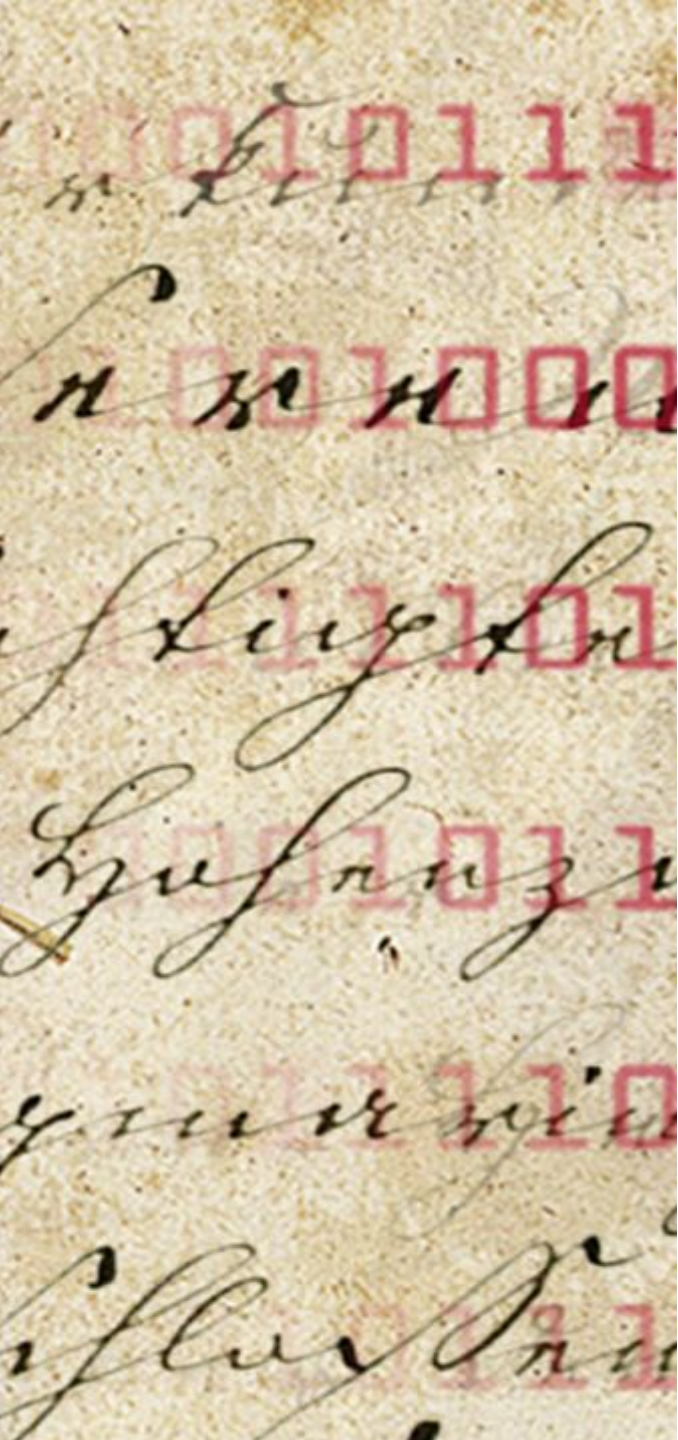


Ein digitaler Werkzeugkasten für historische Forschung mit Archivgut Status quo und Perspektiven

53. Deutscher Historikertag, Fachsektion 6.10.2021
Prof. Dr. Harald Sack (FIZ Karlsruhe), Daniel Fähle (Landesarchiv BW)



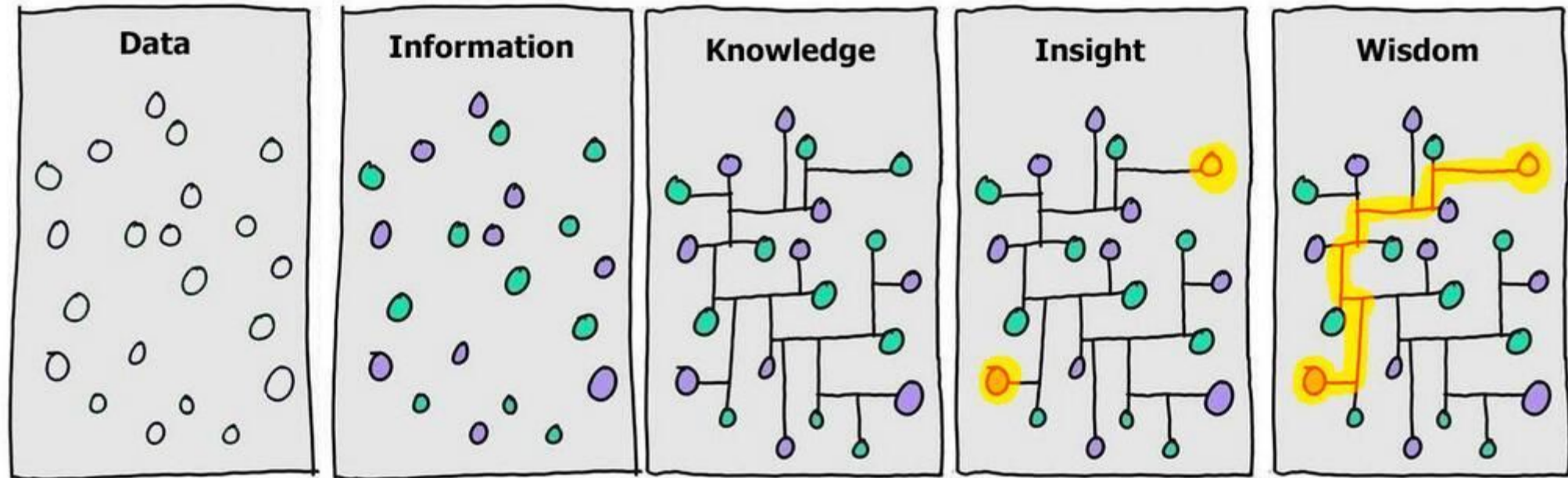
01 Archivportal-D und
4Memory Data Space

02 Datenqualität und
Schnittstellen

03 Anwendungen und
Kompetenzen

04 Wissensextraktion
aus historischen
Forschungsdaten

05 Angewandte
Wissensgraphen



Quelle: Internet Meme, Herkunft: ?

01

Datenangebot: Archivportal-D und 4Memory Data Space



*Historiker*innen brauchen für eine effiziente Forschung aggregierte Datensammlungen, die eine effiziente Suche und einen Zugriff auf Inhalte aus der Ferne erlauben.«*

Jörg Wettlaufer, Welche Services braucht die digitale Geschichtswissenschaft von Bibliotheken, Archiven, Museen und Datenzentren?, 5.10.2021

Archivportal-D

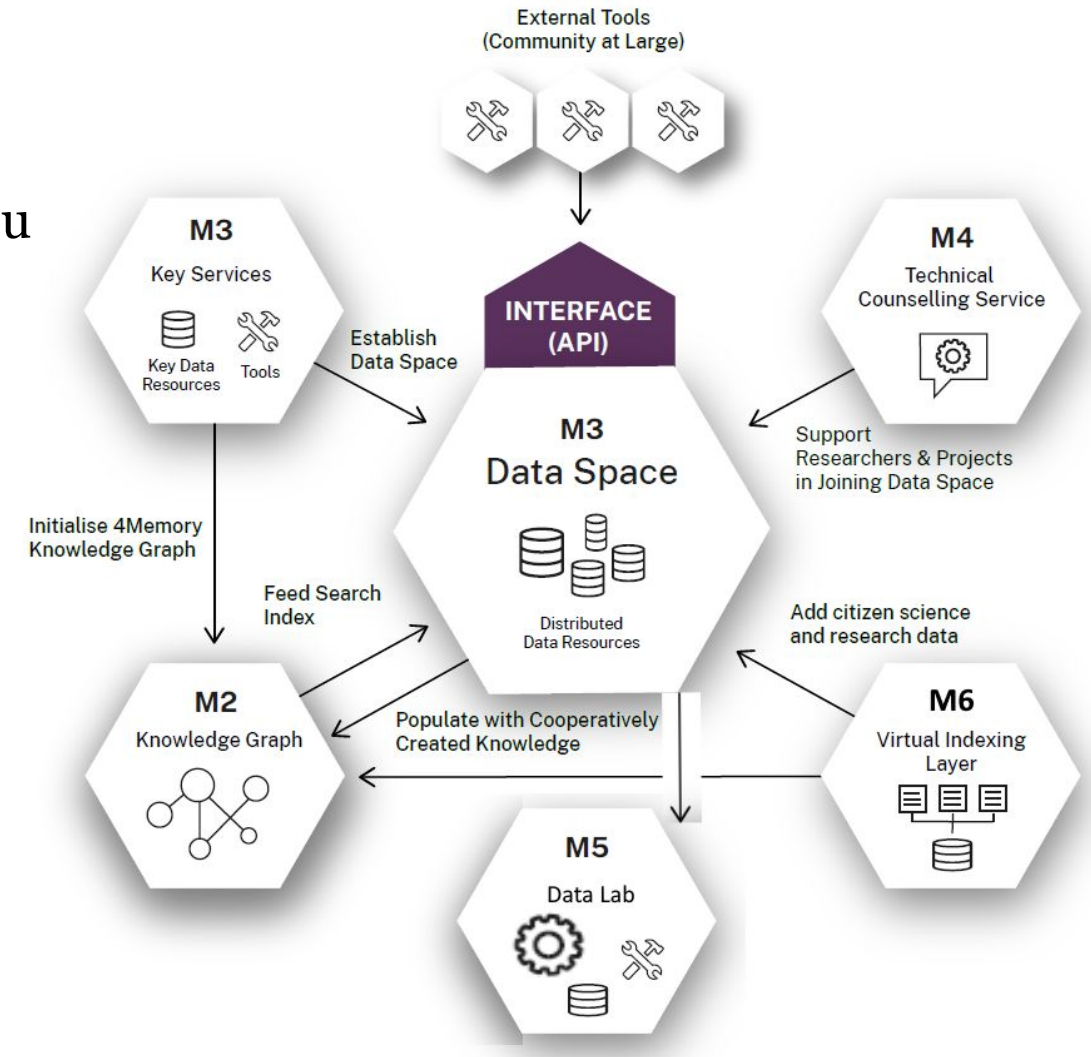
- seit 2014 “Data Hub” der Archive, eng angelehnt an Deutsche Digitale Bibliothek
- Ziel: vollständiges Nachweissystem für Archivgut in Deutschland
- sachthematische Zugänge (Weimar, Wiedergutmachung)
- zentraler Andockpunkt für Services (z.B. Normdatenanreicherung)
- API der DDB nutzbar



Quelle: pixabay

NFDI4Memory - Data Space

- Bestehende relevante Repositorien, DH-Tools und Datenangebote interoperabel und FAIR machen
- Vernetzung bestehender Dienste/Infrastrukturen zu einem übergreifenden »Ökosystem« (Data Space)
- Bedarfsgerechte Optimierung und Erweiterung zentraler Angebote
- Services: übergreifende Suche auf Basis eines Knowledge Graph; kooperative Erschließung
- Data Lab: Erprobung von Anwendungen und Technologien



