Landesarchiv
Baden-Württemberg
Zentrale Dienste

KI im Archiv? Aber wie!

Werkzeuge für die Erschließung und darüber hinaus

EDV-Tage 2022 am 29.09.2022

Benjamin Rosemann

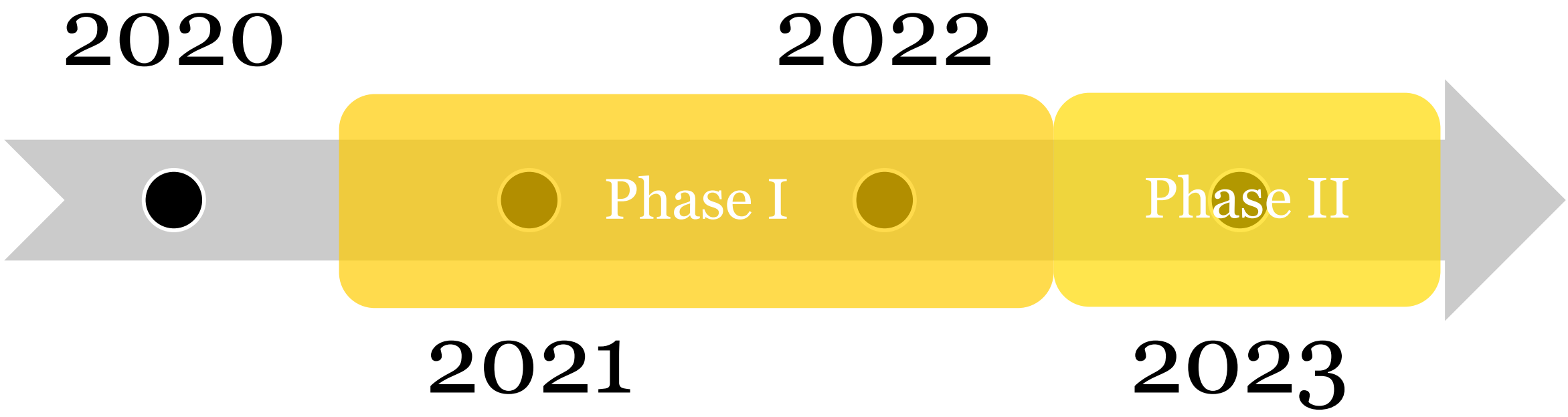
Landesarchiv
Baden-Württemberg
Zentrale Dienste

