

Papers and Collections at the Department of Contemporary History

Modelling and Description



Agenda

- Introduction
- Transkribus - Layout and Text Recognition
- RNAB - Data Description
- Goobi, Phaidra - Publication and Storage
- Machine Learning and Data Reconciliation

Introduction: Our Project

- Department of Contemporary History
- 2023/2024
- personal papers, correspondences, newspaper articles
 - description
 - storage
 - processing
 - publication

“Gaupresse”- Archiv at the Contemporary History Library



PROJEKT- BESCHREIBUNG

Das sogenannte „Gaupresse“-Archiv Wien (Archiv des Gaupresse-Amtes der NSDAP-Gauleitung Wien) ist im wesentlichen eine Zeitungsausschnittsammlung der NSDAP-Gauleitung Wien, die in über 600 Kartons als Sondersammlung der Fachbereichsbibliothek

Indexes of the “Archive” at the Department of Contemporary History



Archiv - Nr.: Do-19/Nl-7

Tit.: Sammlung Otto Molden

Umfang: 19 Mappen (3 Kartons)

Zeitraum: 1938-45

Inhalt: Materialsammlung Otto Molden zur Geschichte der österr. Widerstandsbewegung gg. Hitler (O.Molden, Ruf d. Gewissens).

Inventarmappe (unvollständig, v. O.Molden angelegt).

Mappe:	Bericht	Autor (od. Bez. d. Nr.):	Hauptaktionsgebiete:	Tätigkeitsdauer:
I	1	Molden, Fritz	Wien, Tirol, Salzburg, Schweiz, Südtirol, Italien, Frankreich	1938-45
	2	Molden, Otto	Wien, Tirol, Salzburg, Schweiz, Südtirol	1938-45

Written Estate of Hugo Portisch



©ORF

Transkribus

- Optical Character Recognition (OCR), layout analysis and structure recognition of historical documents
- Based on Artificial Intelligence
- For the “Gaupresse”- Archiv our objective is keyword search
- Optical Character Recognition (OCR) is largely successful within the “Gaupresse”- Archiv

GPA collection, Transkribus OCR and Kraken: Transcription, heading, date and newspaper reference are searchable

1 "Völkischer Beobachter", Wien, 4. April 1939.

Die neue Brüsseler Kammer

Zur innerpolitischen Lage Belgiens

Von Ernst Veit

Die internationalen Spannungen der letzten Wochen haben die Vorgänge in der belgischen Innenpolitik, denen sonst sicher mehr Aufmerksamkeit gebührt hätte, in den Hintergrund gedrängt. Am Sonntag fand nun in Belgien die Wahl zu einem neuen Parlament statt, das nach dem Wunsch und Willen des Königs „Die Lösung des wesentlichen Problems bringen soll: inneres Gleichgewicht, einträchtige Zusammenarbeit im gegenseitigen Verständnis und gegenseitige Loyalität im Dienste der nationalen Einheit“. Die Auflösung der alten Kammer war überraschend gekommen, besonders da sie tatsächlich unter den Bedingungen erfolgte, daß der König auch seine beauftragten Minister mit größter Anstrengung verhindern wollten: der flämischen Frage.

und Parteien wochenlang mit mehr oder weniger Verantwortung mit der Regierungsbildung nicht fertig wurden, war hinter der parlamentarischen Systemkrise der wahre Spannungsherd immer sichtbar geworden, das alte Grundproblem Belgiens: der Gegensatz zwischen den Flamen und dem belgischen Staat.

Die Krise wurde durch einen an sich scheinbar glänzenden Vorfall ausgelöst. Der flämische Professor M a t t e e n s war bei Eröffnung der flämischen Akademie der Wissenschaften zum Mitglied der neugegründeten flämischen Akademie für Medizin ernannt worden. Der damalige Ministerpräsident Spaak hatte diese Ernennung bestätigt, vielleicht in dem Glauben, daß diese Geste — sicher war ihm bekannt, daß Professor M a t t e e n s während des Weltkrieges dem nach staats-

Region 1

1 "Völkischer Beobachter" Wien, 4. April 1939.