Erweiterung der Datengrundlage

Grundproblem:

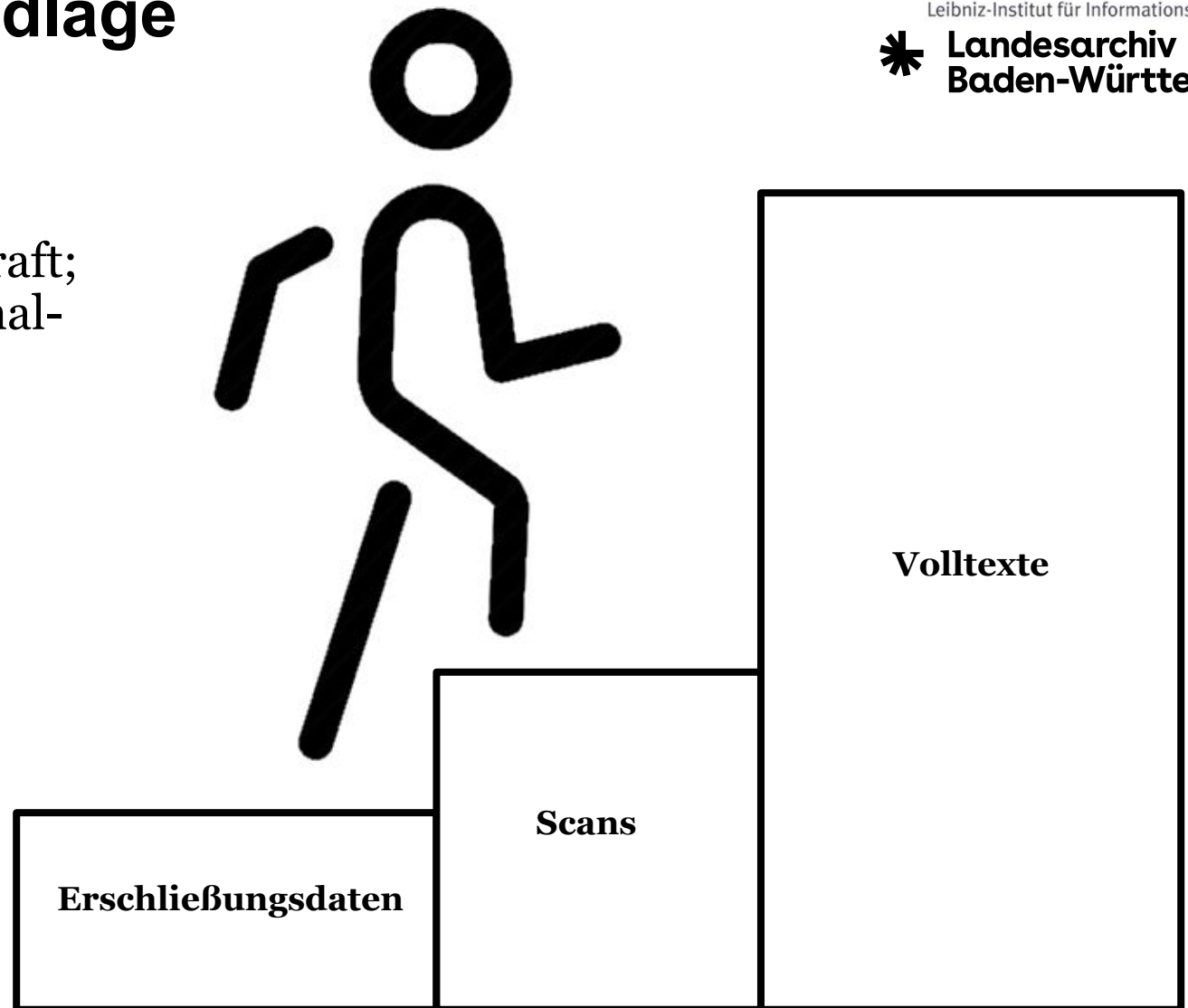
Erschließungsdaten mit wenig Aussagekraft;
flache Erschließung aufgrund von Personal-
und Ressourcenmangel

Abhilfe:

(semi-)automatisierte Verfahren zur
Generierung von Metadaten durch
Datenextraktion und Mining-Verfahren

Quantensprung:

Bereitstellung von Volltexten



02

Bedarfe: Datenqualität und Schnittstellen

FAIR-Prinzipien für Datenqualität

Zentrale Leitplanken

- stabile Identifikation/Zitierfähigkeit
- standardisierte und maschinenlesbare Metadaten
- Verwendung von Normdaten
- rechtssichere Nachnutzung durch Einsatz gängiger Lizenzen
- Bereitstellung von Schnittstellen/API
- Images in ausreichender Qualität



Quelle: pixabay

Datenbereitstellung - Schnittstellen

Vorhanden:

- Bereitstellung von Metadaten und Primärdaten im Download sowie Exporte im EAD(DDB)- und METS/MODS-Format

Bedarfe:

- Datenbereitstellung für individuelle Datensets und Corpora
- RDF-Exporte für semantische Nutzungsszenarien

Lösungsansätze:

- Erweiterung der DDB-API
- Records in Contexts (RiC): neuer Erschließungsstandard unterstützt semantische Datenmodellierung (Umsetzung im DFG-Projekt EEZU)

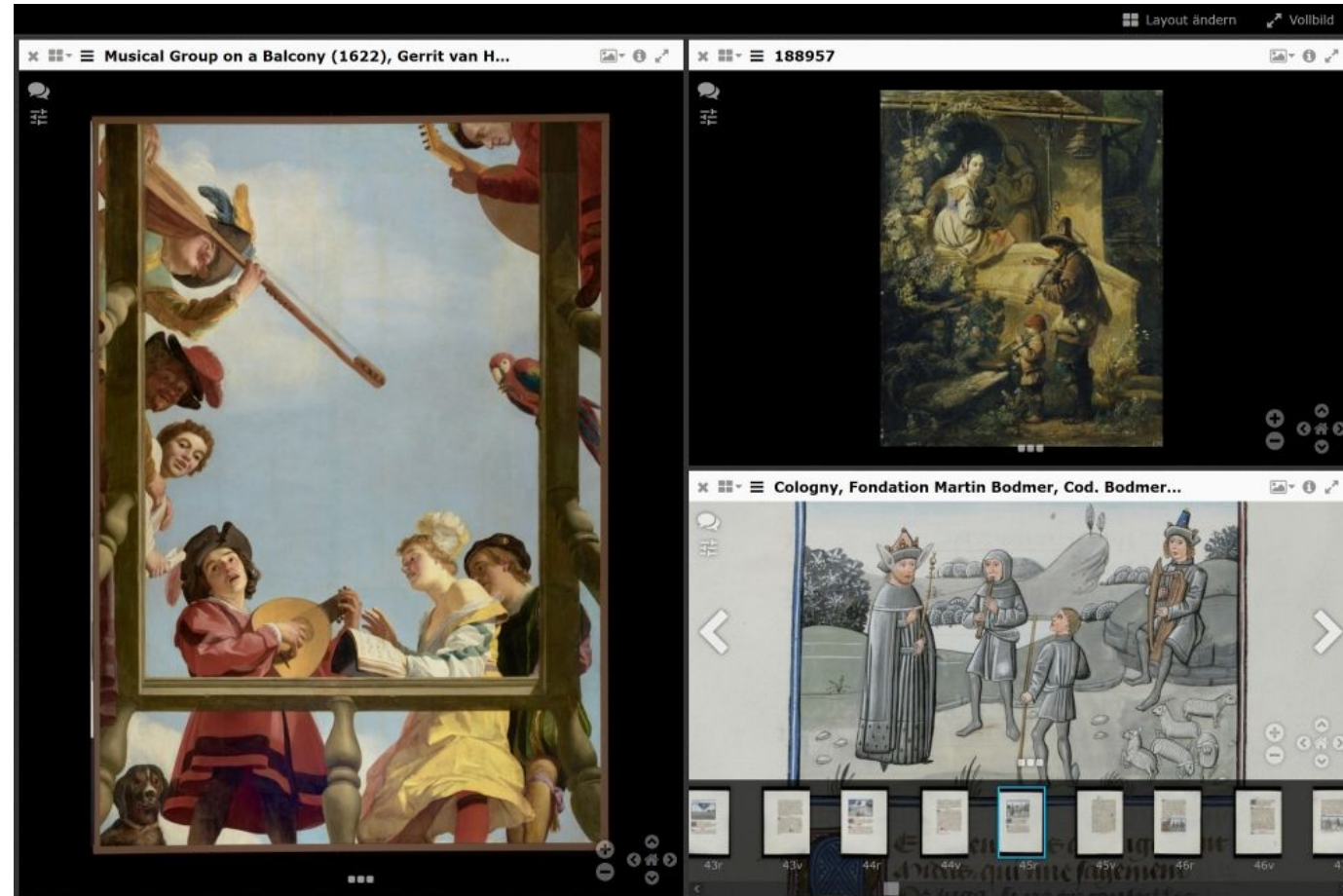


Quelle: Twitter; @hapyresearchers

Schnittstellen



- hochperformant
- standardisierte Schnittstellen für Metadaten, Images und Präsentation
- Zusammenführung von Bildern aus unterschiedlichen Quellen
- ermöglicht Vielzahl an Nutzungsmöglichkeiten in eigenem Viewer (Annotation etc.)



Quelle: Sonja Gasser; Mirador-Viewer
<https://www.artlog.net/de/notebook/iif-kleines-experiment-mit-einem-bildstandard>

