# Transkribus

# A Platform for Automated Text Recognition and Searching of Historical Documents

Sebastian Colutto s.colutto@innovation-Innsbruck.at University of Innsbruck, Digitisation and Digital Preservation Group



# Agenda

- Introduction
- Technical Platform Overview
- Business Platform Overview READ-COOP
- Outlook

## Introduction

- What is Transkribus?
  - Digital Platform for Text Recognition & Enrichment
  - 2 EU Projects: Transcriptorium & READ



# Technical Platform Overview

instru fin mögte mut ob tis singsprifte. Ora



Artificial Neural Network

innere seyn mögte und ob die eingereichte. Druck.

#### Layout Analysis - How to produce line images?

und Huge Verankaltung/ Des Rapferl. General Feld. Marichall Lieutenants

#### und kluge Veranstaltung/des Käyserl.General Feld=Marschall Lieutnants

innow fin might mut of his singsprifte. coul.

innere seyn mögte und ob die eingereichte. Druck.



יוחנן בן נורי וכי מה אכפת להם הע

#### Transkribus v1.3.6.7-SNAPSHOT (31\_10\_2017\_13:56), Loaded doc: TUC, ID: 28551, Page 13, file: Noord-Hollands Archief2\_Nico Vriend.jpg [Image Meta Info: (Resolution:300.0, w\*h: 3619 \* 4844) ]

😑 🗉 🗔 🕐 🔤 🎉 🛱 🖑 Search current document... 🔍 🚺 🖣 13 /27 🕨 🍂 🖓 🖶 🖓 Page status: In Progress 🔍 🗋 🔄 🎭 😌 🔍 🧠 🔍 🚔 🚔 🚛 📟 🏈 🥹