Region 2

Region 3

1 Die neue Brüsseler Kammer
2 Zur innerpolitischen Lage Belgiens
3 Von Ernst Veit

Region 4

1 Die internationalen Spannungen der letzten
2 Wochen haben die Vorgänge in der belgischen
3 Innenpolitik, denen sonst sicher mehr
4 Aufmerksamkeit gebührt hätte, in der
5 Hintergrund gedrängt. Am Sonntag fand
6 nun in Belgien die Wahl zu einem neuen
7 Parlament statt, das nach dem Wunsch und
8 Willen des Königs „Die Lösung des
9 wesentlichen Problems bringen
10 soll: inneres Gleichgewicht, einträchtige Zusammenarbeit
11 im gegenseitigen Verständnis
12 und gegenseitige Loyalität im Dienste der
13 nationalen Einheit“. Die Auflösung der
14 alten Kammer war überraschend gekommen
15 besonders da sie tatsächlich unter dem
16 Lösungswort erfolgte, daß der König und
17 auch seine beauftragten Minister mit größter

GPA collection, Transkribus OCR and Kraken: Transcription is searchable, but the heading, date and newspaper reference need to be revised



Das bevorstehende Verfahren gegen Schuschalgg und seine Mitarbeiter.

Berlin, 22. (DNB.) Zu dem von der Reichsregierung erlassenen Gesetz über die Verantwortlichkeit von Mitgliedern ehemaliger österreichischer Bundes- und Landesregierungen und ihrer Mitarbeiter schreibt der „Deutsche Diens“: Das Gesetz hat die Möglichkeit, System-Größen des völkseindlichen Schuschalgg-Regiments zur Rechenschaft zu ziehen und schwere Verstöße sehr maß führender Persönlichkeiten gegen Recht und Verfassung zu ahnden. Diese Maßnahme stellt eine billige Rache an einstigen politischen Widersachern dar, die ohnehin keine Gelegenheit mehr finden werden, ihre gefährlichen Machenschaften fortzusetzen. Sie erfolgt im Namen des Volkes das Jahre hindurch gequält und durch Gewaltmethoden, die oft genug den Tatbestand eines

Auf diese Weise seien in Schuschalgg-Österreich viele tausende Personen, die irgendwelchen einflussreichen Männern verhasst waren, nicht nur wirtschaftlich, moralisch und physisch ruiniert, sondern zum Teil auch lebenslanglich in die Gefängnisse gesperrt worden. Zumal selbst Angehörige der Bundesregierung nach dem vorgefundener Altenmaterial und nach eigenen Aufzeichnungen Recht und Verfassung gebrochen haben, um persönlichen Nachgelust an ihnen mißliebigen Persönlichkeiten freien Lauf zu lassen, war in Österreich eine Rechtsunsicherheit entstanden, die wohl in keinem anderen Staat der Welt eine Parallele hat.

Es sei sichergestellt worden, daß das neue Gesetz mit äußerster Vorsicht angewandt



region 1	region 2	region 3	region 4	region 5
1 „OBSERVER“				
2 I. österr. behördl. konzessioniertes				
3 Unternehmen für Zeitungs-Ausschnitte				
4 WIEN, I., WOLLZEILE 11				
5 TELEPHON R 23-0-43				
	Region 2			
1 vom ;				
	Region 3			
1 Pni				
2				
3	Ausschnitt ausbrannel lagosbete, Erunb			
4 25. Aug. 1938				
	Region 4			
1 Jas beborstrheade Berfahren ge				
2 gen				
3 seine Mitarbeiter.				
4 Schuschalgg un				
	Region 5			
1 Berlin, 22. (DNB.) Zu dem von der Reichs-				
2 regierung erlassenen Gesetz über die Verantwort-				
3 lichkeit von Mitgliedern ehemaliger österreichi-				
4 scher Bundes= und Landesregierungen und ihrer				
5 Mitarbeiter schreibt der „Deutsche Diens				
6 „Das Gesetz hat die Möglichkeit, System=Größen				

Layout Recognition

- Quality of text recognition is depending on layout recognition
- Testing of models “Block Detection” and “Kraken”:
no significant searchability difference
- “Kraken” slightly superior regarding transcription