03

Werkzeuge: Anwendungen und Kompetenzen

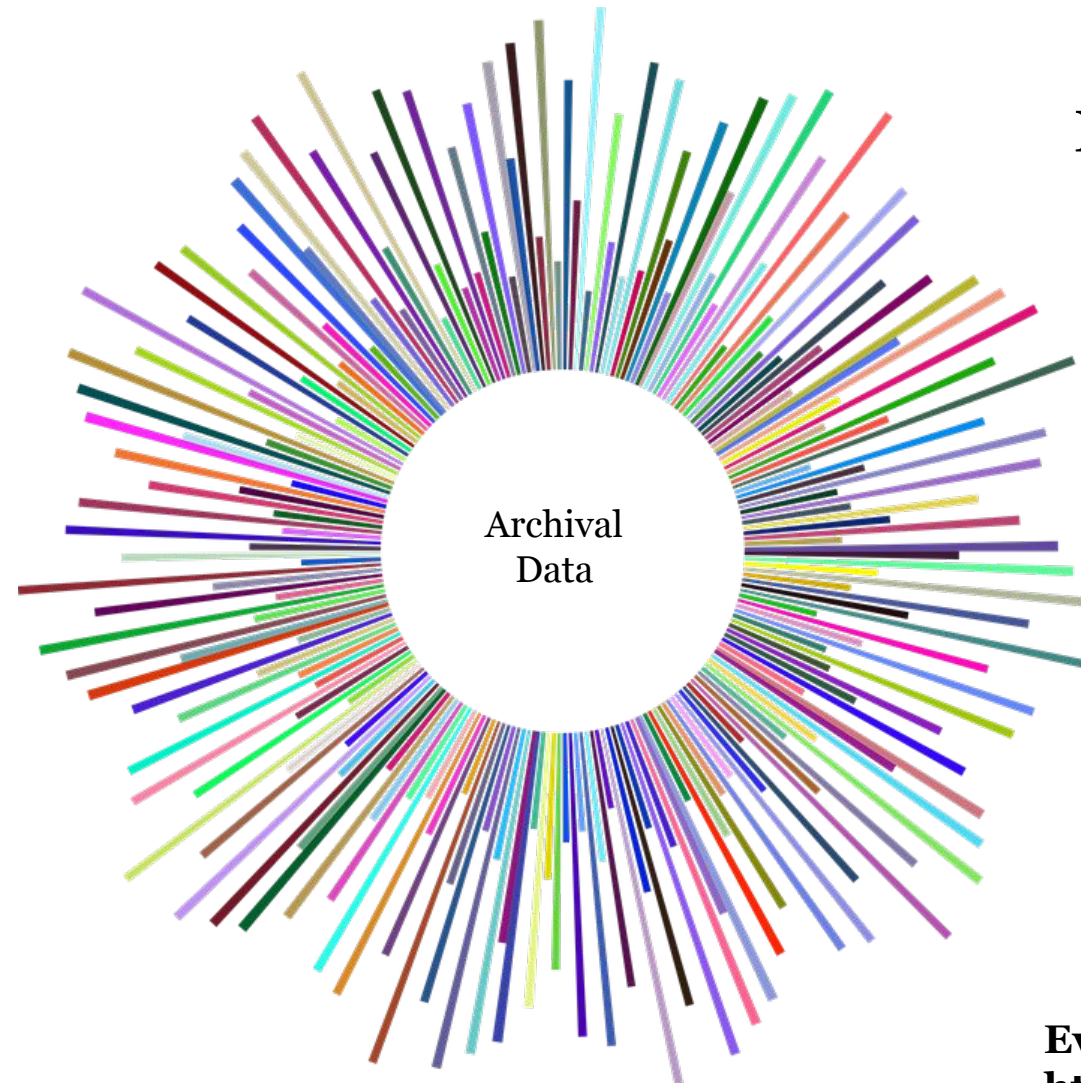
Anwendungen

Bereinigung

Transkription

Annotation

Geoinformationssystem



Datenanalyse

Datenstrukturierung

Visualisierung

Extraktion/Mining

Evaluierung u.a. im FDMLab:
<https://fdmlab.landesarchiv-bw.de/>

Kompetenzen

Data Literacy auf der Nutzerseite:
“the ability to collect, manage, evaluate,
and apply data in a critical manner” (C.
Ridsdale)

Aufgabenverteilung:
Archive stellen Daten über
Schnittstellen bereit, Forscher*innen
setzen passende Tools ein

Pflege des Werkzeugkastens:
gemeinsame Aufgabe von Archiven,
Forschung, Digital Humanities und
Informatik



E-Mail und Telefon

Ein größerer Anteil an Findmitteln und der größte Teil des Archivguts wird auch künftig nur analog zur Verfügung stehen.

Archivarinnen und Archivare beraten und unterstützen gerne!



04

Wissensextraktion von historischen Forschungsdaten

OCR und Texttranskription

Landratsamt Kehl
Fernsprecher 295/297
Nr. 1164 Abt. I a

Kehl a. Rh., den 23. Juni 1952

Wiedergutmachung nat. soz. Unrechts an Angehörigen des öffentlichen Dienstes

hier :
Antrag des Ratschreibers Karl [Name] in Legelshurst.

Blatt II Folio 1, 37
33 Blatt I
Legelshurst, 16. 5. 1952.

Wiedergutmachung nat. soz. Unrechts an Angehörigen des öff. Dienstes
Antrag des Ratschreibers Karl [Name] in Legelshurst.

Nr. 3326/III
024/35/III 19. MAI 1952

An Ministerium des Innern, Freiburg.

Für Ausweitung an den Bereich der Bearbeitung befindlichen Antrag.

I. Von interessanter Seite wurde mir darauf aufmerksam gemacht, daß auch der durch die Dienstentlassung 1933 abschließend verkündete Ruhegehalt aus Wiedergutmachung kommt. Ich bezog 1932/33 ein Gehalt von 1500 RM. jährlich für 4 Dienstjahren. 3 Monate nach meiner Entlassung - im Dezember 1933 - teilte mir der

Landesamt
für die Wiedergutmachung
Stuttgart

8.9.65

7 Stuttgart-1, den
Theodor-Heuss-Str. 26
Tel.: 296051/52
Az.: ST - 11737 u. - I Wei
8. 9. 65

Herrn
An
Landgerichtspräsidenten a.D. Marx
beim Landgericht Stuttgart
Stuttgart.

Betr.: Entschädigungsantrag

Bezug: Dortiges Schreiben vom 13.5.65

Anl.: 1 Bd. Akten 11737
1 " " 11737-1

**Auskunft aus dem Strafregister
der Staatsanwaltschaft zu Mannheim**

Familiennam(e): [Redacted] Vornam(e): Jakob

Geburtsangab(e)n: (Tag, Monat, Jahr) 17. 1. 1878 Gemeinde: Mannheim Landgerichtsbezirk: Mannheim StraÙe: Praden Verwaltungsbezirk: Mannheim

Familienstand: ledig — verheiratet — verwitwet — geschieden

Vor- und Familien-(Geburts-)Name des (bzw. früheren) Ehegatt(en): Der Oberstaatsanwalt b. Landgericht Mannheim Eing 18. JAN 1947

Vor- und Familienname des Vaters: [Redacted] Vor- und Geburtsname der Mutter: [Redacted]

Stand (Beruf): Schneider ggf. des Ehemanns:

Wohnort: Bingen, Klakehardstr. 2 91

ggf. letzter Aufenthaltsort: [Redacted]

StraÙe und Hausnummer: [Redacted]

Staatsangehörigkeit: d. Heimatgemeinde: Mannheim Heimatbezirk: Mannheim

Verurteilung(en) vermerkt:

Nr.	am	durch	wegen	auf Grund von	zu	Bemerkungen
		Urteil				
						Keine Vorstrafen Die Richtigkeit der Personalien ist nicht nachgeprüft. Mannheim, den 20. Jan. 1947. Staatsanwaltschaft Der Strafregisterführer o.B. [Name] Justiz-Überinspektor Angew.

- Archivdokumente und Akten sind besondere Textdokumente
- Formulare
- gedruckte Texte, maschinengeschriebene Texte, handschriftliche Texte
- Stempel und Unterschriften
- Fließtexte, Tabellen und strukturierte Datenfelder.

Eine Wiedereinstellung als Ratschreiber der Gemeinde Legelshurst erfolgte erst wieder auf Grund der Bestellung durch den Herrn Landrat des Kreises Kehl mit Schreiben vom 4. September 1945. Infolge der tatsächlichen Sachlage befürwortet das Bürgermeisteramt Legelshurst die Antragstellung auf Wiedergutmachung.

In der Anlage werden die Akten zur Einsichtnahme übersandt mit dem Bemerken, daß die Akten zur Einsichtnahme durch Dritte oder Aushändigung - auch einzelner Teile - an Dritte bedarf es der Einholung unseres Einverständnisses.

(vollandt)

STR. 9

OCR und Texttranskription

Badischer Beobachter

Donnerstag den fünften September - 18. Jahrgang

Nr. 41

Preis 10. Kreuzer

Der Lehrer der Menschheit

Samt 16. Heftchen der Ehrenbeilage des Hesperus

- Archivdokumente und Akten sind besondere Textdokumente
 - historische Schriftformen,
 - ungewöhnliche Formatierung und Layout,
 - abweichende Sprachmodelle,
 - semantische Drift.

Paulus
Christian
Christoph
Lauterbach
Carl Scheffer
Georg Thoma
Das Königl.
Es wurde anfangs
im Jahre 167

folo Beth the Recapitulacion
pitulo
pica endeth the table of the Ru
esent book.

LE
TRES EXCEL
LENT ET ANCIEN
ieu Pythagorique, dict Rythmoma
chie, fort propre & tresvtil à la re
creation de

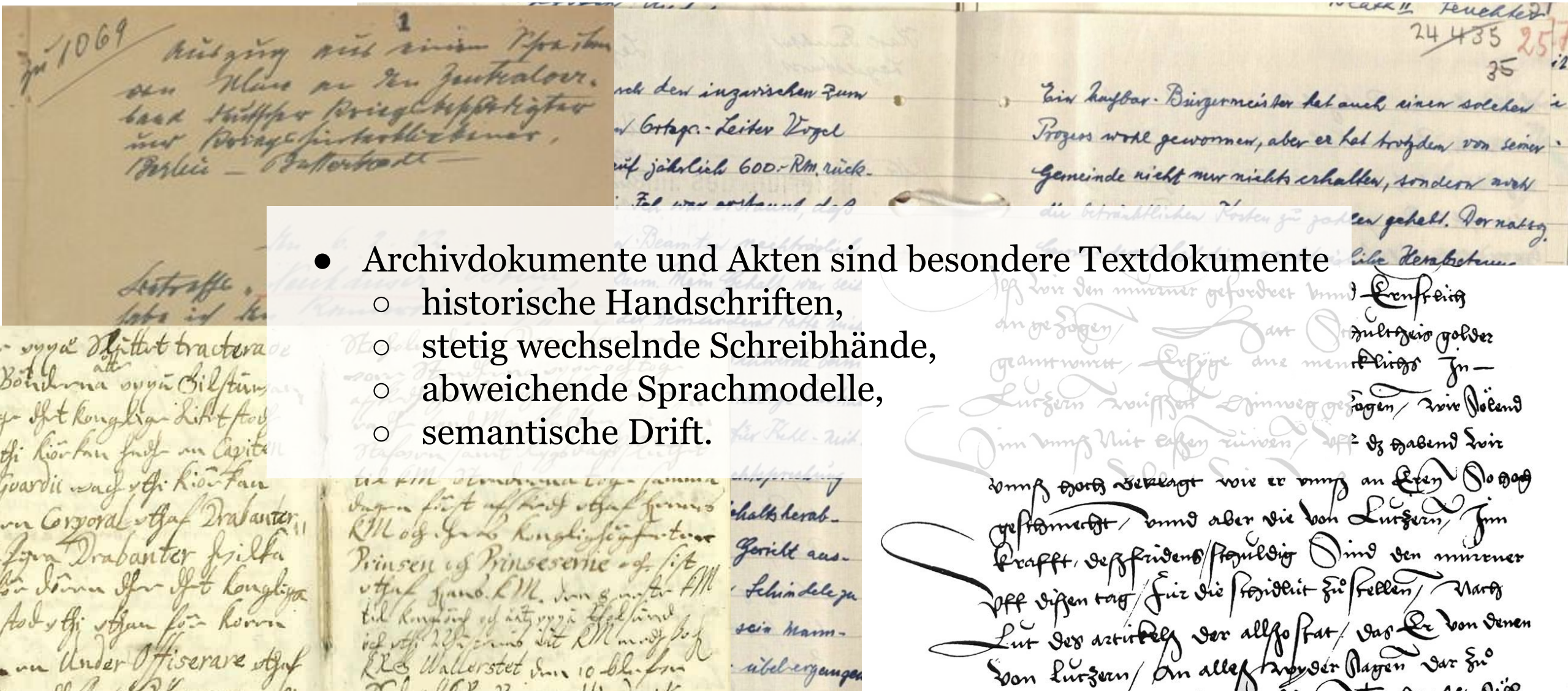
obtenir vra

Valerij Maximi Romane urbis viri illustri
bis Rome ex te
rarumq; gentiu facta simul
memoratu digna/que apud
tius diffusa sut/ut quambre
gnosa possint/ab illustribu
auctoribus/diligere stitui.
menta lumē uolentib; longe mquiliōnis lab
Rec mihi cuncta cōplectēdi cupido incessit. C
omnis eū gesta modico voluminū numero d
derit. Aut quis compos mentis! domestice per
historie seriem felici superiori stilo conditam!
ciorē cura! vel prestanciori facundia: traditur
uerit? Te igitur huic cepto penes quē homini

malz ein alle liam
Der herre pr
gelagt noch dem uecht. Der w
St. Gut. Es wert was sein thū
pitterkeit entpfāt. Der schalē
greiff er der schalē hertikeit. D
geseit. Spracher das ist mir w
ben mir verhouet meinen mu
zu der selben fart. Der lierne de
Dem selben affen sein gleich. U
reich. Die durch liurze pitterke
ge susikeit. wenne mā das te