01

Das FDMLab

Laufzeit des FDMLab@LABW

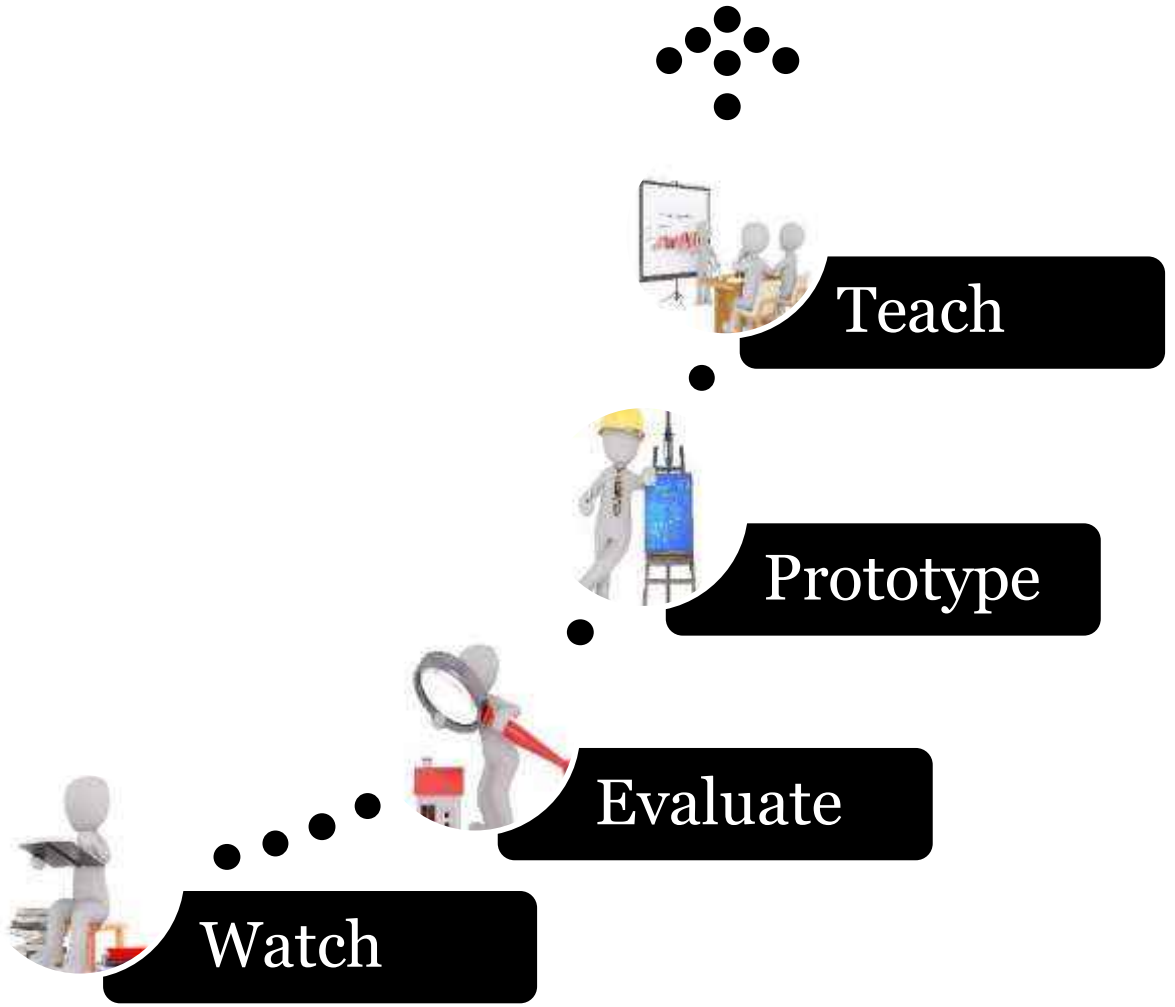
Landesarchiv Baden-Württemberg
Zentrale Dienste



Förderung durch die **Baden-Württemberg Stiftung** im Rahmen der **Zukunftsoffensive III.**

Vorgehen im FDMLab@LABW

Landesarchiv Baden-Württemberg
Zentrale Dienste



Landesarchiv
Baden-Württemberg
Zentrale Dienste

02

Was ist Künstliche Intelligenz?

Die Rechenmaschine

Es ist Sonntag Nachmittag und es regnet. Laut Werbeprospekt kosten 1kg Äpfel im Supermarkt 2 Euro.

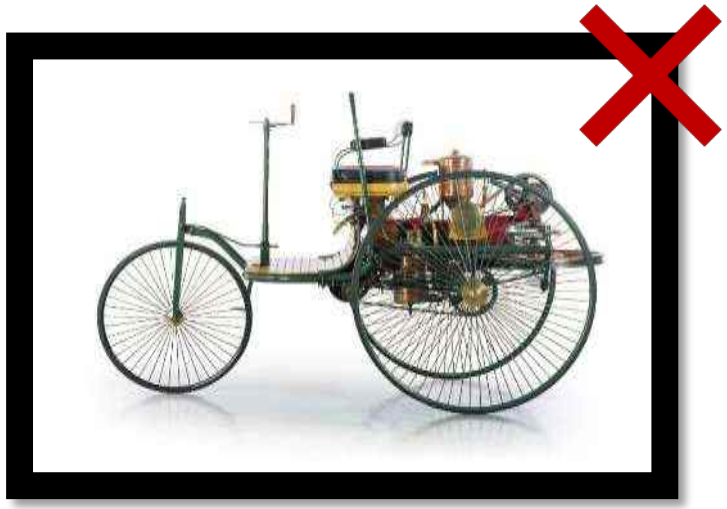
Ich möchte 5kg Äpfel. Wie viel Geld benötige ich zum Einkaufen?

Was ist Künstliche Intelligenz?

