane@gmai...	4136
Noscemus_test_1	Latin	stefan.zatham...	4116
William_Hey_M1	English	louise.seaward...	4097
DehmelM1	German	Tamara.Terbul...	4096
KB Altjesnitz Taufen	German	vit1@gmx.de	4076
Handschriftliche Archivquellen	German Kurre...	andreas.blum...	4059
ziraEarly	Hebrew	sinai.rusinek@...	4058
Bassermann_Manuscripts_0.2	German	karcher@uni...	4057
Zfirah2	Hebrew	sinai.rusinek@...	4056
Wiegand_4th_model	Many	david.lindema...	4036
Devanagari mixed M1	Hindi	guenter	4017
CourtRecords M1	Swedish	lauri.hirvonen...	4016
carte1	English	brown3@tcd...	3996
Prud-Homme_1858_M3	French	tobias.hodel@...	3980
Passau_HTR_plus	Unknown	Unknown	3979
Polish2	Polish	noahscheer@...	3978
Borum-Lyngby_1930-39_v2	Danish	cjk@aarhus.dk	3977

**Details**

Name:	De Ruyter M1	Language:	dutch
Description:	first model	Parameters:	
Nr. of Words:		Nr. of Lines:	
21549		3200	

[Show Train Set](#)   [Show Test Set](#)   [Show Character Set](#)

### Learning Curve

The graph displays the learning progress of the model over 31 epochs. The Y-axis represents Accuracy in CER (0% to 100%) and the X-axis represents Epochs (0 to 31). The blue line (CER Train) starts at ~98% and plateaus around 10% after epoch 15. The red line (CER Test) starts at ~98%, drops to ~25% by epoch 5, and then follows a similar downward trend as the train set, plateauing around 10% after epoch 15. A vertical green line is drawn at epoch 30.

**Accuracy in CER**

**Epochs**

**CER on Train Set: 4.25%**   **CER on Test Set: 10.11%**

Transkribus v1.4.3.4-SNAPSHOT Advanced Compare

All pages  Pages (5): 1-5 ... Options default (case sensitive) Compare

**Advanced Statistics**

Page	Word Error Rate	Char Error Rate	Word Accuracy	Char Accuracy	Bag Tokens Precision	Bag Tokens Recall	Bag Tokens F-Measure
Overall	7.34%	2.45%	94.87%	98.94%	93.35%	95.19%	0.9425992779783394

Page	Word Error Rate	Char Error Rate	Word Accuracy	Char Accuracy	Bag Tokens Precision	Bag Tokens Recall	Bag Tokens F-Measure
Page 1	5.68 %	1.31 %	95.17 %	99.24 %	94.63 %	95.17	0.9490084985835694 %
Page 2	12.03 %	3.16 %	91.6 %	98.49 %	89.77 %	92.36	0.9104385423100679 %
Page 3	6.14 %	1.23 %	94.95 %	99.21 %	94.37 %	95.18	0.9477351916376305 %
Page 4	11.0 %	3.96 %	91.87 %	97.92 %	89.61 %	92.05	0.9081376795076361 %
Page 5	4.61 %	2.02 %	97.56 %	99.55 %	96.05 %	97.96	0.9699350355038525 %

Base folder: C:\Users\c608178\Desktop

File/Folder name: DocId\_52494

Export path: C:\Users\c608178\Desktop\DocId\_52494.xls

Error Rate

Quick Compare

Compare Versions in Textfile

Other Tools

Current page  Pages (5): 1-5 ...

Add Baselines to Polygons

Add Polygons to Baselines

Region:

Desktop c608178  DEU 07:03

S. Publico-

Den 31<sup>st</sup>. Okt 1792.

Abbildung der Capitulations-  
Verhandlungen für den imstrun-  
tiuiss ynfurde & Koenig sind  
auf den vorigen Samstag abends  
zur Verficht stonden, woher  
entstehet die eigentliche Ursache  
dieser Verhandlung

283.-283  
In Publico - Politicis Publico-Politicis  
Den-den 31t Xbr 1792.  
Ausstellung der Gesundheits-  
zeugnisse zeugnisse für die nis-ins Vene-  
tianische gehenden Waaren wied  
auf die vorige Gewonheit werde  
zurückgeführt werden, wozu  
laut diesseitigen Bericht vom  
19ten v. M. Hoffnung vorhanden  
seyn solle.  
Conclusum.  
a Da der angezogene Bericht  
ur nur auf eine vom H: Joseph v  
Guñer dem H. Bürgermeister  
dieses Magistrats ertheilte  
spezielle auskunft Auskunft erstattet  
worden, als ist von demselben  
die Äußerung abzufordern, Außerung abzufordern, wie  
weit die in dieser Sache nach  
Venedig Venedig unterhaltene Korre-  
spondenz bereits gekommen,  
auch ob, auch ob, und bis welche Zeit die  
Wiederherstellung des aus

# Zürcher Zeitung.

Nro. 22.