🗿 TR 🛈 L 🕜 BL () W Ø ... 0

60 6 OD H

OD V opL

\*

8 2

**1**2 ...

Document Manager Versions Vost Recently Used Documents Ilections:  D 1-66 / 66 V 1 1 V Common Version ID Title 26594 TEST_CITlab_PROB 11 2670 TEST_CITlab_Humarec 26267 TEST_CITlab_Humarec 26267 TEST_CITlab_Humarec 26268 TEST_CITlab_PROB 11 25749 TEST_CITlab_PROB 11 25749 TEST_CITlab_PROB 11 25714 TEST_CITlab_Proce 25715 TEST_CITlab_Aranda M 25709 scantent_shots 25601 TEST_CITlab_Aranda M 25709 tEST_CITlab_Aranda M 25709 TEST_CITlab_Aranda M 25709 TEST_CITlab_Aranda M 25000 TEST_CITlab_RA_GEO_J 23000 TEST_CITlab_Ltinera N 23000 TEST_CITlab_STENC								
Versions           Most Recently Used Documents           Illections:           D           1-66 / 66         I         1         1           ID         Title           26594         TEST_CITIab_PROB 11           26270         TEST_CITIab_Humarec           26267         TEST_CITIab_Homarec           26267         TEST_CITIab_Probe 11           25849         TEST_CITIab_Probe 11           25849         TEST_CITIab_Probe 11           25714         TEST_CITIab_Probe 11           25715         TEST_CITIab_Probe 12           25714         TEST_CITIab_Cothic let           25715         TEST_CITIab_Catalogue           25714         TEST_CITIab_Cothic let           25306         TEST_CITIab_Cothic let           25307         TEST_CITIab_Todesurd           25907         TEST_CITIab_RA_GEO_1           25908         TEST_CITIab_RA_GEO_1           25909         TEST_CITIab_RA_GEO_1           25909         TEST_CITIab_RA_GEO_1           25000         TEST_CITIab_RA_GEO_1           25128         TEST_CITIab_RA_GEO_1           25297         TEST_CITIab_RA_GEO_1           25208         TEST_CITIab_RA_GEO_1     <	📸 Document Manager							
Nost Recently Used Documents         Ilections:         D         1-66 / 66 ↓ ↓ 1 1 ▶         ID         Title         26594         TEST_CITIab_PROB 11 1         26270         TEST_CITIab_PROB 11 1         26270         TEST_CITIab_MS A 265         26283         TEST_CITIab_MS A 265         26284         TEST_CITIab_PROB 700         2549         TEST_CITIab_Pairce M1         2573         testdoc_scantent         25714         TEST_CITIab_Gothic let         25709         scantent_shots         25001         TEST_CITIab_Todesurt         25297         TEST_CITIab_Rodesurt         25298         TEST_CITIab_RA_GEO_1         25209         25100         TEST_CITIab_RA_GEO_1         25297         TEST_CITIab_RA_GEO_1         25298         TEST_CITIab_RA_GEO_1         2509         TEST_CITIab_RA_GEO_1         24809         TEST_CITIab_RA_GEO_1         24785         TEST_CITIab_RA_GEO_1	Vost Recently Used Documents v							
Ilections: D Title D Title Constant C								
D 1-66 / 66 V 1 1 1 ID Title 26594 TEST_CITlab_PROB 11 26270 TEST_CITlab_PROB 11 26270 TEST_CITlab_MORA 2655 26238 TEST_CITlab_MAS A265 26238 TEST_CITlab_ABP Rottr 25919 TEST_CITlab_ABP Rottr 25919 TEST_CITlab_ABP Rottr 25714 TEST_CITlab_Gothic let 25715 TEST_CITlab_Gothic let 25715 TEST_CITlab_Hoffneinz 25714 TEST_CITlab_Catalogu: 25714 TEST_CITlab_Catalogu: 25715 TEST_CITlab_Gothic let 25306 TEST_CITlab_Todesurte 25297 TEST_CITlab_Todesurte 25297 TEST_CITlab_Todesurte 25297 TEST_CITlab_Todesurte 25297 TEST_CITlab_Todesurte 25297 TEST_CITlab_Todesurte 25297 TEST_CITlab_Todesurte 25298 TEST_CITlab_RA_GEO_I 2600 TEST_CITlab_RA_GEO_I 2600 TEST_CITlab_RA_GEO_I 2610 TEST_CITlab_RA_GEO_I 2610 TEST_CITlab_RA_GEO_I 2610 TEST_CITlab_RA_GEO_I 2610 TEST_CITlab_RA_GEO_I 2610 TEST_CITlab_Itinera_NA 2610 TEST_CITlab_Itinera_NA 2610 TEST_CITlab_Itinera_NT 26264 TEST_CITlab_Itinera_NT 26207 TEST_CITlab_StarL_Pro.								
1-66 / 66     ↓     ↓     1 1 ▶       ID     Trile       26594     TEST_CITIab_PROB 11       26270     TEST_CITIab_Humarec       26267     TEST_CITIab_MS A 265       26283     TEST_CITIab_MS A 265       26284     TEST_CITIab_MS A 265       26285     TEST_CITIab_MS PROttr       25719     TEST_CITIab_ABP Rottr       25721     testdoc_scantent       25717     TEST_CITIab_Gothic let       25718     TEST_CITIab_Catalogu       25719     TEST_CITIab_Todesurt       25709     scantent_shots       25001     TEST_CITIab_Fraktur fri       25298     TEST_CITIab_Fraktur fri       25278     KWS-DEMO_HS_118_di       25009     TEST_CITIab_RA_GEO_J       26009     TEST_CITIab_RA_GEO_J       28090     TEST_CITIab_RA_GEO_J       28090     TEST_CITIab_RA_GEO_J       28090     TEST_CITIab_RA_GEO_J       28100     TEST_CITIab_RA_GEO_J       28207     TEST_CITIab_Linter_ANC       28207     TEST_CITIab_Linter_ANC       28207     TEST_CITIab_Linter_ANC       28207     TEST_CITIab_Star4_Prov.t       28207     TEST_CITIab_LinterANC       28207     TEST_CITIab_Star4_Prov.t       28207     TEST_CITIab_Linter <anc< t<="" td=""><td>DEMO (1457, Owne</td><td>er)</td><td></td></anc<>	DEMO (1457, Owne	er)						