Transkribus - Metadata

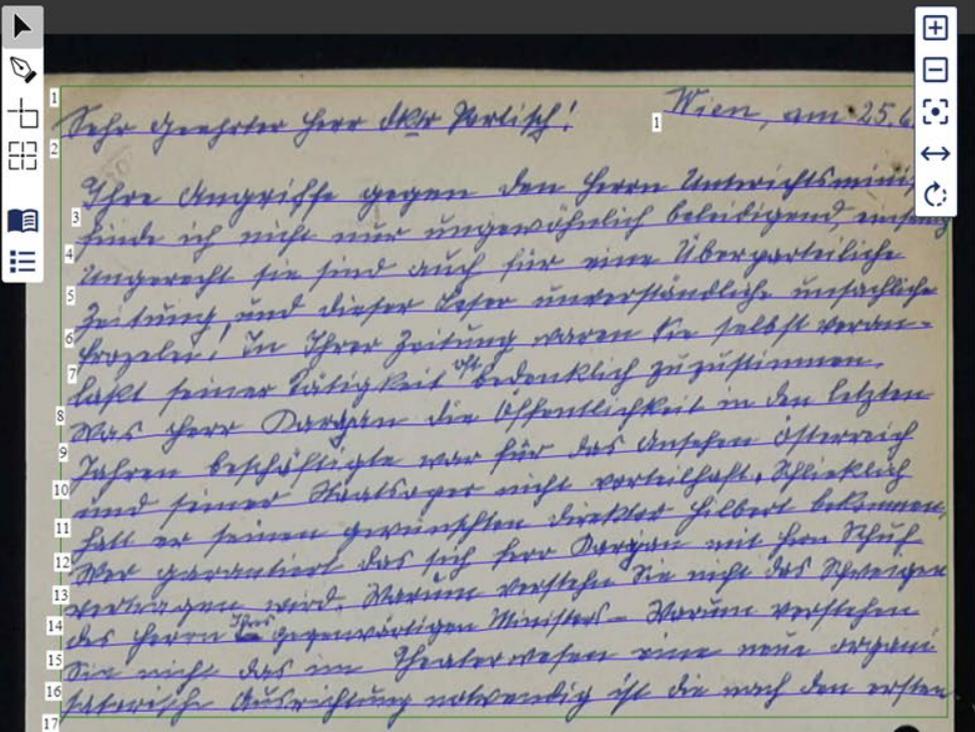
- Metadata extraction poses challenges
- Quality is insufficient for reliable recognition
- Solutions could be:
 - Assigning coordinates
 - Machine learning

Transkribus - Hugo Portisch

- 24,000 pages of typewritten documents and correspondence
- 3,000 handwritten documents and correspondence by different authors are difficult for Handwritten Text Recognition (HTR)
- Text Titan I (GPT-Model) and German Kurrent XIX-XX M6-2 were tested
- Analysis according to visual similarity



Hugo Portisch, Transkribus Layout Analysis, Text Titan I: Transcription is sufficient for keyword search



Wien, am 25.6.69

Sehr geehrter Herr dkg Portisch!

Ihre Angriffe gegen den Herrn Unterrichtsminister finde ich nicht nur ungewöhnlich beleidigend, einseitig

Ungerecht sie sind auch für eine Überparteiliche Zeitung, und dieser Leser unverständliche unsachliche

frozelei. In Ihrer Zeitung waren Sie selbst veranlaßt seiner Kätigkeit oft bedenklich zuzustimmen.