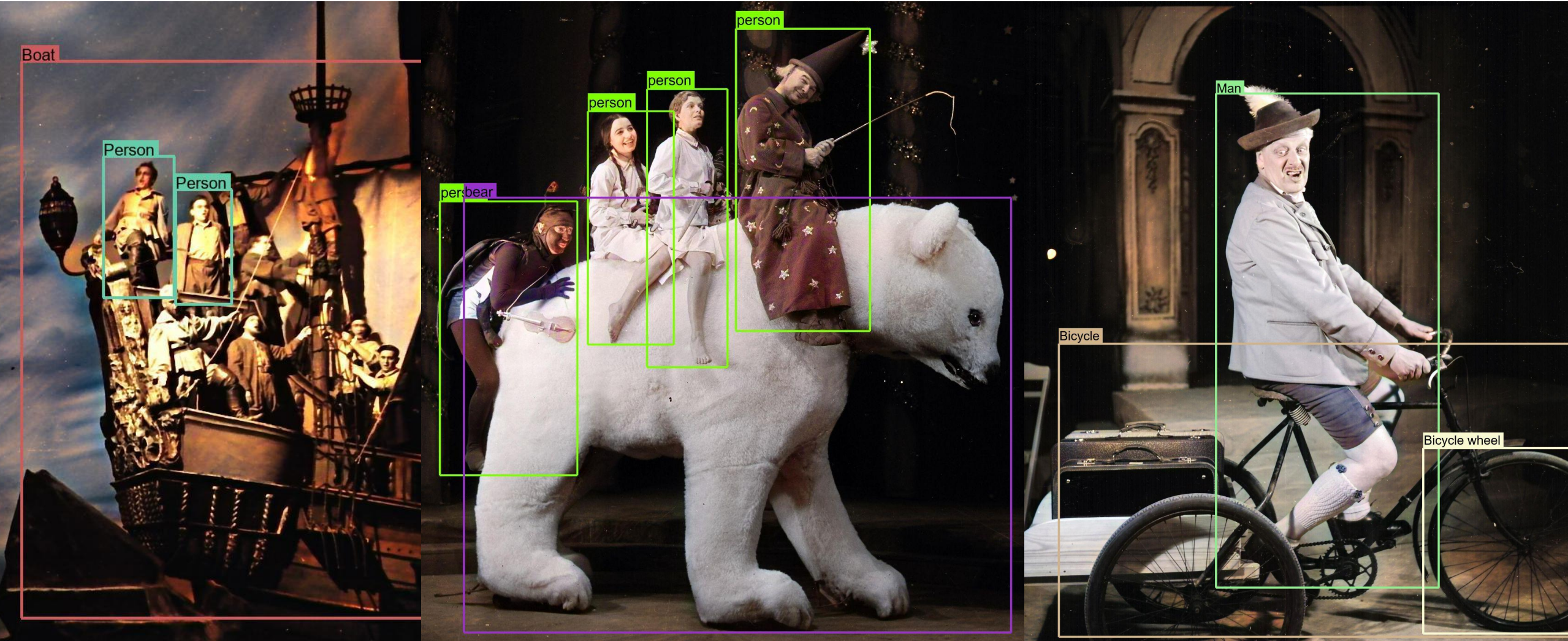
Mr Ludwig Wilhe
Marggraff zu Baa
Graff zu Saufenberg/ Graff zu Spanheim und
Lahr und Walsberg ic. Ob zwären auff Unse
den 3. Decemb. 1699. und 24. Novemb. 1700. ergangene/ und hin/ und wider affigirte l
befindende/ und bey vorgewehnten letzteren laudigen Kriegs-Weß. gänglich der Niere/ un
ser und Wohnungen aufserbauer worden/ Wir aber gnädigst gemeint seind/ und gern seh
völlig außgebracht/ und in guten Stand gesetzt werde. So haben Wir nicht allein bey
sagt Unsere Residenz-Stadt mit der bereits angefangenen Fortification gänglichen zu sch
cher völlig erbaut werde/ die so wol Unsern eingeseßnen Burgern/ als auch denen Fremdb
neue Häuser zu erbauen/ vermög ermeldester Patenten ertheilte Freyheiten hienit widerhol
wohl fremde/ was Nation und deren dreven im Röm. Reich/ vermög Frieden-Schluss
ländische/ welche in mehr ermeldester Unserer Residenz Model-mäßige Häuser bauen werd
blicationis drey Jahr gegeben/ denen Fremdben aber weder Zeit noch gewisser Termin p
gebaut und bezogen seyn wird/ von ihren Häusern und Gütern/ von der Frohn in allem a
rer Haus-Batter von Unsern eingeseßnen Burgern/ so solches Haus erbauer/ mit red abg
solchen Kindern Jeniges/ deme solches Haus an seinem Erbtheil zufallete/ frey/ ingleich
ein vor allemal nicht das geringste an Herrschafft. Gelder/ (worunter aber Reichs- und
Brandschagungen nicht verstanden) außser einer Stuben/ und Cammer zu Logierung Unse
dig seyn. Da aber ermelde Unse eingeseßne Burgere ihre Häuser in denen angezegete
den/ alsdann die oder derselbe obbenandte Freyheit nicht allein nicht zu genießen haben/ so
bauen angegeben/ überlassen werden solle/ zumahlen Wir öfters ersagt Unsere Residenz-
ten Wir auch ferners/ so viel die Leibhügenschaft betrifft/ die Fremdben nicht allein für si
tuum exempt/ nicht weniger auch das zur Erbauung solcher neuen Häusern benötigte W
folgen lassen. Sollten aber die Fremde einige burgerliche Güter an sich erhanden/ so v
gewöhnlichen Anlagen/ wie vorher/ underworfen; Wie es mit denen Inheimischen disfa
schen gemüßam expliciert worden. Welches zu Maniatschens Wissenschaft und l
affigirt wird. So geschehen Rastatt/ den 5. Decemb. 1701.

Handschriften und OCR



- Archivdokumente und Akten sind besondere Textdokumente
 - historische Handschriften,
 - stetig wechselnde Schreibhände,
 - abweichende Sprachmodelle,
 - semantische Drift.

Automatisierte visuelle Analyse



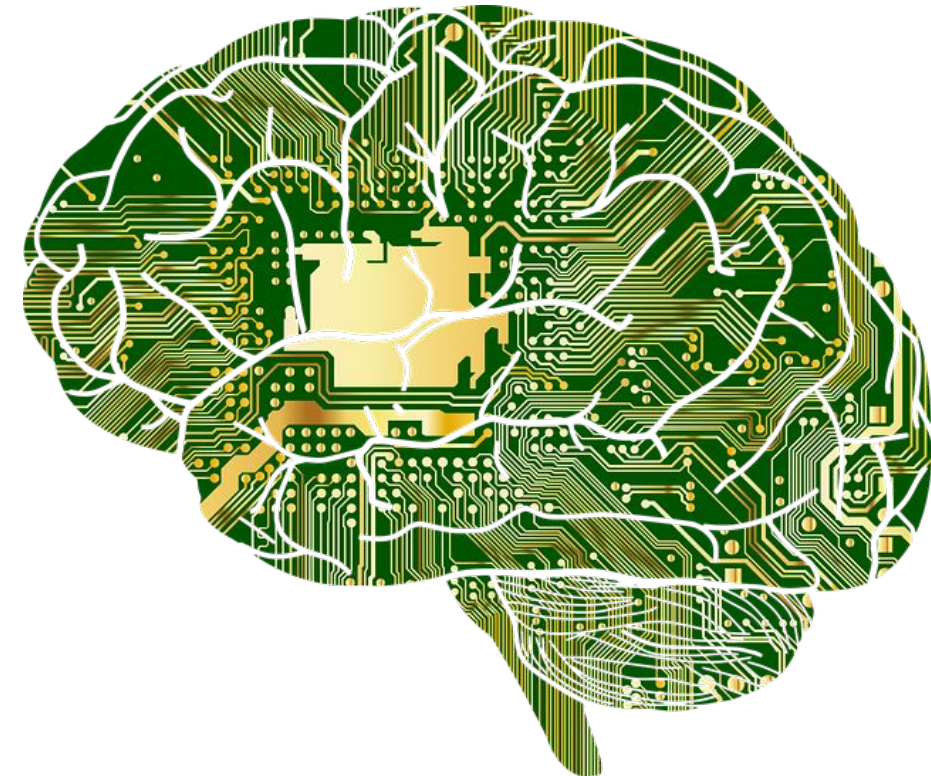
Automatisierte visuelle Analyse

- Archivdokumente sind besondere Bilddokumente
 - historische Fotografien
 - historische Gemälde und Zeichnungen
 - mit oftmals unklarem Kontext
- Deep Learning basierte Bild- und Objekterkennungsverfahren verfügen oft über einen “modernen” Bias