ID         Title           26594         TEST_CITIab_PROB 11           26270         TEST_CITIab_MS A 265           2628         TEST_CITIab_MS A 265           2628         TEST_CITIab_Proce MI           25919         TEST_CITIab_Pairce MI           25738         testdok_guenter           25714         TEST_CITIab_Cothic let           25715         TEST_CITIab_Cothic let           25716         TEST_CITIab_Cothic let           25717         TEST_CITIab_Cothic let           25718         TEST_CITIab_Cothic let           25719         TEST_CITIab_Todesurtu           25709         scantent_shots           25001         TEST_CITIab_Todesurtu           25298         TEST_CITIab_Todesurtu           25297         TEST_CITIab_Todesurtu           25208         TEST_CITIab_RA_GEO_L           25090         TEST_CITIab_RA_GEO_L           26000         TEST_CITIab_RA_GEO_L           28009         TEST_CITIab_RA_GEO_L           28100         TEST_CITIab_RA_GEO_L           28101         TEST_CITIab_RA_GEO_L           28102         TEST_CITIab_RA_GEO_L           28103         TEST_CITIab_RA_GEO_L           28104         TEST_CITIab_RA_GE	N 0 0 E	• •						
26594     TEST_CITIab_PROB 11       26270     TEST_CITIab_MUMARC       26267     TEST_CITIab_MIX A 265       26283     TEST_CITIab_ABP Rotting       25919     TEST_CITIab_ABP Rotting       25849     TEST_CITIab_ABP Rotting       25714     TEST_CITIab_ABP Rotting       25715     TEST_CITIab_ABP Rotting       25716     TEST_CITIab_Gothic let       25717     TEST_CITIab_Hotheinz       25718     TEST_CITIab_Hotheinz       25719     TEST_CITIab_Hotheinz       25714     TEST_CITIab_Todesurting       25060     TEST_CITIab_Todesurting       25287     TEST_CITIab_Todesurting       25298     TEST_CITIab_Todesurting       25209     TEST_CITIab_Todesurting       25209     TEST_CITIab_Todesurting       25209     TEST_CITIab_RAGEO_I       25000     TEST_CITIab_RAGEO_I       25000     TEST_CITIab_RAGEO_I       26000     TEST_CITIab_RAGEO_I       28009     TEST_CITIab_RAGEO_I       28100     TEST_CITIab_RAGEO_I       28101     TEST_CITIab_RAGEO_I       28102     TEST_CITIab_RAGEO_I       28103     TEST_CITIab_RAGEO_I       28104     TEST_CITIab_RAGEO_I       28105     TEST_CITIab_RAGEO_I       28106     TEST_CITIab_RAGE		Pages	Uploader					
26270 TEST_CITIab_Humarec 26267 TEST_CITIab_MS A 265 26238 TEST_CITIab_MS A 265 26238 TEST_CITIab_ABP Rotto 25949 TEST_CITIab_Peirce M1 25738 testdoc_scantent 25711 TEST_CITIab_Peirce M1 25715 TEST_CITIab_Gothic let 25716 TEST_CITIab_Gothic let 25717 TEST_CITIab_Catalogue 25714 TEST_CITIab_Catalogue 25714 TEST_CITIab_Catalogue 25010 TEST_CITIab_Todesurto 25208 TEST_CITIab_Todesurto 25208 TEST_CITIab_Todesurto 25208 TEST_CITIab_Todesurto 25208 TEST_CITIab_Todesurto 25208 TEST_CITIab_Todesurto 25209 TEST_CITIab_Todesurto 25208 TEST_CITIab_Todesurto 25208 TEST_CITIab_Todesurto 25209 TEST_CITIab_Todesurto 25209 TEST_CITIab_RA_GEO_J 23610 TEST_CITIab_RA_GEO_J 23610 TEST_CITIab_RA_GEO_J 23610 TEST_CITIab_RA_GEO_J 23650 TEST_CITIab_Conter_Th 23208 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Start_Pro	11 1840s M1 (25_0	4	guenter.hackl					
26267     TEST_CITIab_MS A 265       2628     TEST_CITIab_ABP Rott       25919     TEST_CITIab_ABP Rott       25849     TEST_CITIab_Perice M1       25711     testdoc_scantent       25712     testdoc_scantent       25713     TEST_CITIab_Contic let       25714     TEST_CITIab_Contalogue       25709     scantent_shots       25001     TEST_CITIab_Todesurt       25297     TEST_CITIab_Fraktur fn       25208     TEST_CITIab_Fraktur fn       25209     TEST_CITIab_Fraktur fn       25209     TEST_CITIab_RA_GEO_       23000     TEST_CITIab_RA_GEO_       23000     TEST_CITIab_RA_GEO_       23010     TEST_CITIab_RA_GEO_       23020     TEST_CITIab_RA_GEO_       23030     TEST_CITIab_RA_GEO_       23610     TEST_CITIab_RA_GEO_       23610     TEST_CITIab_RA_GEO_       23610     TEST_CITIab_RA_GEO_       23610     TEST_CITIab_RA_GEO_       23610     TEST_CITIab_Littore_NA       23610     TEST_CITIab_Littore_NA       23640     TEST_CITIab_Littore_NA       23641     TEST_CITIab_Littore_NA       23641     TEST_CITIab_StazH_Prot       23182     TEST_CITIab_StazH_Prot	ec Arabic M1_dup	. 1	guenter					
26238     TEST_CITIab_WrDieariu       25919     TEST_CITIab_Peirce M1       25849     TEST_CITIab_Peirce M1       25731     testdoc_scantent       25717     TEST_CITIab_Gothic let       25716     TEST_CITIab_Catalogue       25717     TEST_CITIab_Catalogue       25718     TEST_CITIab_Catalogue       25719     TEST_CITIab_Todesurte       25709     scantent_shots       25601     TEST_CITIab_Todesurte       2528     TEST_CITIab_Todesurte       25298     TEST_CITIab_Todesurte       25000     TEST_CITIab_Todesurte       25298     TEST_CITIab_PorvocOc       24009     TEST_CITIab_RA_GEO_I       23009     TEST_CITIab_RA_GEO_I       23610     TEST_CITIab_RA_GEO_I       23610     TEST_CITIab_GNM_typ       23651     TEST_CITIab_Libnera_TH       23664     TEST_CITIab_Libnera_TH       2308     TEST_CITIab_Libnera_TH       2308     TEST_CITIab_Uiber Exte       2307     TEST_CITIab_Librera_TH       2308     TEST_CITIab_Uiber Exte       2307     TEST_CITIab_Start_Pro       23182     TEST_CITIab_Start_Pro	654 M2_duplicated	2	guenter					
