

Transkribus in Practice

# Table Processing in Transkribus

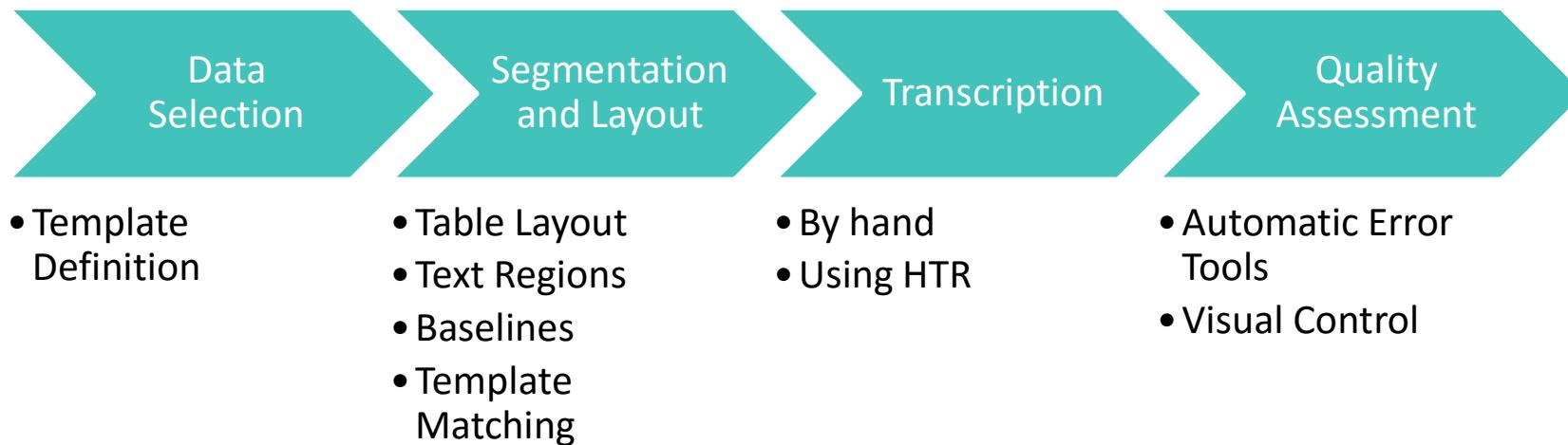


The church registry books present information in table format which is highly intuitive. Automatic recognition must be easy.

”

Common user expectation

# Standard workflow





## ABP death records 1847 – 1878

- Standardized forms with similar content
- More and more standardized writing
- Roughly 700 different hands
- 26,579 scanned pages

# A) Using text regions

18		19								
Name der Vorsteherin	Name in Religion	Lebensgesch. Beschäftigung Name der Gruppe	Leb. oder verstorben	Krankheit, Art. Jahrmutter	Tag und Name der Geburtsfeier	Tag und Ort der Taufe	Alter	Vorname von Mutter	Vorname von Mutter	Erziehung von Mutter
Anna Stift	Elleg. der ledigen Kirche Maria Stift	Schulmeister Herrn	Kind	Brüderlichkeit auf Arzt.	11. Juli 8 d. Margareta	10. Juli	-	-	-	gefürstete Hausgruppe
Rufus Kron	Konkubine	hof	ledig	Weizensäuerling Dr. Brüderlichkeit	15. Juli 5 d. Pauli	17. Juli	01	-	Hautinger, Arzt	Georg gefürstete Hausgruppe
Raffina Kloßberger	gefürstete Spieß Herrn Kloßberger	Schulmeister Herrn	Kind	Brüderlichkeit auf Arzt.	21. Juli 5 d. Margareta	22. Juli	13 Margareta	-	-	Hautinger, Arzt
Evangelika Küller	gefürstete Mutter Herrn Küller	Oppermann Herrn	Kind	Spieß, auf Arzt.	22. Juli 8 d. Hermann	23. Juli	5 August	-	-	Geißel.
Maria Knapf	Jesuforeskates	Schulmeister Herrn	ledig	Caries, Dr. Leopold Knapf	25. Juli 2 d. Pauli	27. Juli	27	-	Hautinger, Arzt	Spießknecht
Karin Lang	oft Lederhosen	Arzt	Kind	Spieß, auf Arzt.	26. Juli 5 d. Pauli	28. Juli	09	-	-	Geißel
Lorenz	Elleg. der ledigen Kirche	Arzt	Kind	Hautärztin	2. September	4. September	06	-	-	Hautärztin