Was Herr Karajan die Öffentlichkeit in den letzten Jahren beschäftigte war für das Ansehen Österreich

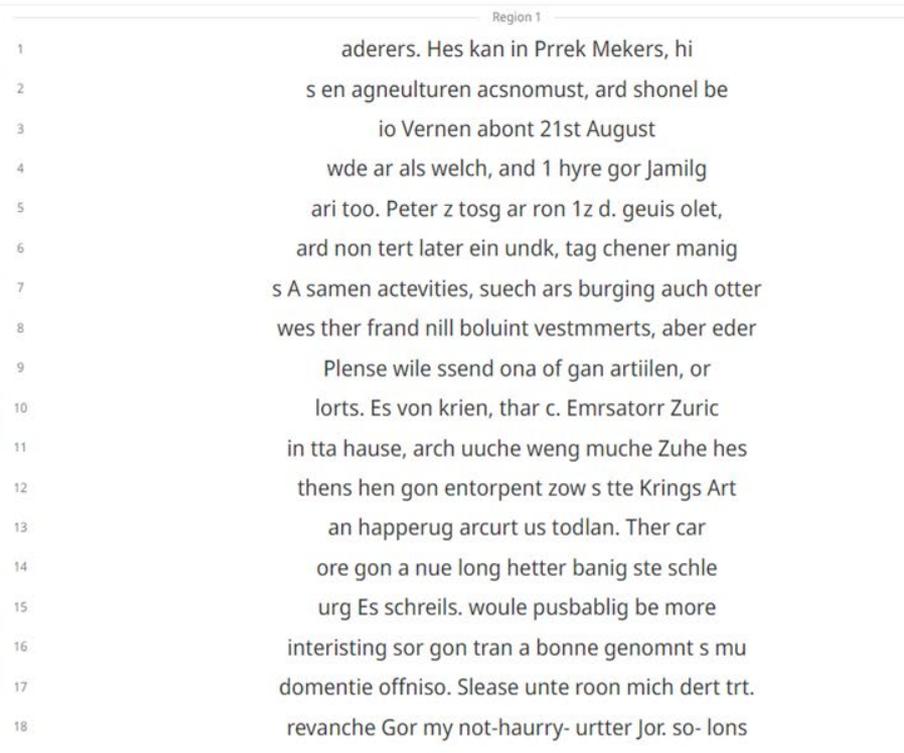
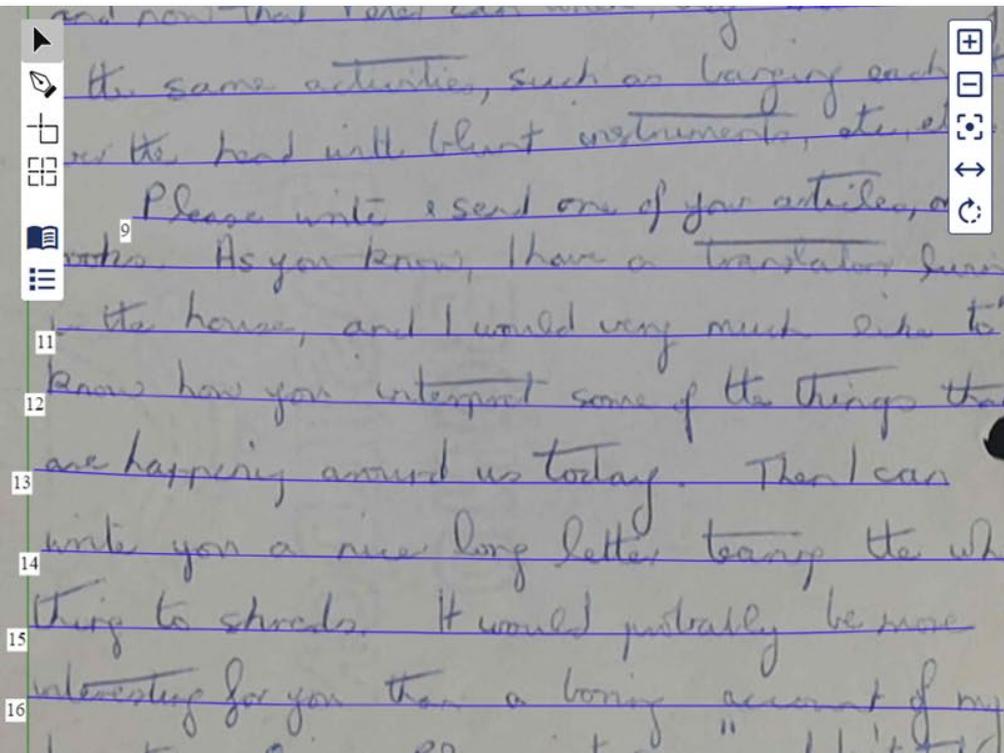
und seiner Staatsoper nicht vorteilhaft. Schlieklich hatt er seinen gewünschten Direktor Hilbert bekommen,