25919 TEST_CITIab_ABP Rottr 25849 TEST_CITIab_Peirce M1 25738 testdoc_scantent 25721 testdok_guenter 25717 TEST_CITIab_Gothic let 25716 TEST_CITIab_Hofheinz 25715 TEST_CITIab_Hofheinz 25715 TEST_CITIab_Hofheinz 25714 TEST_CITIab_Gothic let 25306 TEST_CITIab_Todesurte 25298 TEST_CITIab_Todesurte 25297 TEST_CITIab_Todesurte 25297 TEST_CITIab_Todesurte 25298 TEST_CITIab_PorvooCc 24099 TEST_CITIab_PorvooCc 24099 TEST_CITIab_RA_GEO_I 23601 TEST_CITIab_RA_GEO_I 23601 TEST_CITIab_RA_GEO_I 23609 TEST_CITIab_RA_GEO_I 23609 TEST_CITIab_RA_GEO_I 23609 TEST_CITIab_RA_GEO_I 23606 TEST_CITIab_Itinera_NA 25654 TEST_CITIab_Itinera_NA 23207 TEST_CITIab_Itinera_NA 23207 TEST_CITIab_Itinera_NA 23208 TEST_CITIab_Itinera_TH 23208 TEST_CITIab_Itinera_TH	arium New M1_du	4	guenter					
25849 TEST_CITIab_Peirce M1 25738 testdoc_scantent 25717 TEST_CITIab_Gothic let 25716 TEST_CITIab_Gothic let 25716 TEST_CITIab_Hofheinz 25715 TEST_CITIab_Hofheinz 25717 TEST_CITIab_Hofheinz 25714 TEST_CITIab_Hoftic let 25060 TEST_CITIab_Todesurkt 25298 TEST_CITIab_Todesurkt 25298 TEST_CITIab_Todesurkt 25297 TEST_CITIab_Todesurkt 25297 TEST_CITIab_Todesurkt 25298 TEST_CITIab_Todesurkt 25298 TEST_CITIab_Todesurkt 25298 TEST_CITIab_RA_GEO_I 24785 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_RA_GEO_I 23606 TEST_CITIab_RA_GEO_I 23606 TEST_CITIab_RA_GEO_I 23656 TEST_CITIab_Itinera_NK 23564 TEST_CITIab_Itinera_NK 23208 TEST_CITIab_Itinera_TH 23208 TEST_CITIab_Itinera_TH 23208 TEST_CITIab_Itinera_TST_CITIab_Itinera_TST_CITIab_Itinera_TST_CITIab_Itinera_TST_CITIab_Itinera_TST_CITIab_Itinera_TST_CITIab_StarL_PTC 23182 TEST_CITIab_Ster.Bit	ottmayr_duplicated	3	guenter					
25738     testdoc_scantent       25721     testdok_guenter       25711     TEST_CITlab_Gothic let       25715     TEST_CITlab_Lotheinz       25714     TEST_CITlab_Catalogue       25715     TEST_CITlab_Catalogue       25714     TEST_CITlab_Gothic let       2500     scantent_shots       2500     TEST_CITlab_Todesurt       25298     TEST_CITlab_Todesurt       25209     TEST_CITlab_RacGoL       25000     TEST_CITlab_RA_GEO_       25029     TEST_CITlab_RA_GEO_       25030     TEST_CITlab_RA_GEO_       25040     TEST_CITlab_RA_GEO_       24809     TEST_CITlab_RA_GEO_       23009     TEST_CITlab_RA_GEO_       23010     TEST_CITlab_RA_GEO_       23610     TEST_CITlab_RA_GEO_       23660     TEST_CITlab_Litnera_NA       23661     TEST_CITlab_Litnera_NA       23664     TEST_CITlab_Litnera_NA       23607     TEST_CITlab_Litnera_NA       23640     TEST_CITlab_Litnera_NA       23207     TEST_CITlab_StazH_Prot_       23182     TEST_CITlab_Sterk.Pit	TEST_CITIab_Peirce M1_duplicated							
25721 testdok_guenter 25717 TEST_CITIab_Gothic let 25716 TEST_CITIab_Catalogue 25715 TEST_CITIab_Catalogue 25714 TEST_CITIab_Catalogue 25709 scantent_shots 25601 TEST_CITIab_Gothic let 25306 TEST_CITIab_Todesurkt 25298 TEST_CITIab_Faktur frr 25278 KWS-DEMO_HS_118_d0 25080 TEST_CITIab_Faktur frr 25278 EST_CITIab_Faktur frr 25278 EST_CITIab_RA_GEO_ 24809 TEST_CITIab_RA_GEO_ 23600 TEST_CITIab_RA_GEO_ 23610 TEST_CITIab_RA_GEO_ 23610 TEST_CITIab_RA_GEO_ 23564 TEST_CITIab_Itinera_NC 23564 TEST_CITIab_Itinera_Th 23208 TEST_CITIab_Itinera_Th 23208 TEST_CITIab_Itinera_Th 23208 TEST_CITIab_Itinera_Th 23208 TEST_CITIab_Itinera_Th 23208 TEST_CITIab_Itinera_Th 23208 TEST_CITIab_Itinera_Th 23208 TEST_CITIab_StazL_Pro-	testdoc_scantent							
25717 TEST_CITIab_Gothic let 25716 TEST_CITIab_Cataloguz 25715 TEST_CITIab_Cataloguz 25714 TEST_CITIab_Cataloguz 25714 TEST_CITIab_Gothic let 25306 TEST_CITIab_Gothic let 25306 TEST_CITIab_Todesurt 25297 TEST_CITIab_Todesurt 25297 TEST_CITIab_PorvooCc 24099 TEST_CITIab_PorvooCc 24099 TEST_CITIab_RA_GEO_I 24785 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_RA_GEO_I 23600 TEST_CITIab_GNM_typ 23565 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_StazH_Proc	testdok_guenter							
25716 TEST_CITIab_Hofheinz 25715 TEST_CITIab_Catalogus 25714 TEST_CITIab_Catalogus 25714 TEST_CITIab_Catalogus 25709 scantent_shots 25000 TEST_CITIab_Todesurtd 25298 TEST_CITIab_Todesurtd 25297 TEST_CITIab_Todesurtd 25297 TEST_CITIab_Todesurtd 25297 TEST_CITIab_PorvooCc 24009 TEST_CITIab_RA_GEO_1 24785 TEST_CITIab_RA_GEO_1 23600 TEST_CITIab_RA_GEO_1 23600 TEST_CITIab_RA_GEO_1 23600 TEST_CITIab_RA_GEO_1 23600 TEST_CITIab_GNM_typ 23565 TEST_CITIab_timera_NA 23564 TEST_CITIab_timera_TH 23208 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23207 TEST_CITIab_StazH_Proc 23182 TEST_CITIab_StartB_NC	letter - M1_duplic	23	guenter					