Mittwoch.  
6 Februar, 1799.

Octo*d*.

18 Pluviose, Jahr VII.

Großbritannien. — Irland; Insurrektion. — Veränderungen im Ministerium. — Frankreich. — Spanien. — Deutschland; Wien, Hamburg. — Italien; Rom, Turin. — Helvetien.

Gr̄oßbritannien. London, 8 Januar. (Über Deutschland.) Es verbreitet sich das Gerücht, als ob künftigen Monat die Ausführung einer geheimen Expedition dem Prinzen Eduard anvertraut werden würde. — Zwei neue Regimenter haben Befehl erhalten, sich nebst einer Compagnie Artilleristen nach Gibraltar einzuschiffen, dessen Garnison durch den Abgang der nach Minorca gesegelten Truppen sehr geschwächt ist. Man hatte einige von unsren in Portugall befindlichen Truppen hinzobordert: der dösiae Hof hat sich aber aus Furcht vor

nion. Unter andern ward beschlossen, einen jeden, der eine Maßregel vorschlagen würde, deren Basis die Abschaffung des unabhängigen Parlements dieses Königreichs sey, als einen Feind Irlands, der britischen Verbindungen, und der Constitution zu betrachten.

Ebenda heut, 20 Jan. (Über Frankreich.) Unser Land ist wieder im Begrif in alle Schreden der Anarchie und des Bürgerkrieges zurückzufinden. Am 15 d. ist der Aufstand in der Grafschaft Clare (an der Westküste in der Provinz Munster) welche das vorzüglichste

Zürcher Zeitung.  
Nro. 22.  
Mittwoch.  
6 Februar, 1799.  
Oetodi. Octodi.  
13 Plupiose, 18 Pluviose, Jahr V. VII.

Großbritannien, — Irland; Insurrektion, Veränderungen im Ministerium. — Frankreich. — Spanien. — Spanien.—  
Deutschland; Wien, Wien, Hamburg. — Italien; Rom, Turin. — Hlvetien. Helvetien.

Großbritannien. London, 8 Jannar. 8 Januar.  
(über Deutschland.) Es verbreitet sich das Gerücht, als ob künftigen Monat die Ausführung einer geheimen Expedition dem Prinzen Eduard anvertraut werden würde.—  
Zwei neue Regimenter haben Befehl erhalten, sich nebst einer Compagnie Artilleristen nach Gibraltar einzuschiffen, dessen Garnison durch den Abgang der nach Minorka gesegelten Truppen sehr geschwächt ist. Man hatte einige von unsern unfern in Portugall befindlichen Truppen hin=hinbeordert; der dasige Hof hat sich aber aus Frucht vor Frankreich heftig widersezt.  
Ebendaher, 15. 15 Jan. Lord Camelford ist in Dover, als er eben im Begrif war, nach Frankreich zu entkommen, entdeckt, entdeckt, verhaftet, und nach London gebracht worden. Bei seiner Ankunft ward geheimer Staatsrath gehalten, und Lord Grenville durch einen Expressen nach der Stadt geholt. Er kam Abends um 10 Uhr, hatte eine lange Conferenz-Konferenz mit Pitt, sah aber seinen Verwandten, den Lord Camelford nicht, der noch bei einem Staatsboten in Verwahrung ist.  
In Brasilien hat sich in der Provinz „Bahia di Todos los Santos eine Insurrektion gezeigt, welche völlig auf fränkische Grundsätze gebaut war. Die Geistlichen waren die ersten, welche die Lehre der Freiheit und Gleichheit predigten. Indessen ist sie nach den bei der portugiesischen Regierung eingegangenen Nachrichten bereits wieder gedämpft.  
Das Schiff der Northumberland von 74 Kanon-