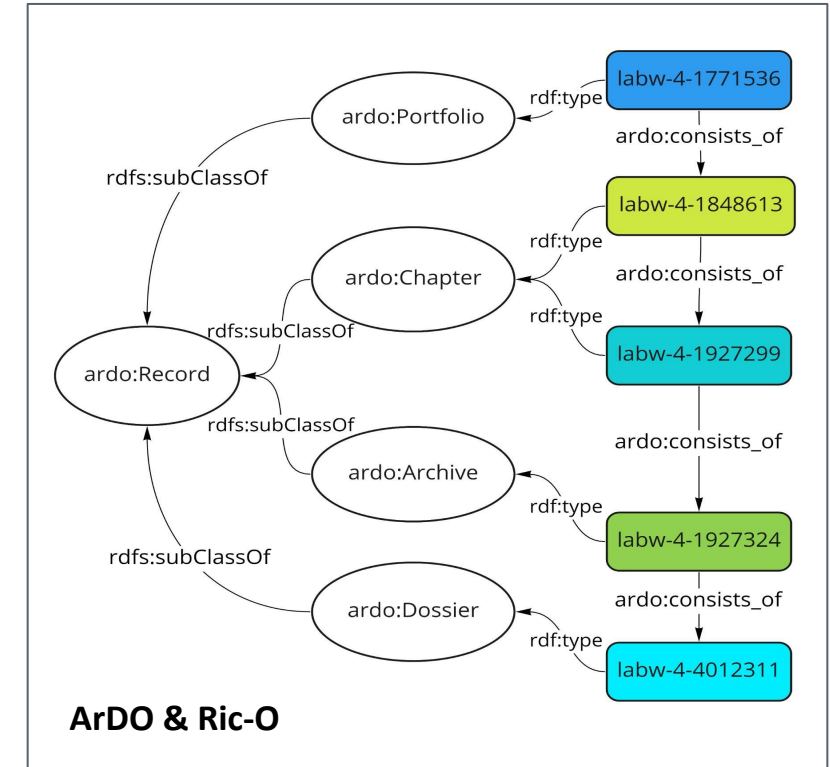
Language Models

- Repräsentation von natürlicher Sprache als statistisches Model
- Nahezu grenzenlos verfügbare Trainingsdaten
- Repräsentation der Semantik natürlicher Sprache durch Einbettung (Embedding) in einen (dichten) Vektorraum
- Potenzielle Einsatzfelder in Archiven:
 - Qualitätsverbesserung von OCR Transkripten
 - Ähnlichkeitsbasierte Suche
 - automatisierte Verschlagwortung
 - automatisierte inhaltliche Zusammenfassung
 - automatisierte Bild- und Objektbeschreibungen
 - automatisierte Übersetzungen
 - inhaltsbasierte Empfehlungssysteme



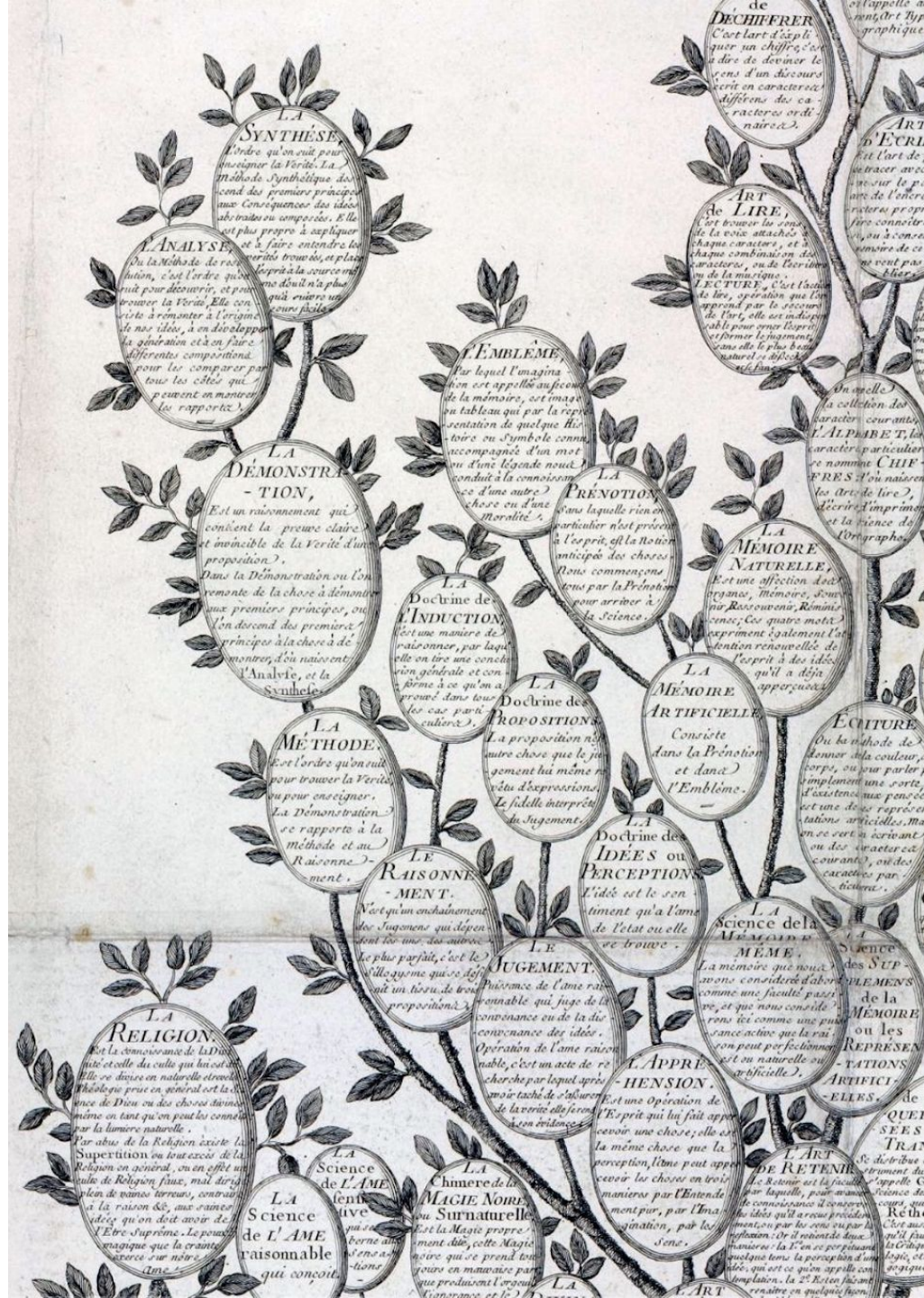
Semantische Repräsentation von Archivwissen

- Ontologien zur semantischen Repräsentation von Archivstrukturen, -vorgängen und -inhalten
 - (statische) strukturelle Organisation
 - Vorgänge und Prozesse
 - zeitliche Dynamik und Entwicklungen
 - Provenienz und Kontext
 - Klassifikation und Klassifikationsdynamik



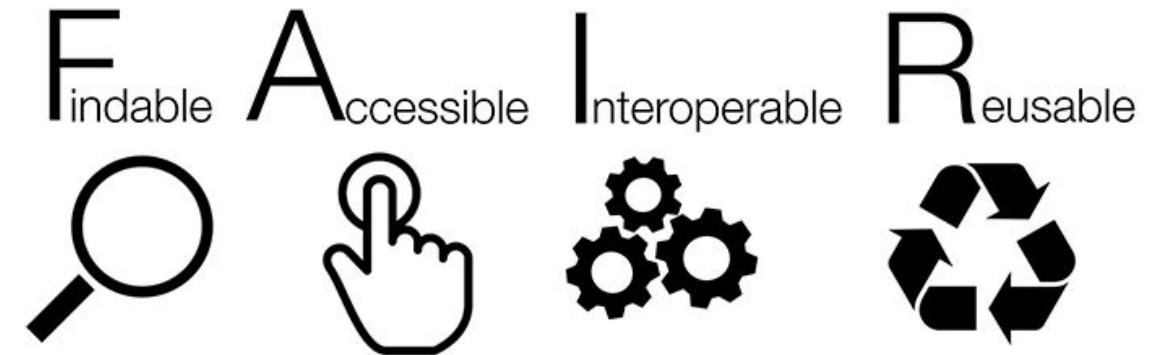
Semantische Analyse und Erschließung

- Wissensgraphen zur Erschließung und Anreicherung von Archivinhalten
- Semantische Annotation von Archivdokumenten
- Potenzielle Einsatzfelder in Archiven:
 - semantische Suche
 - Exploration des Archivbestandes
 - inhaltsbasierte Empfehlungen
 - Visualisierungen



Ontologien, Wissensgraphen und Normdaten

- **Ontologien** formalisieren die Bedeutung der durch sie repräsentierten Konzepte und Relationen
- **Wissensgraphen** verknüpfen Bedeutungsinhalte auf der Basis von Ontologien
- **Normdaten** bilden eine (traditionelle) Datenreferenz
- Miteinander verknüpft bilden sie die Basis zur effizienten Umsetzung der **FAIR Prinzipien** für Archivdaten

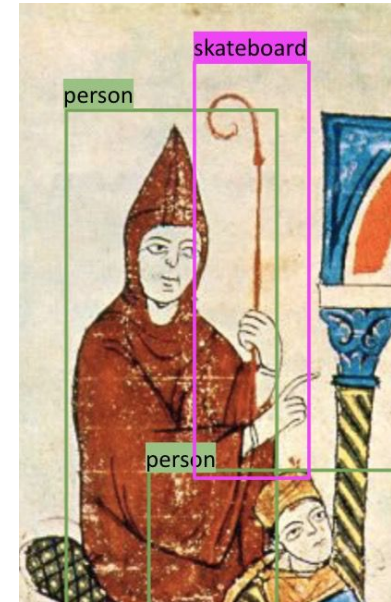


05

Angewandte Wissensgraphen

Zusammenspiel symbolischer und subsymbolischer KI

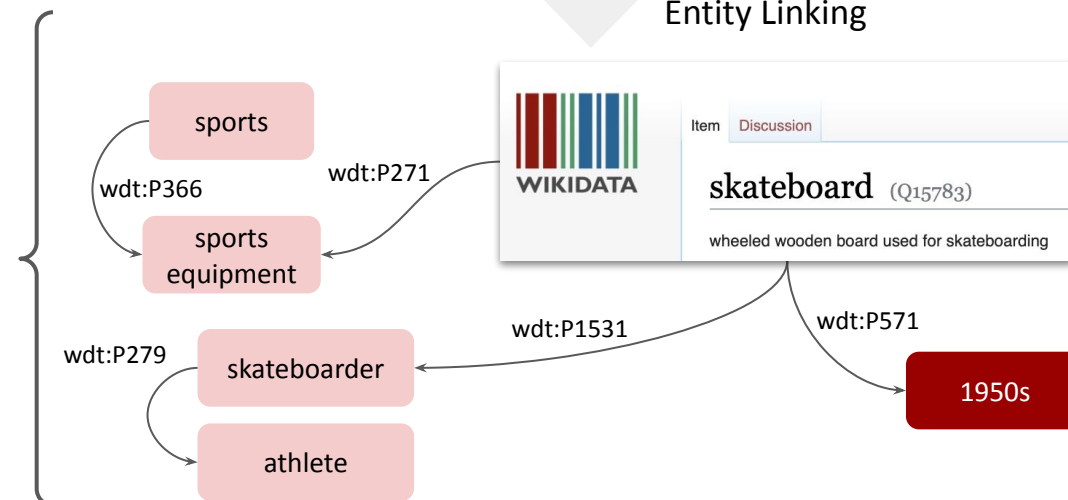
- Automatisierte Erschließungsverfahren erzeugen fehlerhafte Ergebnisse (machine learning = subsymbolische KI)
- Via Ontologien und Wissensgraphen (symbolische KI) können solche Fehler erkannt werden



Machine Learning
(subsymbolische KI)