25715 TEST_CITIab_Catalogue 25714 TEST_CITIab_Aranda M 25709 scantent_shots 25601 TEST_CITIab_Gothic let 25306 TEST_CITIab_Todesurkt 25298 TEST_CITIab_Todesurkt 25297 TEST_CITIab_Todesurkt 25297 TEST_CITIab_Todesurkt 25208 TEST_CITIab_Todesurkt 25080 TEST_CITIab_RA_GEO_I 24099 TEST_CITIab_RA_GEO_I 24099 TEST_CITIab_RA_GEO_I 23609 TEST_CITIab_RA_GEO_I 23609 TEST_CITIab_RA_GEO_I 23606 TEST_CITIab_CMM_stp 23565 TEST_CITIab_Itimera_NK 23564 TEST_CITIab_Itimera_NK 23208 TEST_CITIab_Itimera_NK 23208 TEST_CITIab_Liber Exte 23200 TEST_CITIab_StazH_Proc 23182 TEST_CITIab_Ster.Bit	nz M1_duplicated	2	guenter					
25714 TEST_CITlab_Aranda M 25709 scantent_shots 25601 TEST_CITlab_Gothic let 25306 TEST_CITlab_Todesurts 25297 TEST_CITlab_Todesurts 25297 TEST_CITlab_Todesurts 25297 TEST_CITlab_Todesurts 25080 TEST_CITlab_RA_GEO_1 24809 TEST_CITlab_RA_GEO_1 24809 TEST_CITlab_RA_GEO_1 23600 TEST_CITlab_RA_GEO_1 23600 TEST_CITlab_RA_GEO_1 23600 TEST_CITlab_RA_GEO_1 23564 TEST_CITlab_Itinera_NA 23564 TEST_CITlab_Itinera_NA 23207 TEST_CITlab_Itinera_NA 23207 TEST_CITlab_Itinera_NA 23207 TEST_CITlab_Itinera_NA 23207 TEST_CITlab_Itinera_NA	gus Lectionem M2	. 2	guenter					
25709         scantent_shots           25601         TEST_CITIab_Gothic let           25306         TEST_CITIab_Todesurt           25297         TEST_CITIab_Fraktur fr           25287         FEST_CITIab_Fraktur fr           25298         TEST_CITIab_Fraktur fr           25299         TEST_CITIab_Fraktur fr           25080         TEST_CITIab_RA_GEO_           24809         TEST_CITIab_RA_GEO_           23809         TEST_CITIab_Littorer_ANC           23864         TEST_CITIab_Littorer_ANC           23864         TEST_CITIab_Littorer_ANC           23807         TEST_CITIab_Littorer_MC           23807         TEST_CITIab_StazH_Proc           23820         TEST_CITIab_Sterer_MC	M3_duplicated	5	guenter					
25601         TEST_CITIab_Gothic let           25306         TEST_CITIab_Todesurta           25298         TEST_CITIab_Todesurta           25297         TEST_CITIab_Todesurta           25297         TEST_CITIab_Todesurta           25297         TEST_CITIab_Todesurta           25207         TEST_CITIab_Todesurta           25208         TEST_CITIab_PorvooCc           24090         TEST_CITIab_RA_GEO_1           24785         TEST_CITIab_RA_GEO_2           23809         TEST_CITIab_RA_GEO_2           23610         TEST_CITIab_Tapp_Got           23565         TEST_CITIab_Litinera_NA           23564         TEST_CITIab_Litinera-TH           23200         TEST_CITIab_Litinera-TH           23201         TEST_CITIab_Litinera-TH           23202         TEST_CITIab_Litinera-TH           23203         TEST_CITIab_Litinera-TH           23204         TEST_CITIab_Litinera-TH           23205         TEST_CITIab_StazH_Proc           23182         TEST_CITIab_StazH_Proc		7	guenter					
25306         TEST_CITIab_Todesurta           25298         TEST_CITIab_Traktur fn           25297         TEST_CITIab_Traktur fn           25297         TEST_CITIab_Traktur fn           25298         KWS-DEMO_HS_118_di           25080         TEST_CITIab_RA_GEO_I           24090         TEST_CITIab_RA_GEO_I           24809         TEST_CITIab_RA_GEO_I           23809         TEST_CITIab_RA_GEO_I           23610         TEST_CITIab_GIM_typ           23565         TEST_CITIab_Litinera_NA           23564         TEST_CITIab_Liber Exte           23207         TEST_CITIab_Liber Exte           23208         TEST_CITIab_StazH_Proc           23206         TEST_CITIab_StazH_Proc           23182         TEST_CITIab_Stern Bit	letter - M1_duplic	23	guenter					
25298         TEST_CITIab_Todesurte           25297         TEST_CITIab_Fraktur fm           25278         KWS-DEMO_HS_118_d           25080         TEST_CITIab_PorvoCc           24009         TEST_CITIab_RA_GEO_J           24785         TEST_CITIab_RA_GEO_J           23809         TEST_CITIab_RA_GEO_J           23610         TEST_CITIab_RA_GEO_J           23565         TEST_CITIab_GNM_typ           23565         TEST_CITIab_timera_NA           2364         TEST_CITIab_timera_TH           2308         TEST_CITIab_timera_TH           2309         TEST_CITIab_timera_TH           2310         TEST_CITIab_timera_TH           23207         TEST_CITIab_timera_TH           2310         TEST_CITIab_timera_TH           2312         TEST_CITIab_timera_TH	urteile 1891_duplic	19	guenter					
25297         TEST_CITIab_Fraktur fm           25278         KWS-DEMO_HS_118_dt           25080         TEST_CITIab_RA_GEO_J           24809         TEST_CITIab_RA_GEO_J           24809         TEST_CITIab_RA_GEO_J           23809         TEST_CITIab_RA_GEO_J           23809         TEST_CITIab_RA_GEO_J           23610         TEST_CITIab_RA_GEO_MM_typ           23565         TEST_CITIab_Itinera_NX           23564         TEST_CITIab_Ltinera_TH           23208         TEST_CITIab_Ltinera_TH           23208         TEST_CITIab_StazH_Proc           23182         TEST_CITIab_Staren_MSternera_NX	urteile 1849_duplic	7	guenter					