News – READ Project | Transkribus als Werkzeug | ProjektePhilologisch – U | Project - DEMO - Transkribus | https://transkribus.eu/r/ | https://transkribus.eu/r/ | +

← → C https://transkribus.eu/r/read/sandbox/application/?mihyo=1&collid=1457&docId=94682&pageId=2

Apps 6 Kalender Home - Research Par READ Call READ WEB READ WUI KWS Kalender PROJEKTE - Transkribus Sandbox

ruppn mynynz ganz allein.  
Als illymnum Cris y abn gab  
Gehalben mir:

1. <sup>geno</sup> Die Gehalde von vier Functionären 9000 fl  
2. <sup>geno</sup> Der Gehalt des Archivars des Curatoriums 400 fl  
3. <sup>geno</sup> Die Gehalte des Actuars, Dieners und Hausknechts 1335 fl  
4. dann die Livree des letzteren 46 fl  
5. Das Diurnum des Aushilfsindividuums in der  
Kanzlei vom 1ten Jänner an 300 fl  
6. Die Reise- und Präsenzaelder der zur feierlichen

x<sup>2</sup> x<sub>2</sub> B I U abc Special Characters Annotate ... ?! Unclear ▲

Text Region 3	
1	1tens Die Gehalte der vier Functionäre _____ 9000 fl
2	2tens Der Gehalt des Archivars des Curatoriums _____ 400 fl
3	3tens Die Gehalte des Actuars, Dieners und Hausknechts _____ 1335 fl
4	dann die Livree des letzteren _____ 46 fl
5	4tens Das Diurnum des Aushilfsindividuums in der Kanzlei vom 1ten Jänner an _____ 300 fl
6	5tens Die Reise- und Präsenzaelder der zur feierlichen
7	

In Bearbeitung

Save Changes

**Search for...**

Documents Fulltext (Solr) Tags KWS

Search in: Current document  Partial Matches  Case-sensitivity  Expert Syntax

Confidence Threshold: 0.05 <  >

Queries

Keyword 1 Mitterlehner

Keyword 2 Franz

Keyword 3 Thening

Keyword 4 Pasching

Keyword 5

Search

Search Results

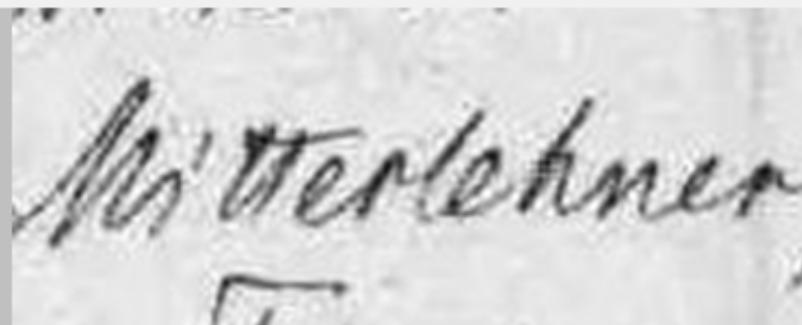
Created	Status	Queries	Duration	Scope
23.01.18 18:13:18	Completed	"MITTERLEHNER" (23), "FRANZ" (30), "MA...	4.54 sec.	Document 300...
29.12.17 13:22:14	Completed	"Mitterlehner" (17), "Thening" (4)	3.34 sec.	Document 300...
05.12.17 16:21:21	Completed	"Linz-Waldegg" (0)	1.15 sec.	Document 300...
23.11.17 14:26:20	Completed	"Thening" (4)	4.02 sec.	Document 300...
23.11.17 12:49:47	Completed	"Mitterlehner" (14)	1.39 sec.	Document 300...
23.11.17 12:47:49	Completed	"Krankenhaus" (3), "Thening" (16)	2.02 sec.	Document 300...
23.11.17 12:45:30	Completed	"Franz Mitterlehner" (3), "Franz" (5), "Mitter...	3.23 sec.	Document 300...

## Keyword Spotting Results

"Franz Mitterlehner" (3 hits) "Franz" (5 hits) "Mitterlehner" (7 hits) "Linimeir" (2 hits)

Confidence	Page Nr.	Line transcription	Prev...
0.4983	3	des Mitherlehaer	
0.4707	1	dranz Miterlekner, dienskaachtsein	
0.2080	1	Moiveshekner Maurrts ded siusbesitzas	
0.1445	4	Jofuch M strlofet	
0.1429	5	Josiuch Mberlehent	
0.0574	1	z e Mensrgaaeeghdnetuhes. Ms mernensgriduach Mti	
0.0530	4	t keeste seht Mnet lesnnehen Eni Me	

Preview



[Close](#)

# Mitterlehner - Moiveshekner

Activity

Page 12

ca... 26  
ated 5  
ed 388  
14-... 101  
322  
lic... 522  
79  
79  
236  
603

H V L ...