## B) Using table cells

12	Name der Bekleidungen	Name der Religion	Leinwand ausgeschaut Rückseit der Kleidung	Länge der ausgeschaut	Kaufpreis. Art. Gebraucht.
10.	Blauer Mantel mit weißem Kragen	illegal bei Leinen drapiert weiss! Weißer Mantel	Leinenzeug	Kind	Leinenstoff auf Art.
11.	Antique Krone	Leinenstoff	hof	lebig	Wollgewebe Dr. Seufzler
12.	Röffchen Kleidung	oft weiß Kind bei Spindel Johann Kleider	Leinenzeug	Kind	Wollstoffstück auf Art.
13.	Scampische Kleider	oft weiß Kind bei Wollkleid Johann Kleider	Wollstoff	Kind	Scampi; auf Art.
14.	Mosai Kleid	Juwelenkleider	Leinenzeug	lebig	Caries; Dr. Seufzler kleider
15.	Kopfumhang	oft Leinenkleid	Art	Kind	Scampi; auf Art.
16.	Leinen Kleid	illegal bei Leinen drapiert	Art	Kind	Gallbladder;
17.	Leinen	illegal bei Leinen drapiert	Art	Kind	Gallbladder;

12	Tag und Stunde der Zinsabrechnung	Tag und St. der Abrechnung	Alter	Haushalt oder Name?	Signaturen oder Name?
11. Juli 8. u. Morgen	12. Juli	-	-	-	prof. Hartinger Haushalt
13. Juli 8. u. Morgen	13. Juli	31	-	Hartinger, Rong	Georg Seufzler Haushalt
14. Juli 5. u. Morgen	14. Juli	13. August	-	-	Hartinger, Rong
15. Juli 8. u. Morgen	15. Juli	8. August	-	-	Seufzler
16. Juli 2. u. Morgen	16. Juli	27	-	Hartinger, Rong	Seufzler
17. Juli 2. u. Morgen	17. Juli	27	-	Hartinger, Rong	Seufzler
18. Juli 5. u. fünf	18. Juli	9	-	-	Seufzler
19. September	19. September	6	-	-	Hartinger, Rong

# Setting the baselines

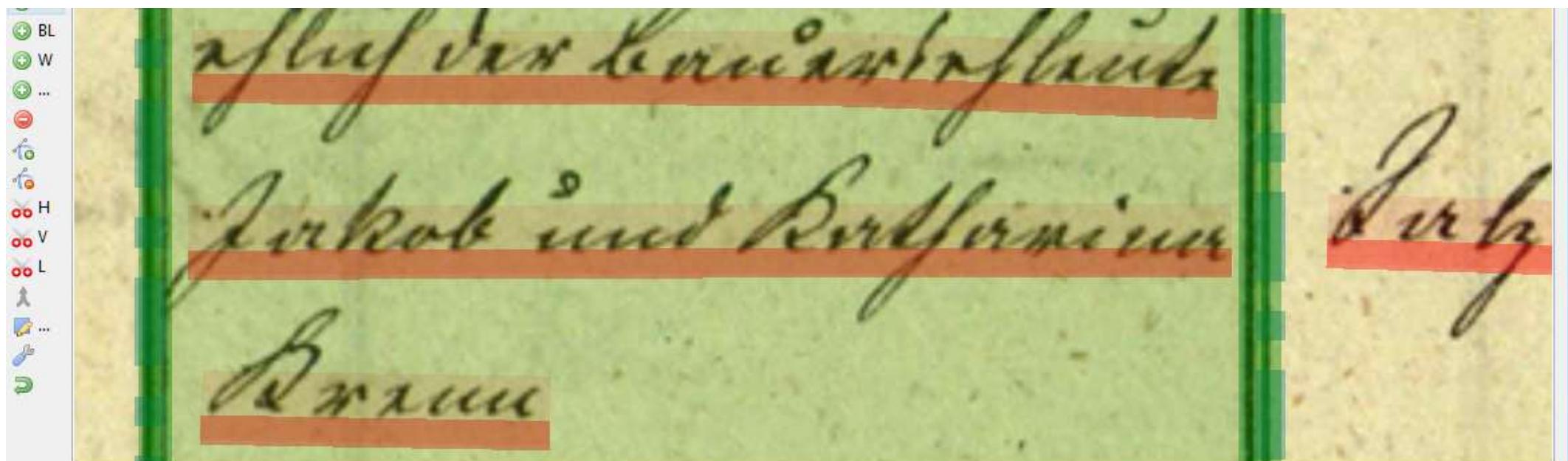
Woch.	Samstag	Sonntag	Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag
18	17	18	19	20	21	22	23	24
18	17	18	19	20	21	22	23	24
19	18	19	20	21	22	23	24	25
20	19	20	21	22	23	24	25	26
21	20	21	22	23	24	25	26	27
22	21	22	23	24	25	26	27	28
23	22	23	24	25	26	27	28	29
24	23	24	25	26	27	28	29	30
25	24	25	26	27	28	29	30	31
26	25	26	27	28	29	30	31	1. September
27	26	27	28	29	30	1. September	2. September	3. September