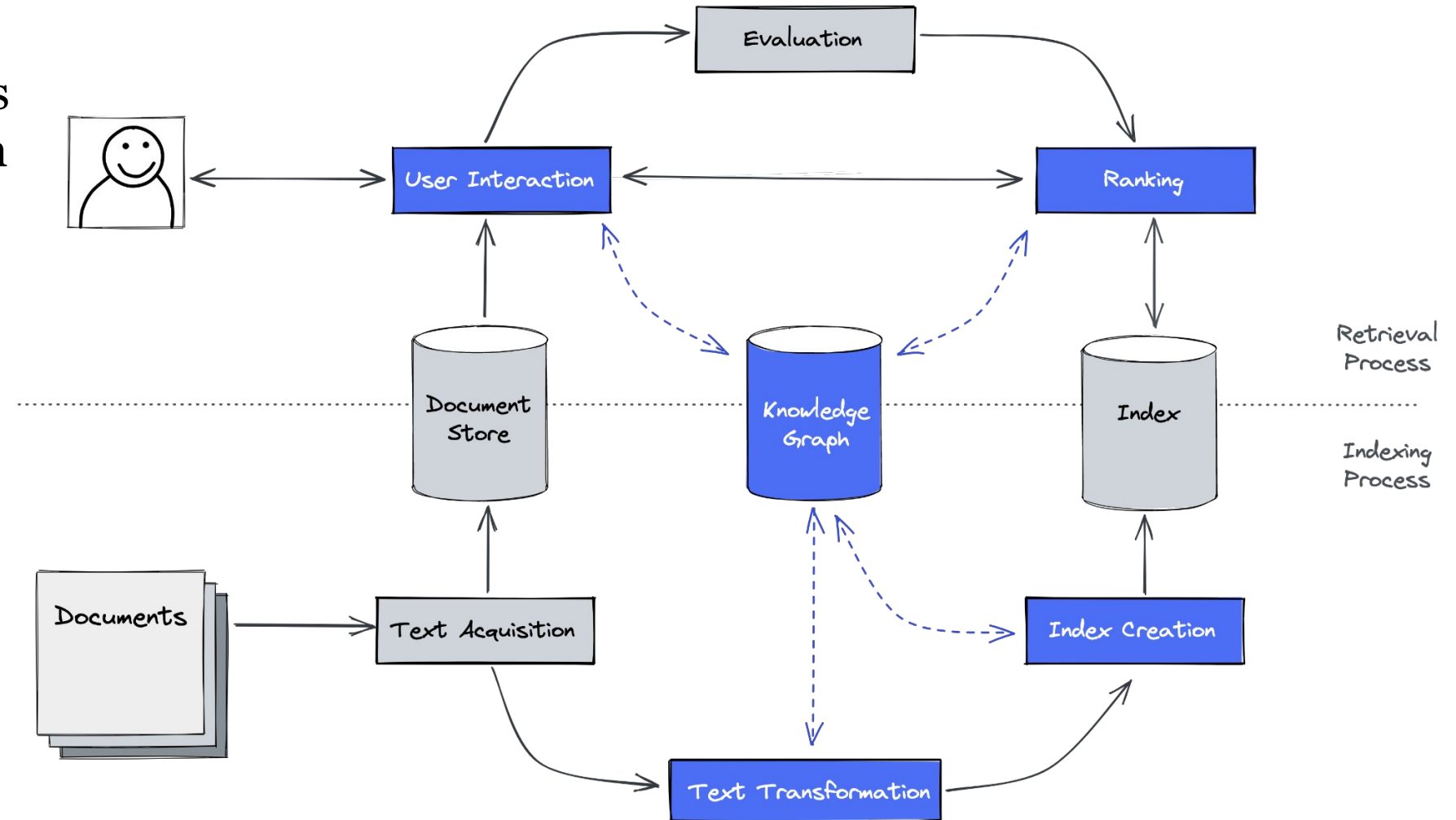
Metadaten
 Titel: Vita Mathildis
 Autor: Donizio
 Datum: 1115
 Ort: Bibliotheca Apostolica Vaticana...

externes Wissen
(symbolische KI)



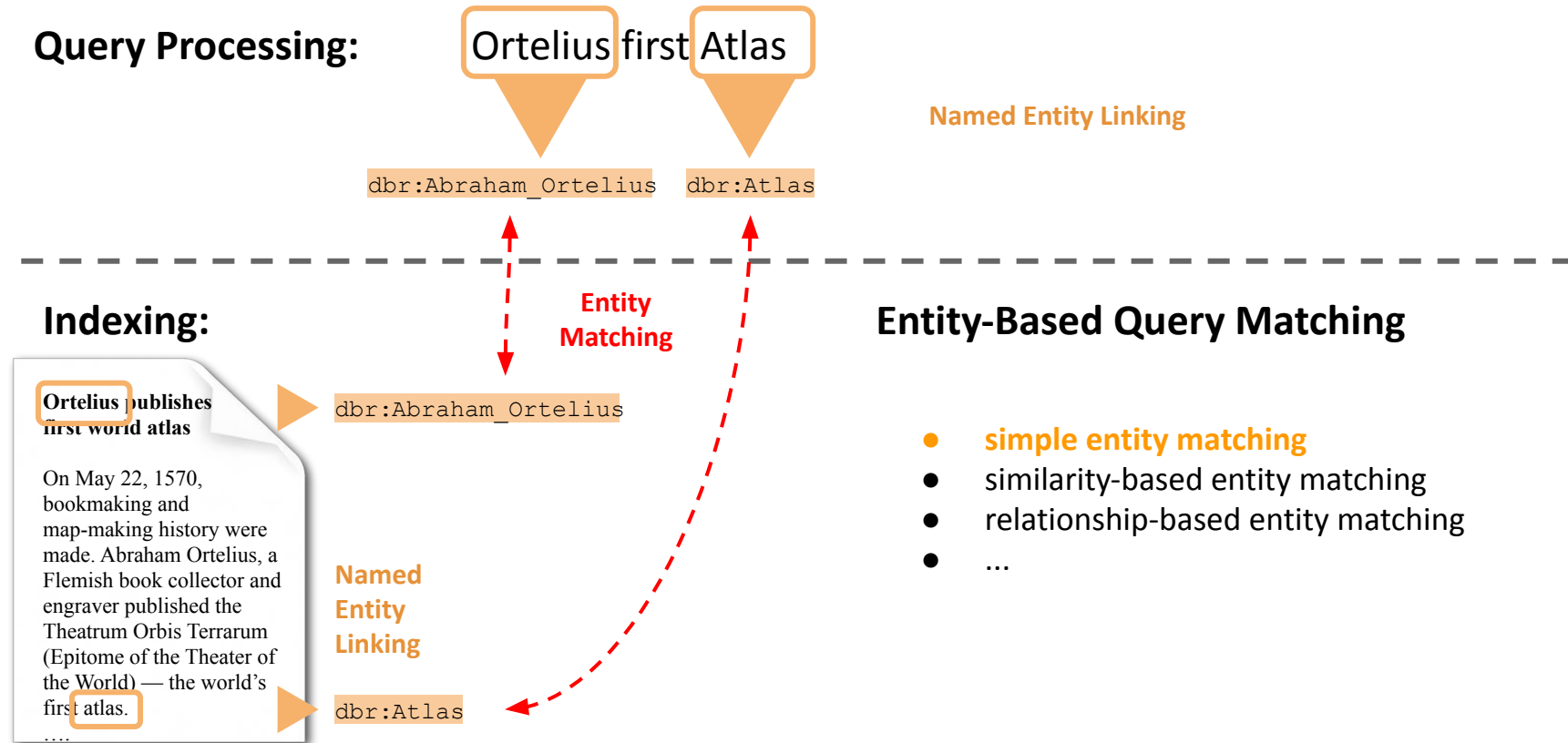
Semantisches Information Retrieval

- Knowledge Graph als zentrales Element im Retrieval Prozess
- steuert Indexerzeugung und -nutzung, Ranking und Benutzer Interaktion
- ermöglicht semantische und explorative Suche



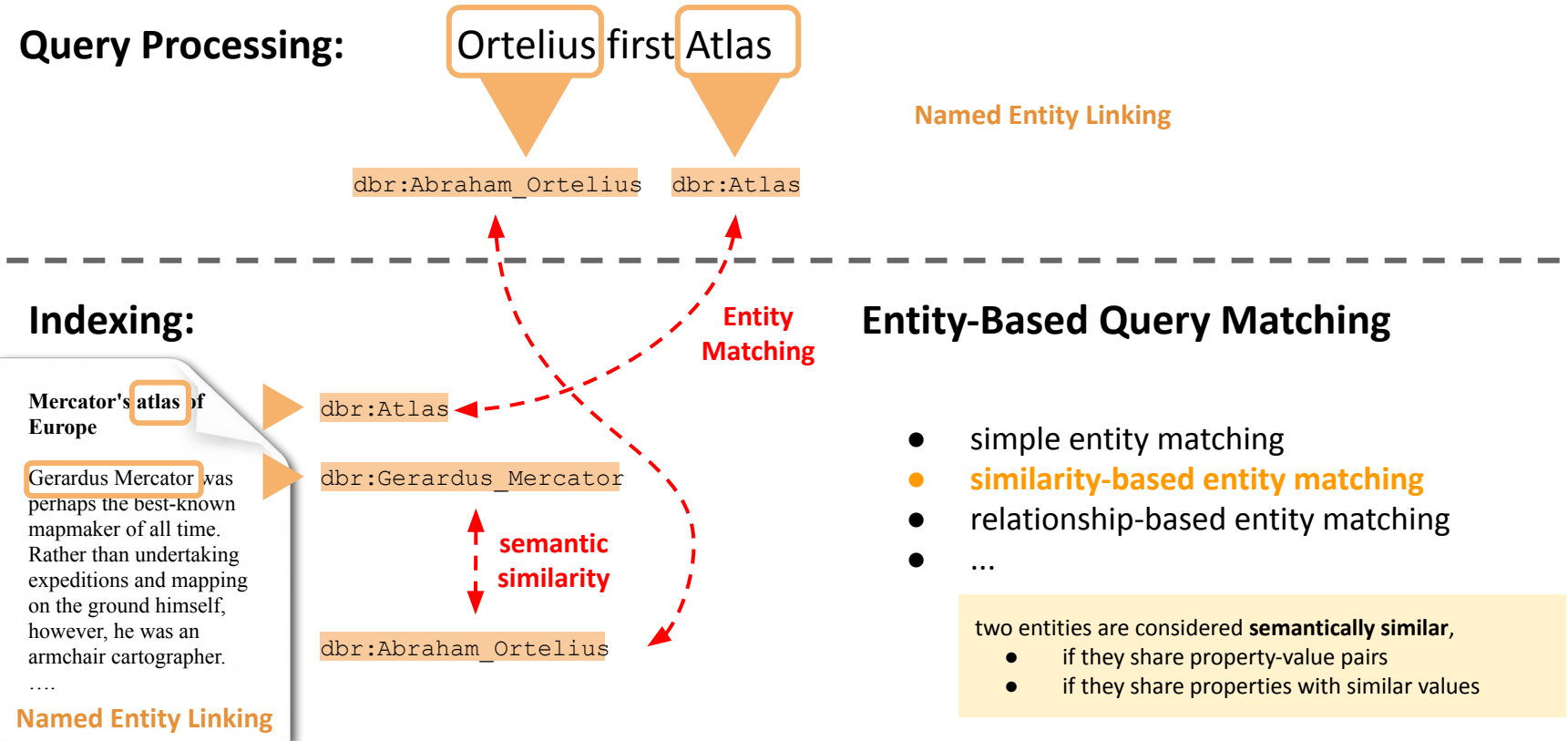
Entitäten-zentrierte Suche

- Suche nach bedeutungstragenden Entitäten
- “Things, not strings”