Wer garantiert das sich Herr Kargau mit Herrn Schuh vertragen wird. Warum verstehen Sie nicht das Schweigen

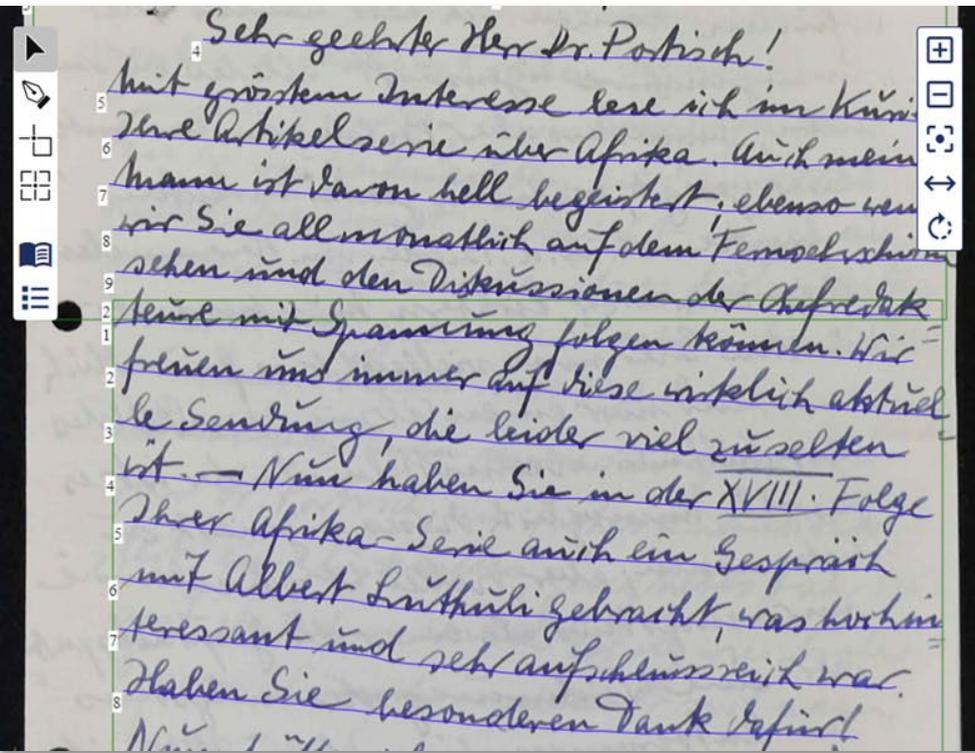
des Herrn Herr gegenwärtigen Ministers - Worum verstehen

Sie nicht das im Theaterwesen eine neue Organisationsatorische Ausrichtung notwendig ist die nach den ersten

Hugo Portisch, Transkribus Layout Analysis, Text Titan I: Transcription is not sufficient since the language could not be identified correctly



Hugo Portisch, Transkribus Layout Analysis, German Kurrent XIX-XX M6-2: Transcription is sufficient for keyword search