Landesarchiv Baden-Württemberg
Zentrale Dienste

Training

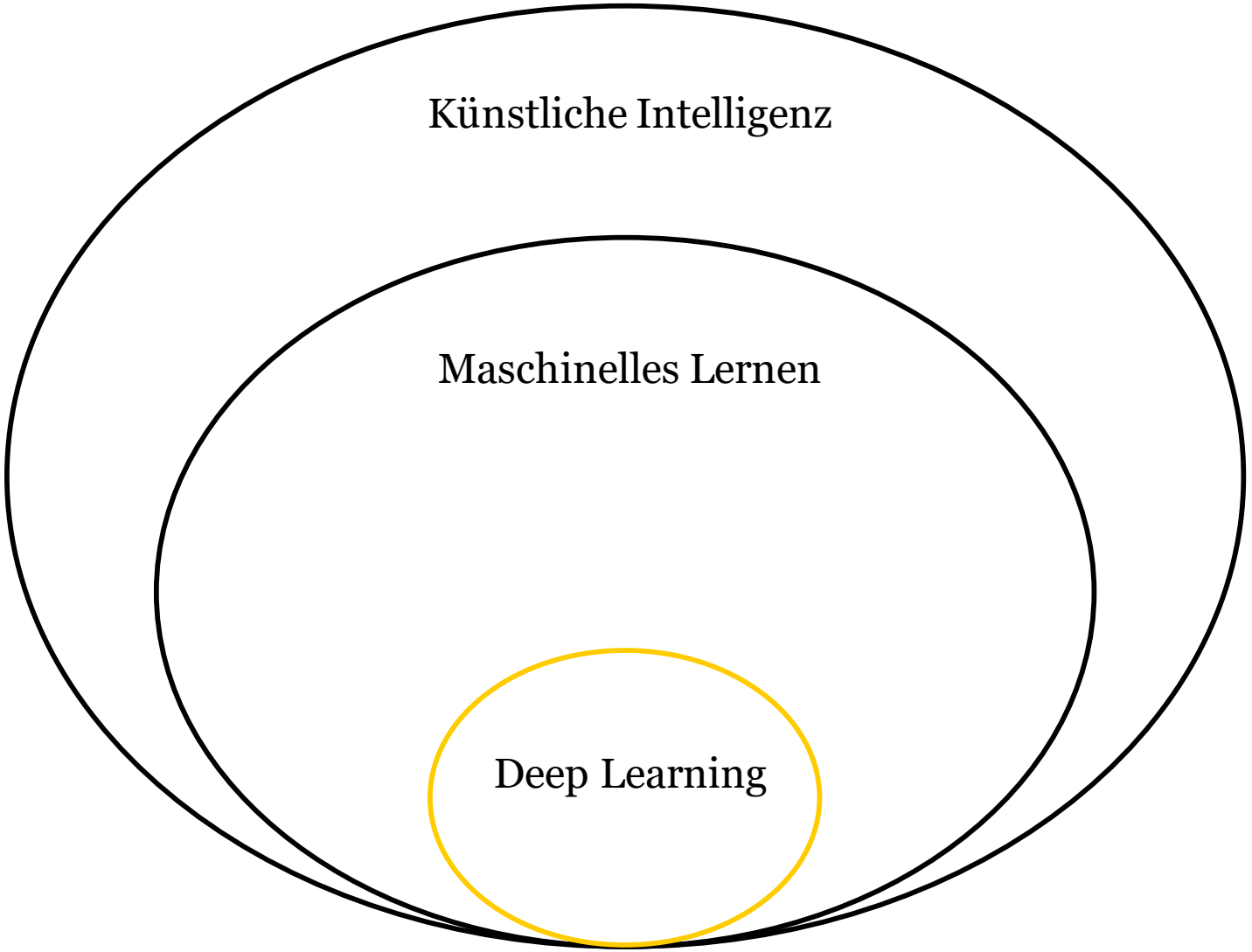
"Aptera 3"
von Chris Anthony, CEO, Aptera Motors Corp.
ist lizenziert unter einer CC BY-SA 4.0 Lizenz.



"Carl Benz Patent-Motorwagen Nummer 1"
von DaimlerChrysler AG
ist lizenziert unter einer CC BY-SA 3.0 Lizenz.