BISTUM PASSAU  
ARCHIV

**READ**

# Using the text editor



- 1 ehlich der Bauersehleute
- 2 Jakob und Katharina
- 3 Krenn

There are different technical steps to be carried out before we can fully automate the whole process.  
We're on a good way!

”

Response by tech. partners

# Defining table templates

The image displays four historical ledger pages from the Bistum Passau Archiv, illustrating various table templates used for record keeping. The pages are arranged vertically, each showing a different layout of columns and rows.

- Top Left:** A page with two main sections. The left section has columns for "Name", "Vorname", "Geburtsjahr", "Todesjahr", and "Geschlecht". The right section has columns for "Tag", "Monat", "Jahr", "Vorname", "Nachname", "Alter", "Hausnr.", and "Wohnsitz".
- Top Right:** A page with columns for "Name", "Vorname", "Religion", "Geburtsjahr", "Todesjahr", "Geschlecht", "Alter", "Hausnr.", and "Wohnsitz".
- Middle Left:** A page with columns for "Name", "Vorname", "Religion", "Aufenthaltsort", "Lebig", "Raumzahl", "Ziel", "Ziel", "Name", and "Bemerkung".
- Middle Right:** A page with columns for "Name", "Vorname", "Religion", "Wohnsitz", "Geburtsjahr", "Todesjahr", "Geschlecht", "Alter", "Hausnr.", and "Wohnsitz".
- Bottom Left:** A page with columns for "Name", "Vorname", "Religion", "Wohnsitz", "Geburtsjahr", "Todesjahr", "Geschlecht", "Alter", "Hausnr.", and "Wohnsitz".
- Bottom Right:** A page titled "S t e r b" and "R e g i s t e r". It has columns for "Name", "Vorname", "Religion", "Geburtsjahr", "Todesjahr", "Geschlecht", "Alter", "Hausnr.", and "Wohnsitz". Below this, specific entries are listed: "Wieell Anton", "Lacim Jäppel", "Landau", "Eggendorf", "Reinhard", "Pouwra", "Steine", "18. Febr. 1851", "gegraben", "20. Feb. Addorf", "16. Febr.", "Expositus", "Haferten", and "pro. I".

# Matching the right table template READ

1. Tabelle					2. Tabelle				
Name der Bürgerin	Name der Bürgerin	Angestellte Zeitpunkt Name der Gruppe	Alter der Angestellte	Arbeitsort, Abt. Zimmer	Tag und Monat der Zeitpunkt	Tag und Ort der Zeitpunkt	Alter der Angestellte	Arbeitsort oder Wohnort	Signature oder Name?
20. Februar 1868 Katharina	H. Kuntzelmann Katharina	Katharina	30	Schiffbauamt Dr. Schmidleiter	20. September 5. d. Kreis	20. September 5. d. Kreis	30	-	Passauer Reg.
21. April 1868 Katharina	Katharina	Katharina	30	Wärter offiziell nicht angestellt	25. September 10. d. Kreis	27. September 10. d. Kreis	30	Passauer Reg.	Passauer Reg. Nach dem Abgang der Bürgerin wurden aufgetrennt gelagert
22. April 1868 Katharina	Katharina	Katharina	30	Morbilli; Dr. Schmidleiter	1. Oktober 10. d. Kreis	2. Oktober 10. d. Kreis	-	-	Passauer Reg.
23. April 1868 Katharina	Katharina	Katharina	30	Krebs; Dr. Schmidleiter	5. Oktober 5. d. Kreis	7. Oktober 5. d. Kreis	1	-	Passauer Reg.
24. April 1868 Katharina	Katharina	Katharina	30	Wohl nicht angestellt	7. Oktober 5. d. Kreis	9. Oktober 5. d. Kreis	-	-	Passauer Reg.
25. April 1868 Katharina	Katharina	Katharina	30	Passauer Reg.	25. Oktober 5. d. Kreis	27. Oktober 5. d. Kreis	1	-	Passauer Reg.
26. April 1868 Katharina	Katharina	Katharina	30	Passauer Reg.	27. Oktober 5. d. Kreis	29. Oktober 5. d. Kreis	1	-	Passauer Reg.
27. April 1868 Katharina	Katharina	Katharina	30	Passauer Reg.	29. Oktober 5. d. Kreis	31. Oktober 5. d. Kreis	1	-	Passauer Reg.