25278         KWS-DEMO_HS_118_dd           25080         TEST_CITIab_RA_GEO_J           24099         TEST_CITIab_RA_GEO_J           24785         TEST_CITIab_RA_GEO_J           23809         TEST_CITIab_RA_GEO_J           23610         TEST_CITIab_Tapp_Got           23565         TEST_CITIab_GNM_typ           23564         TEST_CITIab_Litinera_Nd           23565         TEST_CITIab_Litinera_Th           23208         TEST_CITIab_Litinera_Th           23207         TEST_CITIab_Litinera_Th           23208         TEST_CITIab_Litinera_Th           23206         TEST_CITIab_Litinera_Th           23207         TEST_CITIab_Litinera_Th           23208         TEST_CITIab_Litinera_Th           23209         TEST_CITIab_Starp.Pro           23182         TEST_CITIab_Starp.Pro	25297 TEST_CITIab_Fraktur from OCR_duplic							
25080         TEST_CITIab_PorvooCc           24809         TEST_CITIab_RA_GEO_J           24785         TEST_CITIab_RA_GEO_J           23809         TEST_CITIab_RA_GEO_J           23801         TEST_CITIab_Tapp_Got           23565         TEST_CITIab_CNM_typ           23564         TEST_CITIab_Ltimera_Nc           23505         TEST_CITIab_Ltimera_Th           23006         TEST_CITIab_Liber Exte           23207         TEST_CITIab_Laber Average           23208         TEST_CITIab_Liber Exte           23206         TEST_CITIab_Start_Proc           23102         TEST_CITIab_Start_Proc	_duplicated	603	guenter					
24809         TEST_CITIab_RA_GEO_J           24785         TEST_CITIab_RA_GEO_J           23809         TEST_CITIab_289recto_           23610         TEST_CITIab_Tapp_Got           23555         TEST_CITIab_timera_NA           23564         TEST_CITIab_timera_NA           23565         TEST_CITIab_timera_TA           23208         TEST_CITIab_timera_TA           23207         TEST_CITIab_timera_TA	80 TEST_CITIab_PorvooCourtRecords_du							
24785         TEST_CITIab_RA_GEO_J           23809         TEST_CITIab_289recto_           23610         TEST_CITIab_Tapp_Got           23556         TEST_CITIab_Inter_A_N           23564         TEST_CITIab_timera_N           2308         TEST_CITIab_timera_N           2308         TEST_CITIab_timera_N           2308         TEST_CITIab_timera_N           2309         TEST_CITIab_timera_N           2300         TEST_CITIab_timera_N           2301         TEST_CITIab_timera_N           23020         TEST_CITIab_timera_N           23206         TEST_CITIab_timera_N           23207         TEST_CITIab_timera_N           23208         TEST_CITIab_Start_Pro           23182         TEST_CITIab_Start_P	TEST_CITIab_RA_GEO_M1_duplicated							
23809         TEST_CITIab_289recto_           23610         TEST_CITIab_Tapp_Got           23566         TEST_CITIab_Itinera_Nc           23567         TEST_CITIab_Litinera_Nc           23564         TEST_CITIab_Litinera_Nc           23264         TEST_CITIab_Litinera_Nc           23207         TEST_CITIab_Litinera_Nc           23206         TEST_CITIab_Litinera_Nc           23207         TEST_CITIab_Litinera_Nc           23206         TEST_CITIab_StazH_Proc           23182         TEST_CITIab_Stere.Bit	5 TEST_CITIab_RA_GEO_M1_duplicated							
23610 TEST_CITlab_Tapp_Got 23566 TEST_CITlab_GNM_typ 23565 TEST_CITlab_Itinera_No 23564 TEST_CITlab_Itinera-Th 23208 TEST_CITlab_Liber Exte 23207 TEST_CITlab_Liber Exte 23206 TEST_CITlab_StazH_Pro 23182 TEST_CITlab_Östern-Bil	23809 TEST_CITIab_289recto_210717_duplica							
23566 TEST_CITIab_GNM_typ 23565 TEST_CITIab_Itinera_No 23564 TEST_CITIab_Itinera-Th 23208 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23206 TEST_CITIab_StazH_Pro 23182 TEST_CITIab_Stare.Bin 23182 TEST_CITIab_Stare.Bin	othic_duplicated	6	guenter.hackl					
23565 TEST_CITIab_Itinera_No 23564 TEST_CITIab_Itinera-Th 23208 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Liber Exte 23206 TEST_CITIab_StazH_Pro 23182 TEST_CITIab_Start. Pin	yposkript_duplica	10	guenter.hackl					
23564 TEST_CITIab_Itinera-TH 23208 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Hebrew_t 23206 TEST_CITIab_StazH_Pro 23182 TEST_CITIab_Österr, Bil	Nova_M1_duplica	. 5	guenter.hackl					
23208 TEST_CITIab_Liber Exte 23207 TEST_CITIab_Hebrew_t 23206 TEST_CITIab_StazH_Pro 23182 TEST_CITIab_Österr. Bil	Thun-M1_duplica	. 17	guenter.hackl					
23207 TEST_CITIab_Hebrew_t 23206 TEST_CITIab_StazH_Pro 23182 TEST_CITIab_Österr. Bil	xtended M8_dupli	4	guenter.hackl					
23206 TEST_CITIab_StazH_Pro 23182 TEST_CITIab_Österr. Bil	v_t2i_test_duplicat	4	guenter.hackl					
23182 TEST CITIab Österr, Bil	Protokolle_t2i_du	9	guenter.hackl					
	TEST_CITIab_Österr. Bibelübersetzer M							
22599 TEST_CITIab_UB Belgra	grade Card Catalo	11	guenter					
22595 TEST_CITIab_Berlin Hs1	Hs17_duplicated	3	guenter					
20457 TEST CITIsh Realizatal	aloa M1 dunlicat	22	quenter					