KI und Teilbereiche I

Landesarchiv Baden-Württemberg
Zentrale Dienste



KI und Teilbereiche II

Landesarchiv Baden-Württemberg
Zentrale Dienste

Natural Language Processing	Knowledge Representation	Automated Reasoning
Machine Learning	Computer Vision	Robotic

Quellen:

- Stuart Russell und Peter Norvig, "What is AI?", in *Artificial Intelligence: A Modern Approach* (Upper Saddle River, N.J.: Prentice Hall, 2010), 2-3.

5 Zutaten für KI- und Daten-Projekte

Landesarchiv Baden-Württemberg
Zentrale Dienste

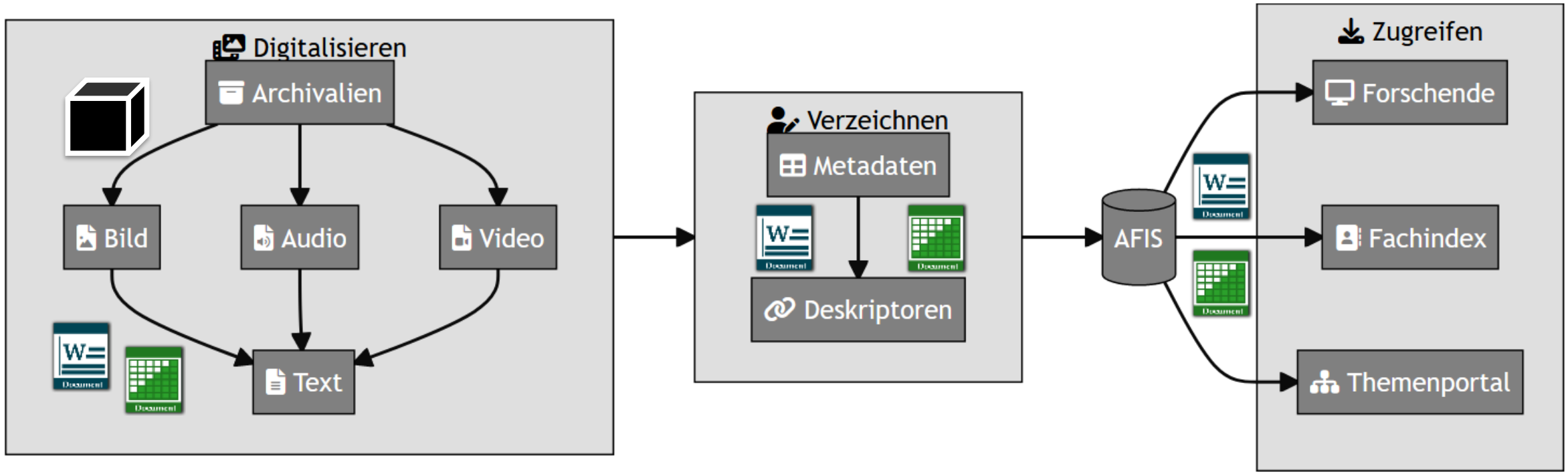
Landesarchiv Baden-Württemberg
Zentrale Dienste