# Using line finding tools

# Detecting the rows

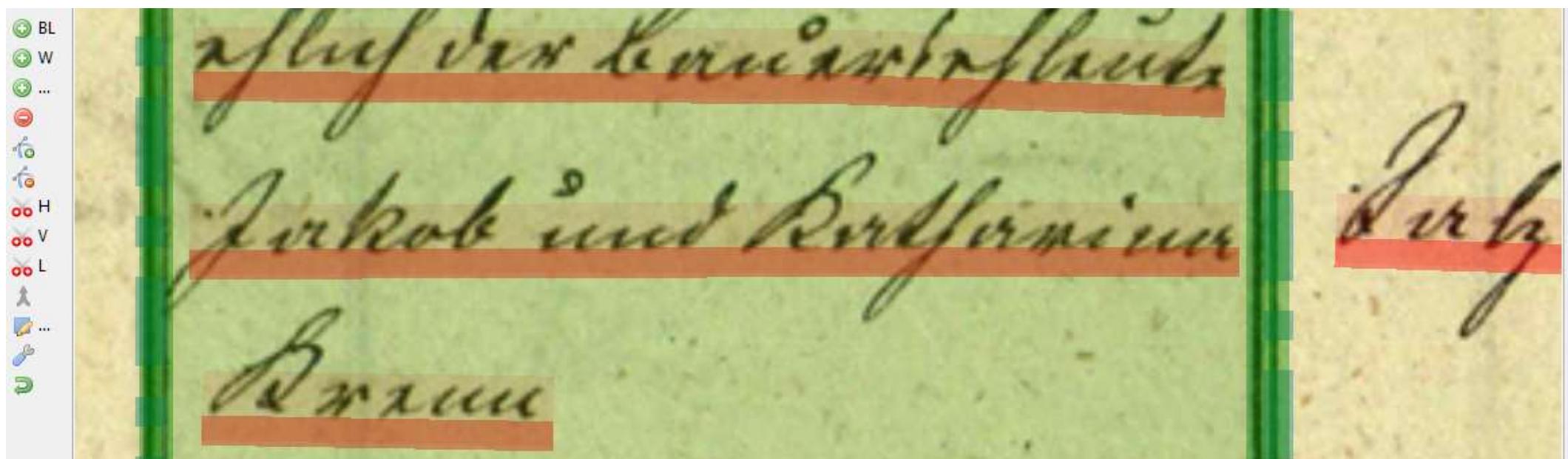
125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145
Plauer	Raxat	Lungenentz.	Lung	Knorpelk. Aug.	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
der Brockenau	mit	abgängig	abwärts	abgewandt	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141
Theresa	stark leidend	klarfüßig	klar	krampfhaft	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142
Maurer	leidenschaftl. Mutter				128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
Maria	juv. fuscus	frisch	klar	lebhaft	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
Uhlzinger					130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145
Silber	lauer z. d. d.	schwach	verdächtig	leidenschaftlich	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146
Neumüller	krankh. g. g.				132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147
Johann	stark betroffen - stark unstetiges Mutter	abgängig	klar	lebhaft	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148
Uhlzinger	abgängig				134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149
Theresa	Stom. Villain	abfall	unreinig.	Vitium recte. Sanguinea pro. Corp.	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
Bauer					136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151
Matthias	juv. fuscus	abgängig	unreinig.	Tuberculosis Dr. Siegl	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152
Hölle					138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
Joseph	stark betroffen - stark unstetiges Mutter	abfall	klar	Catarrh.	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154