16 hetaes

17 1 Jäerer

18 Moiveshekner Maurrts ded siusbesit

19 if

### Export document

Server export  Client export

**TEST\_CITlab\_NZZ 0.1\_duplicated (52494)**

Current document  Current collection Choose documents to export...

**Finished server exports (not older than 2 weeks)**

```
id: 393496, finished: 04.10.2018 11:54:59, link: https://dbis-thure.uibk.ac.at/f/Get?id=DVMPKBNOMGPVQUSJRVXYWGHY  
id: 392906, finished: 03.10.2018 18:19:51, link: https://dbis-thure.uibk.ac.at/f/Get?id=COKQBAAOBVCSDTXVPKKXZWG  
id: 392906, finished: 03.10.2018 21:24:17, link: https://dbis-thure.uibk.ac.at/f/Get?id=VV1DVM7PDE071GUCWMB07SPCK
```

**Choose export formats**

Transkribus Document  
 PDF  
 TEI (restricted to server export)  
 DOCX  
 Simple TXT  
 Tag Export (Excel)  
 Table Export into Excel

---

Export ALL formats

---

Export Selected as ZIP

**Export options:**

Mets  PDF  TEI  DOCX

Export Page  
 Export ALTO (Line Level)  
 Export ALTO (Word Level)  
 Export Image

Image type:

Filename pattern

pageNr + filename  
 filename (warning: filenames must be unique for document)  
 docId + pageNr + pageId

Pattern: \${filename}

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

**Version status**

Latest version

Word based  
 Do blackening  
 Create Title Page

Pages (5):

# Future of Transkribus

# Future of Transkribus

- Governance model
  - Association?
  - Company with Limited Liability?
  - Foundation?
- Let's go for a model
  - Where cooperation is at the heart of its regulations
  - Which is open to institutions and not a closed circle
  - Which is able to keep a connection to all involved sectors – archives/libraries, scholars, volunteers and computer scientists
  - Which is flexible enough to provide services in a competitive environment
  - Where profit can be made in order to invest into new ideas, new projects, new partnerships
- Suggestion to set up a European Cooperative Society

# Business model – sources of income (1)

- Under discussion!
  - We want to find solutions which fit to the communities and their needs
  - We will not surprise our users
- Transkribus platform
  - All services will remain free for a limited amount of pages
  - For professional use some subscription or processing fees will be applied
- Customer projects
  - Large scale projects with archives and libraries
  - Models can be tuned and adapted to the specific needs of the project
  - High Performance Computing clusters allow for large scale processing – e.g. 200.000+ pages per day
  - Keyword spotting of large collections (millions of pages) as the main focus

# Business model – sources of income (2)

- Participation in grant applications
  - We are extremely thankful that we got the chance to develop Transkribus in two EU research projects
  - Of course we try to get involved in further research projects as well – either as service providers for text recognition or as research partners
  - Talk with us once you prepare your next grant application
- ScanTent – DocScan
  - Selling the ScanTent
  - Supporting archives and libraries to organize Scanathons
  - Linking digital repositories from archives and libraries with user generated scans

# 2019 and beyond

- Foundation of the READ-COOP
- NewsEye Project
  - EU Research project – University of LaRochelle (5/2018-4/2021)
  - Transkribus will be extended with layout analysis and article recognition developed by CITlab Team of the University of Rostock
- First large scale project with National Archive Finland
  - About 800.000 pages of court records
  - Recognition and Keyword Spotting
  - 250.000 EUR for implementing Transkribus and setting up a search site
- Project grant from Tyrolean government
  - 200.000 EUR for making cadastre documents from Austro-Hungarian Empire available via keyword spotting
- Negotiations with several (large) institutions from the Netherlands, Ireland, Canada and the US to use Transkribus in their environment or for specific projects

# Thank you a lot for your attention!

More information

<https://read.transkribus.eu/>

<https://transkribus.eu/>

<https://scantent.cvl.tuwien.ac.at/en/>

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943.