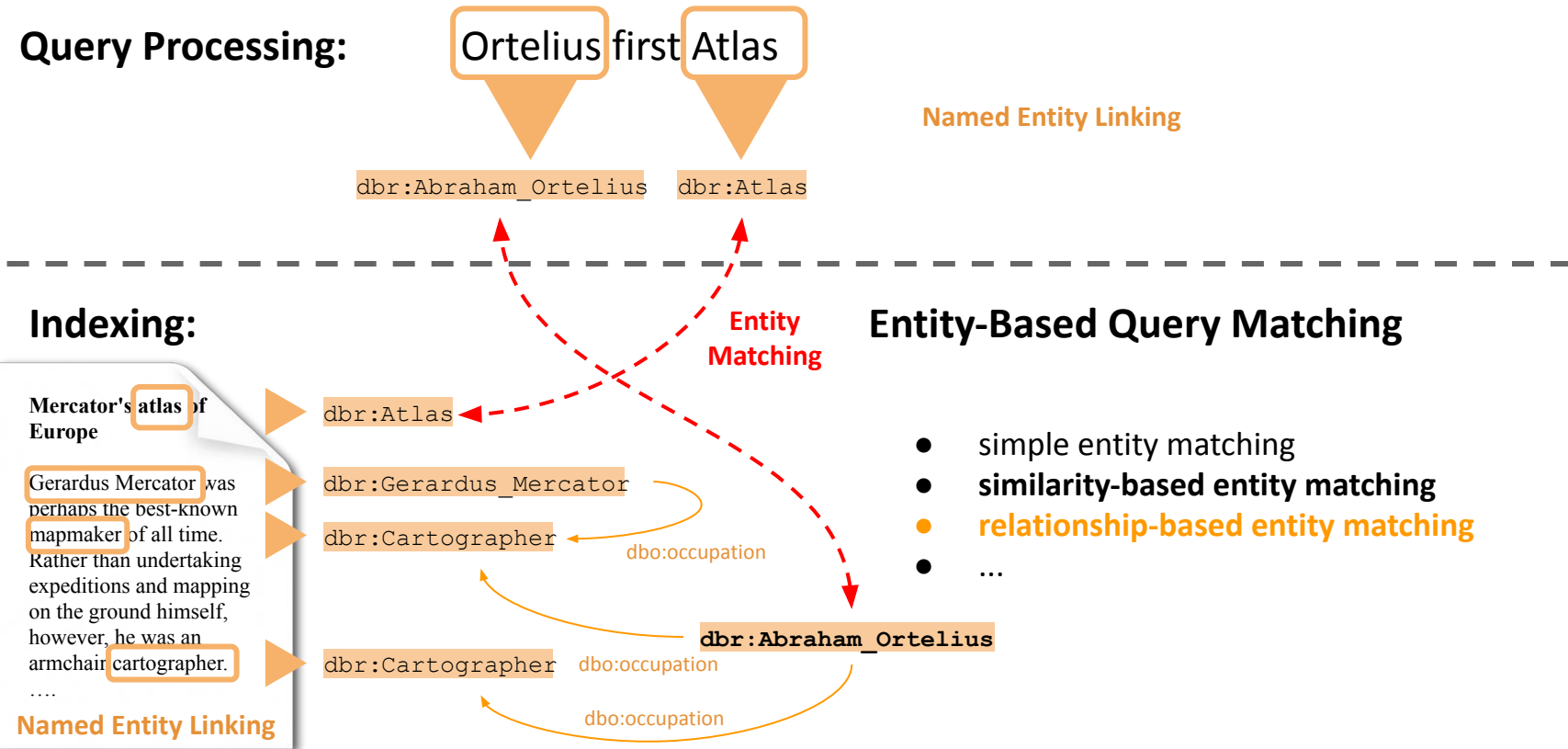
Entitäten-zentrierte Suche

- Suche nach bedeutungstragenden Entitäten
- “Things, not strings”
- **Einbeziehung semantischer Ähnlichkeiten**



Entitäten-zentrierte Suche

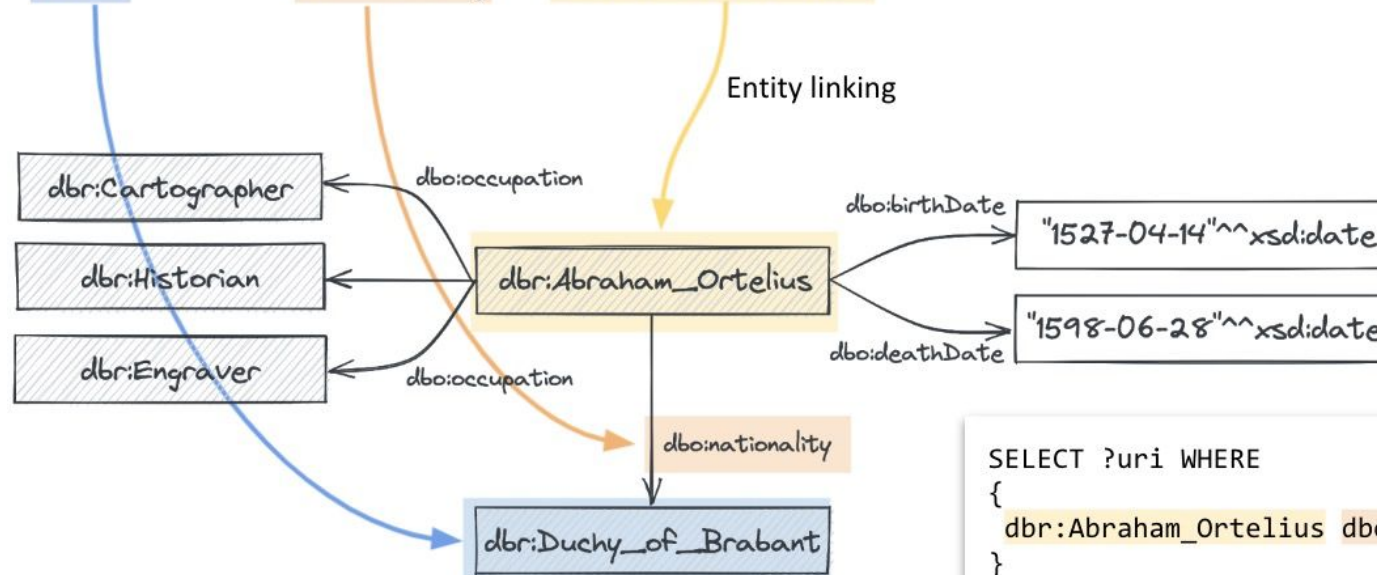
- Suche nach bedeutungstragenden Entitäten
- “Things, not strings”
- Einbeziehung semantischer Ähnlichkeiten
- **Einbeziehung von semantischen Beziehungen und Relationen**



Question Answering

- Beantwortung sachbezogener Fragen
- semantisch Analyse (NLP, Entity Linking, Relation Extraction)
- Formulierung von Knowledge Graph Abfragen (SPARQL)

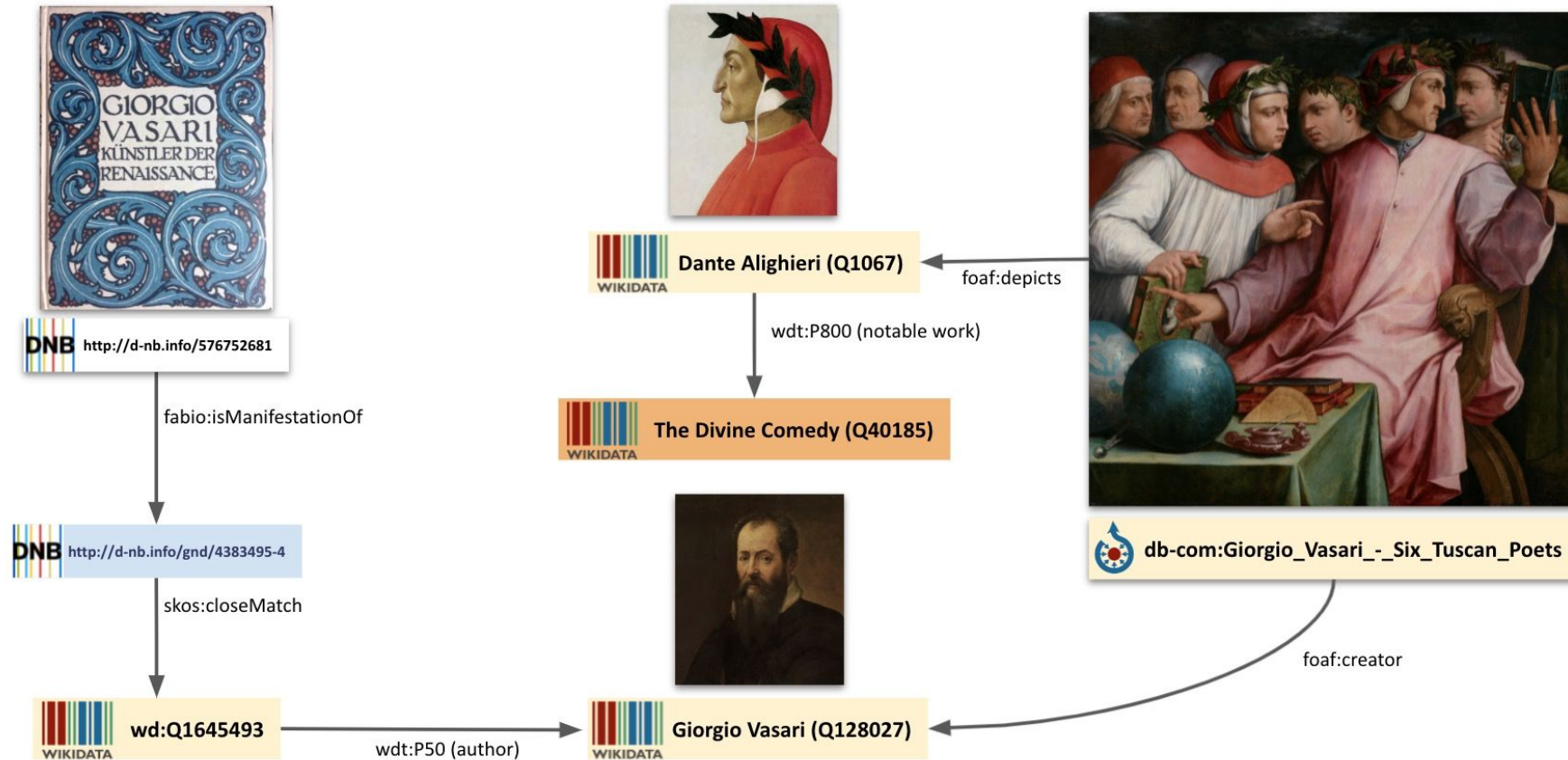
What was the nationality of Abraham Ortelius?



```
SELECT ?uri WHERE
{
  dbr:Abraham_Ortelius dbo:nationality ?uri.
}
```