BISTUM PASSAU  
ARCHIV

**READ**

# Running automatic recognition



1 ehlich der Bauersefhlut-

2 Jakob und Katharina

3 Krnnn



# Checking the quality

## Fraenleiten Frauenleiten

ledig  
Wassersucht.  
Ohne Arzt  
7. Februar 5½ Uhr  
Morg. rite provisa. provis.  
9. Februar  
72 Jahre  
derselbe.  
Genoß dīturaute die turnus-  
wesen Vterpflegung weise  
Verpflegung in der  
Gemeinde u. starb bei beim  
Unterhubermünfcfershale  
Unterhuber im Pferdstalle.  
H4) Joseph  
Fürstberger  
ehl. Kind des Math. Matth.  
Fürstberger, Kochseder Raßöder-  
Bauers bauers  
do do  
Langeneck  
ledig  
Keuchhusten;  
ohne Arzt.  
8. März 9 Uhr Vorm.

## Polhamer

illeg. der Ansölhauer; Aña  
Polhamer,  
Vorderschacstnertochter  
Vorderschachtnertochter  
do do  
Schachten  
ledig  
Zehrfieber;  
ohne Arzt.  
19. März 3 Uhr  
Nachts. Nachm.  
21. März  
7.4 Wochen  
derselbe  
v—  
do 6) A. Maria  
Heiß Geiß  
Austräglerin auf  
den Schneidranwehe dem  
Schneideranwesen  
do do  
Juni Zaun  
Wittwe  
Allgemeine Wassersucht  
Wassersucht.

## derselbe

v—  
7) Walburga  
Huber  
ehl. ehl. Kind des Leimer Klein-  
Gütlers gütlers Joseph Huber  
do do  
Grasensee  
ledig  
Zehrfieber, Zehrfieber;  
ohne Arzt.  
23. 22. Juni 9 Uhr  
Vorm Abends  
25. Juni  
1 Jahr v 4 Monate  
derselbe  
v—  
Stunde. 8) Aloys u.  
9) Otto  
Hartmañegruber Hartmañsgruber  
ehl. Zwillinge des  
Schreiners Aloÿs Hat Aloys Hart-  
mantgauber mañsgruber  
do  
Walburgskirchen  
ledig

## 6-8) 9 Tage

M. 9) 11. Tage  
derselbe  
v—  
10) Amalia  
SKainzsberger Kainzlsberger  
St Schiefer ehl. Kind des Müllers  
G Kainzberger Gg Kainzlsberger  
do do  
Haunddor Hunddorf  
ledig  
Convulsionen; Konvulsionen;  
ohne Arzt.  
7. Juli 1 Uhr  
gleich früh  
9. Juli  
1 Jahr 2 Mon.  
derselbe  
v—

## Summary of ABP experiments

- Table Editor is a useful tool for segmentation
- Handling of transcription is experienced to be more user friendly when using text regions
- HTR quality is quite good already and keeps improving
  - when trained on the current scribe
  - when (one or more) proper dictionary is used
  - when more training material is supplied
  - from the current CER of 13-17%

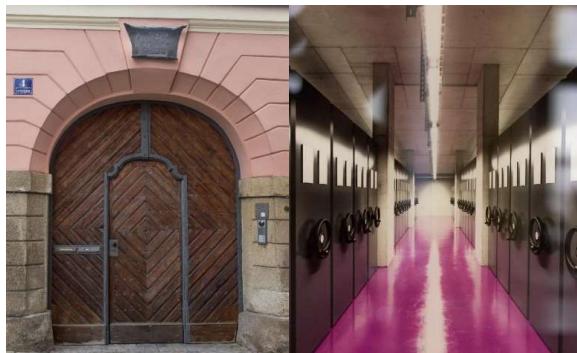


BISTUM PASSAU  
ARCHIV

READ



# Thank you!



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

## Details of presented data

- Source of images / data
  - ❖ ABP, Strasskirchen, S, 012\_0123
  - ❖ ABP, Strasskirchen, S, 012\_0125
  - ❖ ABP, Walburgskirchen, S, 004-01\_0112
- Software used
  - ❖ Transkribus Expert Client
  - ❖ nomacs Image Lounge
  - ❖ Transkribus PyClient

## Special thanks to

- Naverlabs Europe
- Computer Vision Lab, TU Wien
- CITLab, Universität Rostock
- DEA, Universität Innsbruck
- and the other READ partners