03

Praktische Beispiele aus dem FDMLab

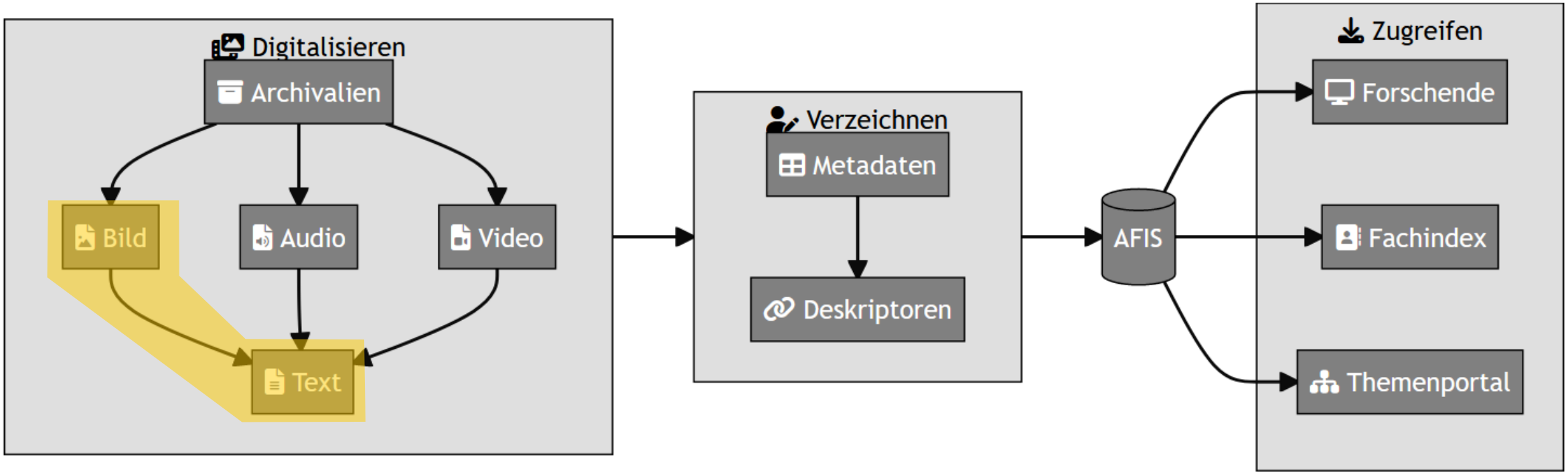
Workflow Digitalisierung

Landesarchiv Baden-Württemberg
Zentrale Dienste



Workflow Digitalisierung

Landesarchiv Baden-Württemberg
Zentrale Dienste



Volltexte erstellen mit KI Tools

Landesarchiv Baden-Württemberg
Zentrale Dienste