\*

\_ 🗇 🗙

some del doi 48 14 + Mingelo and Jamie Burch more Berliner Balandolina U. In Pratien to fallow the adare mediane & que fi little by som and have the son, hope at this will shade in firming future Part Schugun gren m THE NET THE PERSON OF THE she as itere forestion beathy eter this the set fores and in few this first to good case as well any a few shaps as go from the to be weathing was a See anylo aleque about his - I server fin Toger any get a + forth fait grageres land tring ten of stigle grapen fate, bes men to grapes to age and and to be a plant to the star escudie mediante die rolpon esterio et practice die m Sidhan to to the to make a Salentinered Sugardus - Sylinites The 20 Section Carfadone ilso to trades the spifter - not linterester - - that an another days the inst mill quarter wind Callman the for the group · Carpenner to Section -humpering in Sufficher The form the alarida all invitigation performed the move minforce of his contar she aligness and top an eje present of orber seen for the times has fighing sime the Bulle being

								CONG	O I	ELG	E.					
man de le Estanjen 3						OBSERVATIONS mois de Janués			METEOROLOGIQUES					rumme de straquithestaille rum au Jaudin Belinique d'Ester marcane J. Denie		
	A DAROMETRE			THERM S.A.		PSYCHROMETRE		ACTINOMETRE GIROUETTE NUI					1025 TA	EAU		
	MEME	HADTEUR	NAUTEUR	A PRIMITA	EVAP	SEC	HUMIDE	HUMIDE HUMIDITE				DIRECTION FORCES MED.		FORMES	TIMBE	E .
Maper		ALANDAS DA TANA DA TAN	二日本の「日本の「日本の」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	<b>在一時常常在自己的生活的的是在我们的是我们的是我们的是我们的</b>	E BERERERERERERERERERERERERERERERE	E REEREREEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	在一位在在此在在在在在在在在在在在在在在在在在在在在在在在在在在在在	「 「「「「「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」	在 在在在此中在在在在在在在在在在在在在在在在在在在在在在在在在	E GEERBREEREEREEREEREEREEREEREEREEREEREE	that a so that a	CANDANAN CENTERDE LA LONGONALLAN	ちろうのないないないないないないないないないないない あいろう		1111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Permilland bigen Banes de Alemilland Quelque que te se à se à se de Banes de Secontand Plus de 4 & Te a Ji bernan Plus de 4 & Te a Ji bernan Plus de 4 & Te a Ji bernan Annes de Anni Hand Maine de Secontand Maine de Anni Hand Maine de Anni Hand Maine de Anni Hand Maine de Anni Hand Maine de Anni Hand
	T 100 113		r. A.	Landa .			1 the				TOTAL	LAU TO	S DE P	NE S	(h)	
	EREQUES DAS	CE JANA		NNE NE EI	4 1.15L	1.55K.8	2312.31K W	Surve une				12 13 10		DE BROUILL	480 J.	