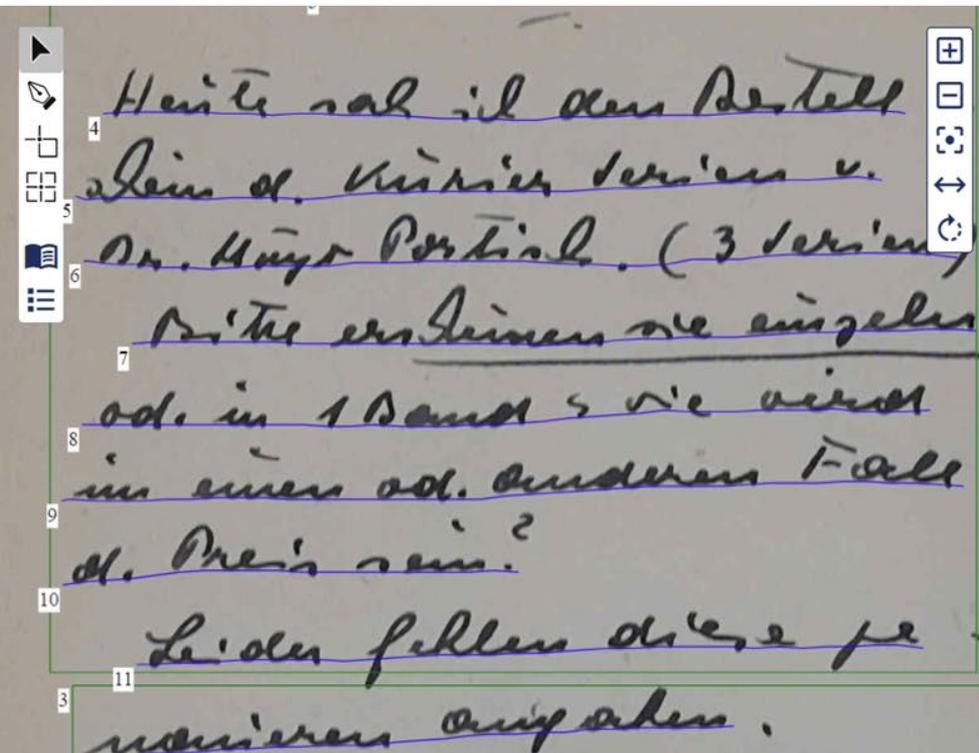
Region 1

1 Maria Lotz
2 Wien den 12.12.62.
3 Wien XIV. Bumgartenstr. 66
4 Sehr geehrter Herr Dr. Portisch
5 mit größtem Interesse lese ich im Kurier
6 Ihre Artikelserie über Afrika. Auch mein
7 Mann ist davon hell begeistert, ebenso wenn
8 wir Sie allmonatlich auf dem Fernsehschirm
9 sehen und den Diskussionen der Chefredak-

Region 2

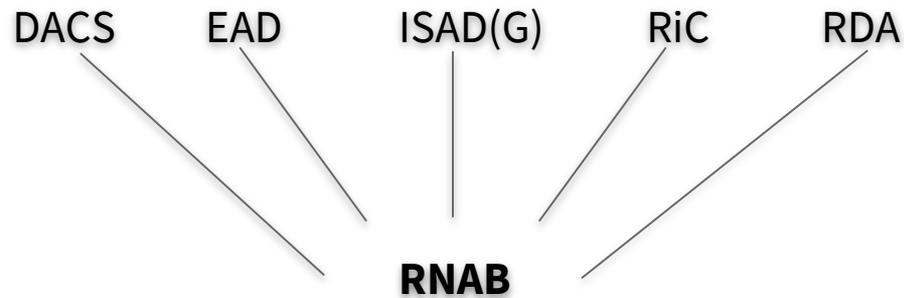
1 teure mit Spannung folgen können. Wir
2 freuen uns immer auf diese wirklich aktuel-
3 le Sendung, die leider viel zu selten
4 ist. Nun haben Sie in der XVIII Folge
5 Ihrer Afrika- Serie auch ein Gespräch
6 mit Albert Luthuli gebracht, was hochin-
7 teressant und sehr aufschlussreich war
8 Haben Sie besonderen Dank dafür!
9 Nun hätte ich persönlich eine sehr grosse

Hugo Portisch, Transkribus Layout Analysis, German Kurrent XIX-XX M6-2: Transcription is not sufficient



	Region 1
1	möder, 8/
2	68
	Region 2
1	dess a.
2	Redaktion d. Kurien
3	bei ein dr
4	Heule sah ich aus Bertels
5	Dein d. Risnies terien v.
6	dr. kagr Portial. (3 darien
7	Bite ersheinen wie einsales
8	od. in 1 Bandt we venat
9	in einen od. aanderen Kals
10	al. Preis sein.
11	Leides fallen diese je
	Region 3
1	manieren auch aben
2	Sebala il darüker
3	anderlies s bei verdun
4	an vielleicht amh bestellen