OCR-D

https://ocr-d.de/

Transkribus

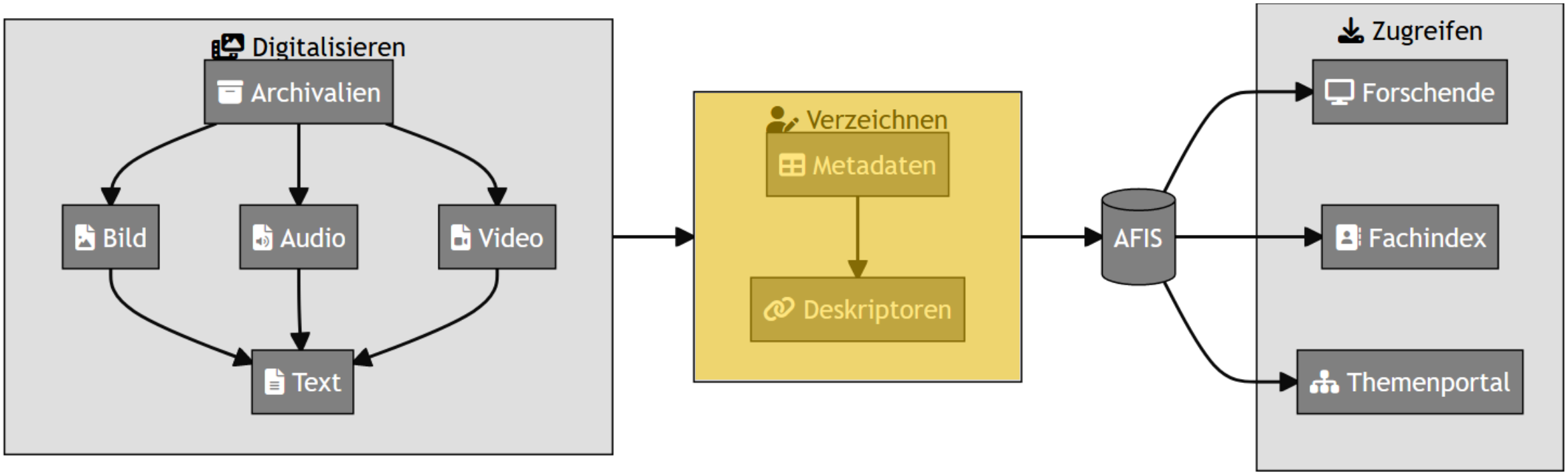
https://transkribus.eu/

eScriptorium

https://gitlab.com/scripta/escriptorium

Workflow Digitalisierung

Landesarchiv Baden-Württemberg
Zentrale Dienste



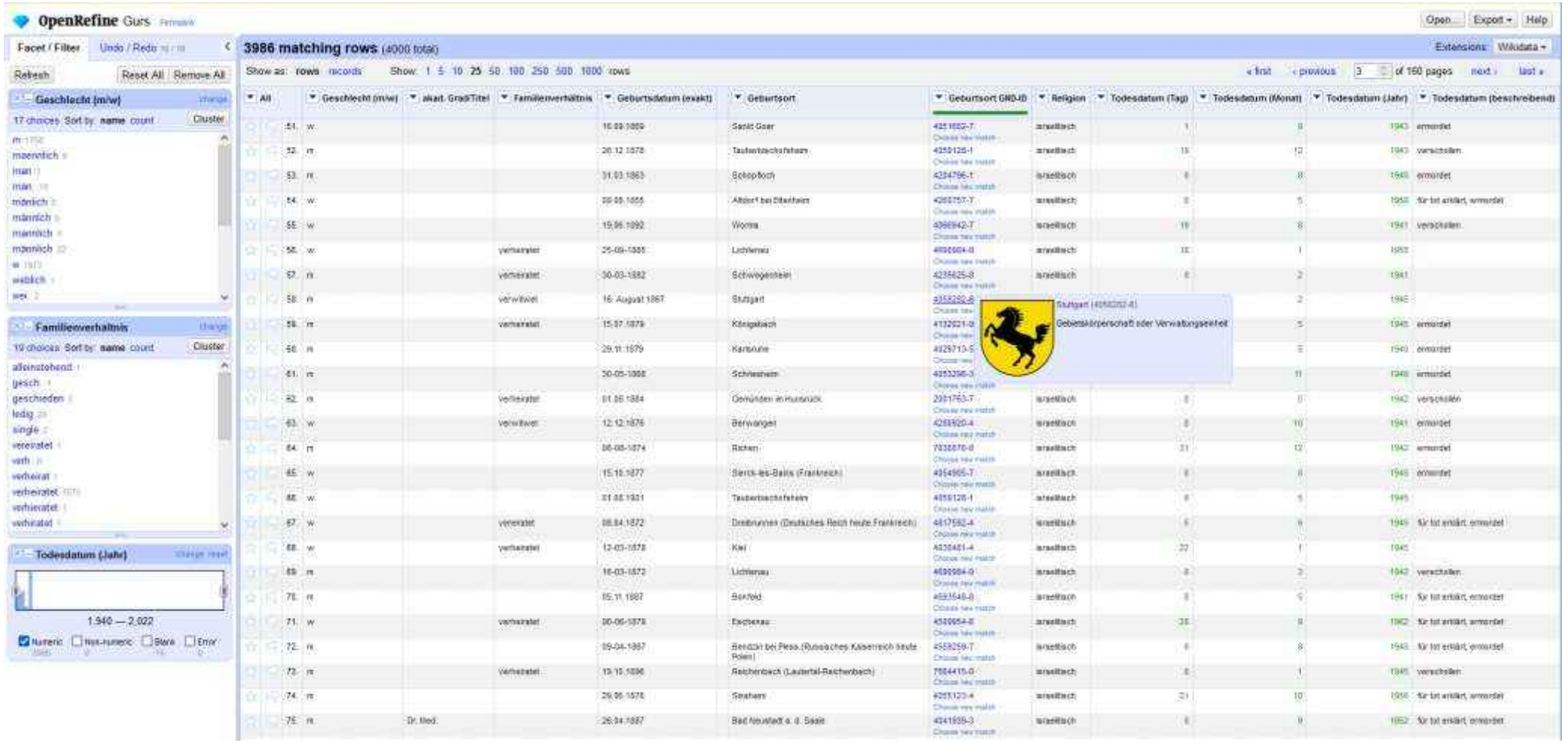
OpenRefine is like Excel on steroids!