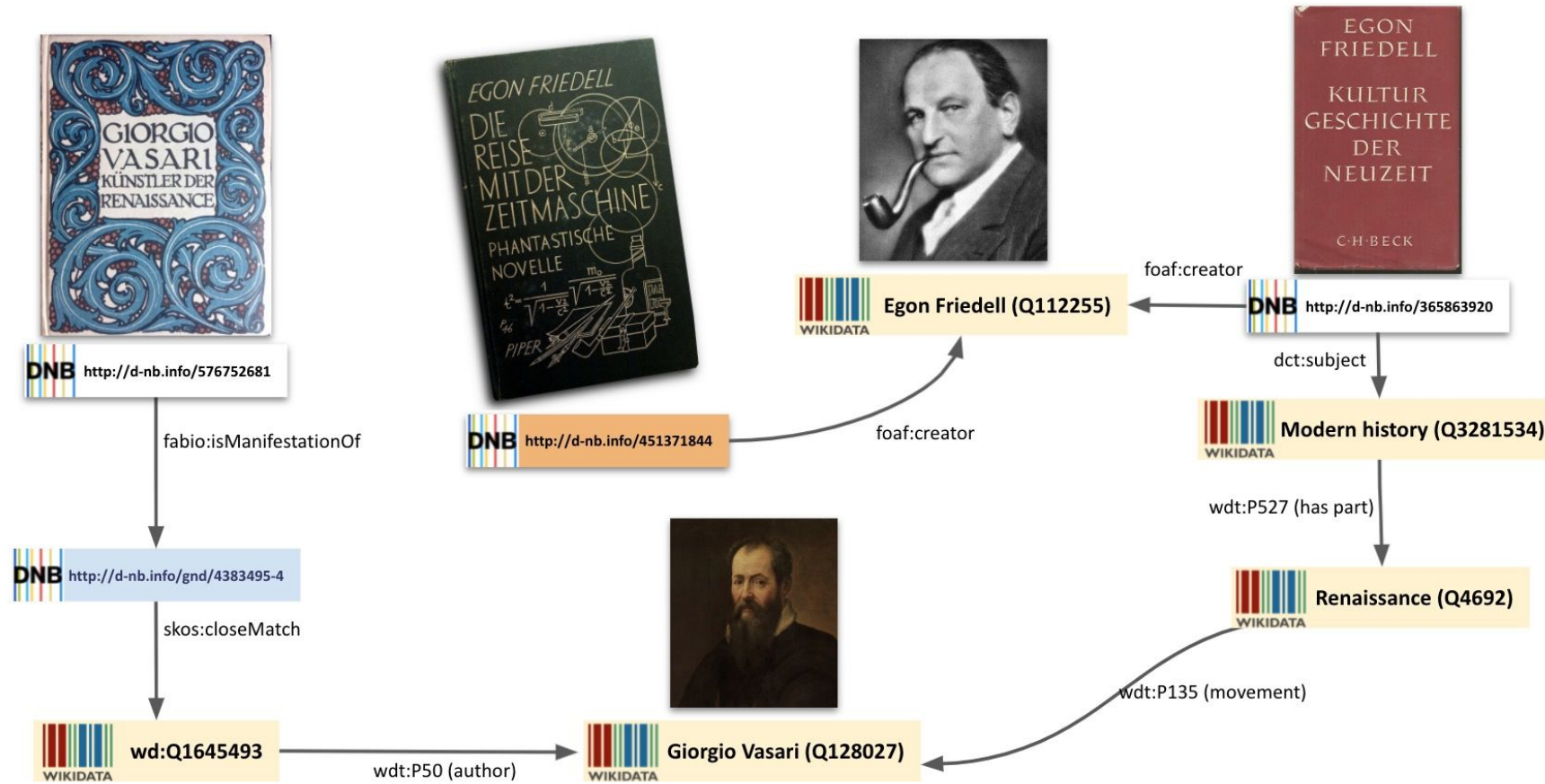
Explorative Suche

- Suche in unbekanntem Terrain
- unklares Suchziel
- komplexe Informationsbedürfnisse
- “Stöbern” statt zielgerichtetes Suchen



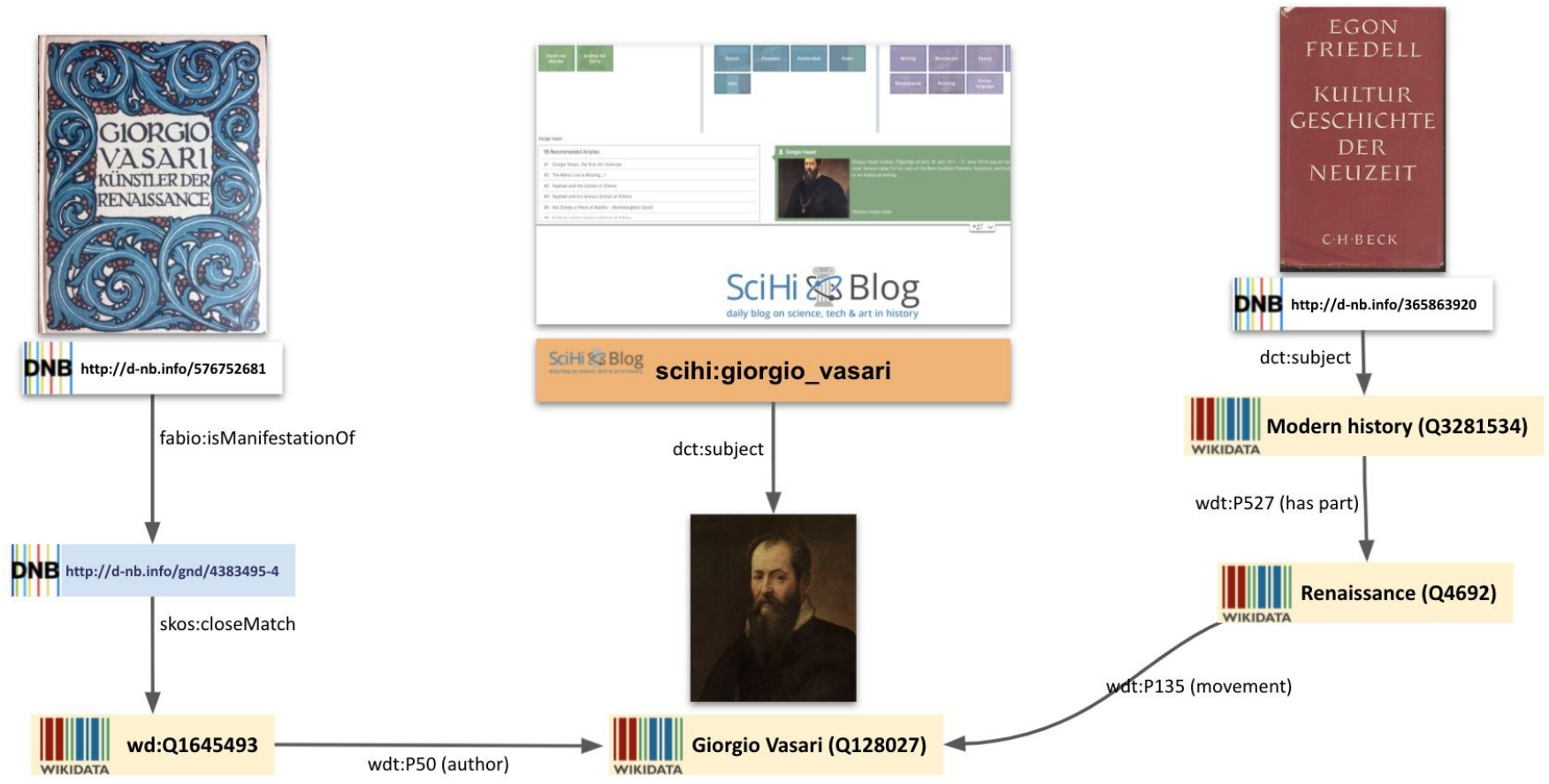
Explorative Suche

- überraschende Funde (“Serendipity”)



Explorative Suche

- überraschende Funde (“Serendipity”)
- Problem der Visualisierung (“Google-Effekt”)



Visualisierung und Exploration

<http://scih.org/>



On July 30, 1511, Italian Renaissance painter, architect, writer and historian **Giorgio Vasari** was born. He is best known today for his *Lives of the Most Eminent Painters* considered the ideological foundation of art-historical writing.

One of the Lesser Known Renaissance Artists

From all the great Renaissance artist, Giorgio Vasari might be one of the less this might be that although an artist of considerable reputation, there were so many Renaissance artists that Vasari's art didn't stand out among them. But, on the Vasari's major merits that we know about so many Renaissance artists authored and published the very first collection of biographies of famous artists.

Born in Arezzo

Vasari was born in Arezzo, Tuscany. Recommended at an early age by his cousin Luca Signorelli, he became a pupil of Guglielmo da Marsiglia, a skillful painter of stained glass. Sent to Florence at the age of fifteen by Cardinal Silvio Passerini, where he enjoyed the friendship and patronage of the Medici.

Giorgio Vasari	
Giorgio Vasari (Italian: [dʒordʒo vaˈzari]; 30 July 1511 – 27 June 1574) was an Italian painter, architect, writer and historian, most famous	
birth year	1511
death year	1574
is influenced of	Andrea del Sarto
field	Architect
birth place	Arezzo
death place	Florence



- Notwendige Voraussetzungen:
 - semantische Annotation
 - geeignete Werkzeuge zur Visualisierung

ivgut
BW)

Visualisierung und Exploration

The screenshot displays a semantic network visualization tool. On the left, a node for 'Karel van Mander' is connected to various nodes representing his attributes and relationships. These include 'deathPlace' (Italy), 'field' (writing, music), 'movement', 'field' (writing, music), 'Poetry', 'Architect', and 'Painting'. A green line connects 'Karel van Mander' to a node 'is influenced of', which points to a node 'is field of', which in turn points to a node for 'Giorgio Vasari'. On the right, a 'Relation Browser' and 'Timeline' are visible, with the timeline showing years 1511, 1574, 1861, and 1946. Below the main visualization, there is a section for 'Giorgio Vasari' with a list of 15 recommended articles and a detailed profile card featuring a portrait and a biographical summary.

Giorgio Vasari

15 Recommended Articles:

- #1 Giorgio Vasari, the first Art Historian
- #2 The Mona Lisa is Missing...!
- #3 Raphael and the School of Athens
- #4 Raphael and his famous School of Athens
- #5 Not Simply a Piece of Marble – Michelangelo's David
- #6 Raphael and his famous School of Athens

Giorgio Vasari

Giorgio Vasari (Italian: [ˈdʒordʒo vaˈzari]; 30 July 1511 – 27 June 1574) was an Italian painter, architect, writer and historian, most famous today for his *Lives of the Most Excellent Painters, Sculptors, and Architects*, considered the ideological foundation of art-historical writing.

DBpedia: Giorgio Vasari

- z.B. via refer.cx WordPress PlugIn at <http://scihi.org/>

Vielen Dank!

Prof. Dr. Harald Sack
Bereichsleiter Information Service
Engineering
FIZ Karlsruhe
harald.sack@fiz-karlsruhe.de

Daniel Fähle M.A.
Referatsleiter IT, digitale Dienste
Landesarchiv Baden-Württemberg
Abteilung Zentrale Dienste
daniel.faehle@la-bw.de