RNAB - Resource Cataloguing with Authority Data in Archives and Libraries



RNAB

	A	B	E	F	G	H	I	J	K
1	D1 Aufbewahrende Institution	D1 Signatur	D2 Stufe der Verzei	D3 Angaben	D4 Titel	D5 Akteur:in 1	Beziehungske	Akteur:in 2	Beziehungskennzeichnung
2	Archiv ÖGZ/IZG (IfZ)/Archiv des HP		Bestand	Der Nachlass	Nachlass Hug	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
3	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_01		Teil eines Bestands	Der Nachlass	Manuskripte	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
4	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_02a		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
5	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_02b		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
6	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_03		Teil eines Bestands	Der Nachlass	Dokumente "	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
7	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_04		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
8	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_05		Teil eines Bestands	Der Nachlass	Kurier-Intern	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
9	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_06		Teil eines Bestands	Der Nachlass	Unterlagen zu	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
10	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_07		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
11	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_08		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
12	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_09		Teil eines Bestands	Der Nachlass	Unterlagen zu	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
13	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_010		Teil eines Bestands	Der Nachlass	Leser:innenb	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
14	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_011		Teil eines Bestands	Der Nachlass	Reiseplanung	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
15	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_012		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
16	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_013		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner
17	Archiv ÖGZ/IZG (IfZ)/Archiv des HP_014		Teil eines Bestands	Der Nachlass	Korresponde	Universität Wien [GND: ht	Eigentümer:ii	Portisch, Hugo [GND: I	Bestandsbildner

Corporate body (kiz)

Universität Wien. Institut für Zeitgeschichte

03.06.1966-; Wien



EAD-Storage - Archiv_IFZG_Test.xml

- Archiv IFZG
 - Nachlass X
 - Nachlass Fritz Flor
 - Document

Download as EAD file Cancel

Archiv IFZG

Goobi process: - No Goobi process assigned so far - Create Goobi process

Node type: [Icons]

Identity Statement Area:

- Agency code Record ID Unit ID Archive Number
- Reference code(s) D-1 DO Number Title Date(s)
- Structured Date of the Unit Level of description
- Extent and medium of the unit of description (quantity, bulk, or size)
- Standardized extent and medium of the unit of description (quantity, bulk, or size)

Context Area:

- Name of creator(s) Role Person
- Administrative | Biographical history Archival history
- Immediate source of acquisition or transfer

Content and Structure Area:

- Scope and content
- Appraisal, destruction and scheduling information Accruals
- System of arrangement

Condition of Access and Use Area:

- Conditions governing access
- Conditions governing reproduction
- Language | Scripts of material Font
- Physical characteristics and technical requirements
- Finding aids

Allied Materials Area:

- Existence and location of originals Existence and location of copies
- Related units of description Publication note D-21 Edition note

Note Area:

- Note Odd notes

Description Control Area:

- Rules or Conventions Editor

Save and exit record group Save

RNAB - EAD - Goobi

A	B	C	D
1 RNAB	Anforderungen für technische Implementierung	EAD	Goobi (Interpretation der
2	2 Teile („Signatur“ und „Aufbewahrende Institution“)	„Signatur“ <unitid>	Unit ID
3 D-1 Signatur und Institutionsnachweis	Textfelder		
4	Institution aus GND auswählen	„Aufbewahrende Institution“ <repository>	Agency Code
5 DO-Nummer	Textfeld (z.B.: „DO 183“, gibt die Kartonnummer an)	<ead:unitid type="Vorl. Nr.">	Archiv Nummer
6 Standort	internes Vokabular	<originalsloc>	Aufbewahrungsort der Originale
7 D-2 Stufe der Verzeichnung	kontrolliertes Vokabular	<ead:c level="[Stufe]>	Verzeichnungsstufe
8	bildet Hierarchie ab (Bestand, Teil des Bestandes, Konvolut, Einzelressource)		
9 D-3 Angaben zur Erschließung	Textfeld	<arrangement>	
10	[x]	<processinfo>	Ordnung und Klassifikation
11 D-4 Titel	Textfeld	<unittitle>	Titel
12	1. Teil: „Akteur:in“		
13	GND-Suche oder Auswahl aus den definierten Entitäten (Personenverzeichnis)	<persname role="[Beziehungskennzeichnung]">	
14	GND-Metadaten sollen mit-indexiert werden	<famname role="[Beziehungskennzeichnung]">	Person; Rolle
15 D-5 Akteur:in (Person, Familie, Körperschaft)	2. Teil: „Beziehungskennzeichnung / Rolle“	<corpname role="[Beziehungskennzeichnung]">	
16	kontrolliertes Vokabular der RNAB		
17	GND-Suche oder Auswahl aus den definierten Entitäten		
18	existierende codierte Listen? (z.B. MARC Code List for Relators)		
19 D-6 Zeitangabe	2 Teile: ISO 8601 und Textfeld (für angepasst Validation)	<unitdate> <date>	Entstehungszeitraum
20 D-7 Ort	ev. GeoNames Suche oder GND	<geogname>	