# TranskribusX – download from transkribus.eu



#### TranskribusLite - transkribus.eu/r/read/projects/

Tublifo = Koliticis. Hitim Bozen den zweisten Farmer 1.7.9.4. For Title Serrn Bürgermeifter Fofeps v. Demics x<sup>2</sup> x<sub>2</sub> U ab ?! Unclear 🗥 Text Region 1 In Publiko = Politicis. Aktum 🏠 Bozen den 🕒 zweyten Jänner 1794 Vor Titl : etc: Herrn Bürgermeister 🙎 Joseph : v: Remich : kaiserl: : königl: : OÖ<sup>en</sup>= Regierungsrath und : T: L: M: Ready for Review In beyseyn der Herrn Herrn Magistratsräthe " " Oberlieutenant 🙎 Karl : v: Kuglern " Save Changes 2 : D<sup>r</sup> Leopold Spiellmann

#### velehanden.nl

# KWS - Keywordspotting



https://transkribus.eu/r/kansallisarkisto/

# **Core Platform Features**

- Import
- Export of documents: PAGE-XML, ALTO, PDF, TEI, Word etc.
- Clients: TransribusX & TranskribusLite (Web)
- Enriching and Enhancement Transcription & Tagging
- Collection Management
- Layout Analysis & Correction
- HTR Training and Recognition
  - Sharing of models across collections
- Structure Analysis & Tagging
- Search: Fulltext (SOLR) & KWS
- REST API
- Filestore



# Transkribus App-Server

- Custom software to distribute jobs across worker nodes
- Jobs are written into database, worker modules process jobs according to simple scheduling configuration
- Hardware:
  - Blade-Center, 16 nodes, 12 cores each, 48GB RAM
    - For jobs with high CPU load, i.e. Upload, Export, HTR-Recognition
  - 2 GPU server
    - 24 CPU cores, 188 GB RAM & 8 graphicsprocessing units (GPU) each
    - Used for GPU heavy tasks like HTR-Training
  - LEO4 SGE computing cluster at the University of Innsbruck
    - 48 nodes with 28 Intel Xeon cores, 64GB RAM for each node
    - Used for HTR of large datasets about 100.000 pages in 24h

## Usage Statistics - Registrations



— Nr. of Registered Users

## Usage Statistics – User Logins



----- Nr. of Unique User Logins per Day

## Usage Statistics – Image Processing



## Business Plan Overview – READ-COOP

# Cooperatives as a tool

A governance model can be seen as a tool to reach a given objective

- Associations:
  - Realize common goal
  - usually no direct business is involved
  - not-for-profit is the usual model
- *Limited liability companies*:
  - Highest profit for the shareholders of the company

#### Cooperatives

- mixture of the above models
- common goal is reached by doing for-profit business with the member mainly
- $\rightarrow$  Profit is directly for stakeholders, not for shareholders

# Main advantages

- Motivation
  - Customers become members, members become customers
  - Simple inclusion of members → strong appearance of the coop with dozens of member institutions
- Democratic constitution
  - Decisions are taken democratically
- Transparency
  - Business with a cooperative is transparent
  - E.g.: library pays for text recognition in Transkribus → used to sustain and improve Transkribus (generating new services, etc.) → all figures on the table
- Return of investment
  - Business can be organised among members: If 20 members are interested in a specific feature e.g. to develop a specific Search-Interface – they may invest just 10,000 EUR each and will be able to get something which is worth 200,000 EUR
  - In this way also procurement rules will not prevent the investment (which would otherwise be the case)

# Current Status of the READ-COOP

#### • 2018

- Informal agreements with National Archive Finland, State Archive Zuerich and National Archive Netherlands – nearly 500,000 EUR foreseen for READ-COOP SCE
- Formal inclusion into the DFG project "Court records" from the University of Greifswald
- 2019
  - 01-07: Founding of the READ-COOP
  - 01-09: Several requests such as from Swiss libraries, Trinity College Dublin, German libraries, German archives, Austrian archives, Dutch archives, scanning companies, translation companies, etc...
  - Currently procurement processes are prepared for above members/customers
  - Several concrete business requests

# Outlook

# Outlook – Business

- Implement Business Model into platform
- Recognition of text in the Transkribus platform
  - The more pages the more cost-effective, above 500,000 pages  $\rightarrow$  negotiations
  - The longer someone binds himself, the more cost-effective will it be (subscription model)
  - COOP members get a reduction of 25%
  - ~13-24 cent per handwritten page, ~5 cents for printed, depending on above features, no basic fee

→ So the most cost-effective way is to conclude subscriptions as a member institution on large quantities of documents

- Other services are currently on basis of negotiations
  - Support in ground truth generation
  - Setting up web-interfaces for searching
  - Creating probability indexes (inclusion of UPVLC)
  - Creating specific workflows for table recognition (Naverlabs, CVL, UPVLC,...)
  - Structure Analysis
  - ...

# Outlook – Technical & Project based

- Currently: server and all computing resources located at UIBK
- As the platform is growing  $\rightarrow$  distribute!
- Benefit from collaborations with other Universities and their computing power & storing capacities
- IIIF (International Image Interoperability Framework)
  - Data is stored on different servers but made available through common interface
- $\rightarrow$  "READxd" apply for H2020 project for large scale AI
  - tender: "Big Data technologies and extreme-scale analytic", ICT-51-2020
  - Goal: make large archive collections searchable + enrich and link information (named entity tagging & linking → "Facebook of the past")



transkribus.eu

read.transkribus.eu/coop