Machine Learning and Data Reconciliation

- The quality of the OCR will allow for full text search, but probably not for named entity recognition => topic modelling
- The structural metadata of the GPA (hierarchical call marks) need to be reconciled / unified and connected with the relevant information in the contemporary indices
- For the former, we will use Gensim (a Python library) and display the results on the user interface; for the latter, we will collaborate with colleagues from the Data Science network

Name	Änderungsdatum	Typ	Größe
 K004_3-M001-A001-0	11.12.2017 23:12	JPG-Datei	217 KB
 K004_3-M001-A002	11.12.2017 23:12	Adobe Acrobat-D...	2.426 KB
 K004_3-M001-A002-0	11.12.2017 23:12	JPG-Datei	1.158 KB
 K004_3-M001-A002-1	11.12.2017 23:12	JPG-Datei	1.261 KB
 K004_3-M001-A003-0	11.12.2017 23:12	JPG-Datei	919 KB
 K004_3-M001-A004-0	11.12.2017 23:12	JPG-Datei	1.165 KB
 K004_3-M002-A001-0	11.12.2017 23:12	JPG-Datei	143 KB
 K004_3-M002-A002	11.12.2017 23:12	Adobe Acrobat-D...	4.577 KB
 K004_3-M002-A002-0	11.12.2017 23:12	JPG-Datei	2.698 KB
 K004_3-M002-A002-1	11.12.2017 23:12	JPG-Datei	1.871 KB
 K004_3-M002-A003-0	11.12.2017 23:12	JPG-Datei	1.203 KB
 K004_3-M002-A004-0	11.12.2017 23:13	JPG-Datei	1.528 KB
 K004_3-M002-A005-0	11.12.2017 23:13	JPG-Datei	709 KB
 K004_3-M003-A001-0	11.12.2017 23:13	JPG-Datei	136 KB
 K004_3-M003-A002-0	11.12.2017 23:13	JPG-Datei	1.280 KB
 K004_3-M003-A003-0	11.12.2017 23:13	JPG-Datei	1.962 KB
 K004_3-M003-A004	11.12.2017 23:13	Adobe Acrobat-D...	3.065 KB
 K004_3-M003-A004-0	11.12.2017 23:13	JPG-Datei	1.544 KB
 K004_3-M003-A004-1	11.12.2017 23:13	JPG-Datei	1.513 KB
 K004_3-M003-A005	11.12.2017 23:13	Adobe Acrobat-D...	2.430 KB
 K004_3-M003-A005-0	11.12.2017 23:13	JPG-Datei	1.239 KB
 K004_3-M003-A005-1	11.12.2017 23:13	JPG-Datei	1.184 KB
 K004_3-M004-A001-0	11.12.2017 23:13	JPG-Datei	135 KB
 K004_3-M004-A002-0	11.12.2017 23:13	JPG-Datei	1.228 KB
 K004_3-M005-A001-0	11.12.2017 23:13	JPG-Datei	131 KB
 K004_3-M005-A002-0	11.12.2017 23:13	JPG-Datei	973 KB



Thank you!

emily.herkommer@univie.ac.at

e.hofmann@univie.ac.at

thomas.wallnig@univie.ac.at

Emily Herkommer, Eva Hofmann, Thomas Wallnig