Quelle unbekannt

https://openrefine.org/

OpenRefine

Landesarchiv Baden-Württemberg
Zentrale Dienste



Clustering (Unsupervised Learning)

Landesarchiv Baden-Württemberg
Zentrale Dienste

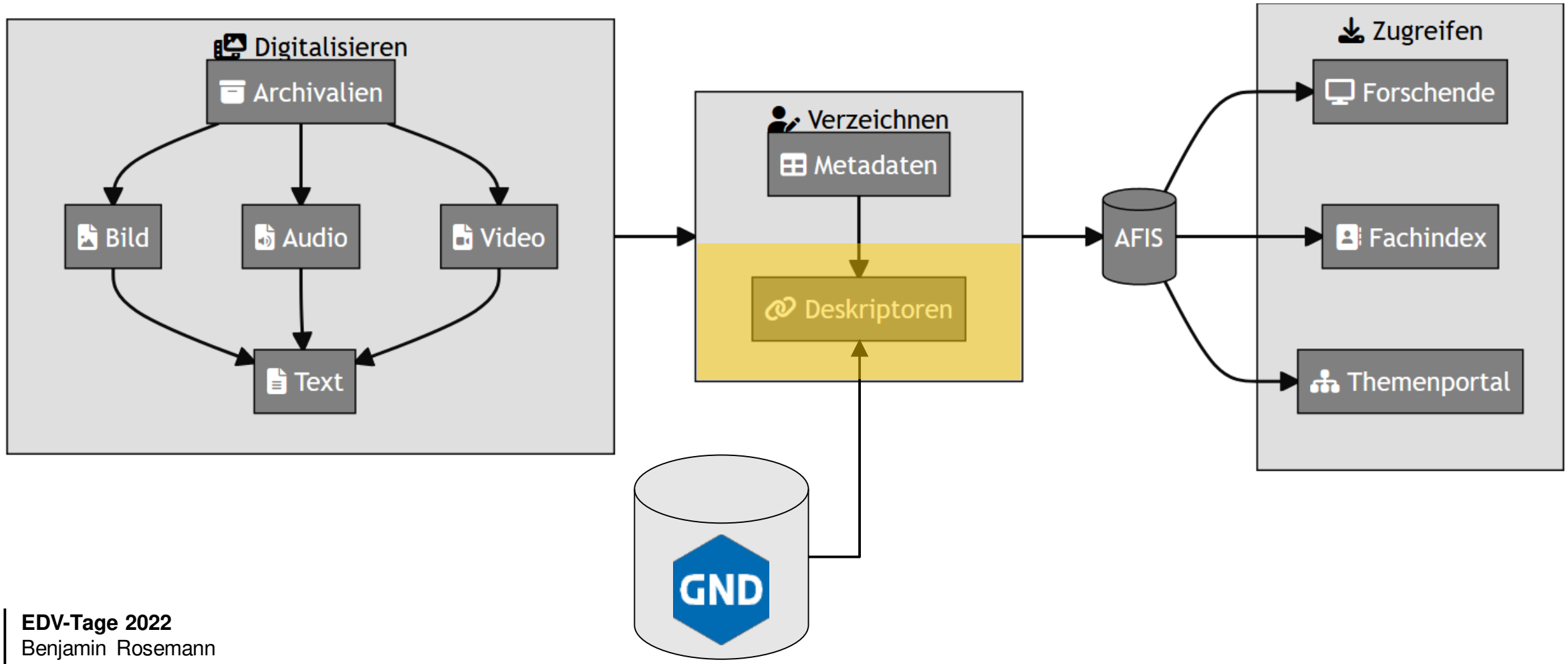
Integrierte Algorithmen

- FP: Fingerprinting
- FP: N-Gram Fingerprinting
- Phon: Cologne-phonetic
- Phon: Methaphone3
- Phon: Daitsch-Moktoff
- Phon: Baider-Morse
- NN: Levensthein
- NN: Prediction by Partial Matching

Name	Orte
Benjamin Rosemann	Nordschwaben (Rheinfelden)
Benjamin Rosemann	Nordschwaben bei Rheinfelden
Rosemann Benjamin	Nordschwaben-Rheinfelden
Beniamjn Rosemann	Rheinfelden-Nordschwaben
Benjamin Roseman	Nordschwaben / Rheinfelden
Benjamin Rosemam	Rheinfelden | Nordschwaben
Benjamin R. Rosemann	Rheinfelden, Nordschwaben

Workflow Digitalisierung

Landesarchiv Baden-Württemberg
Zentrale Dienste

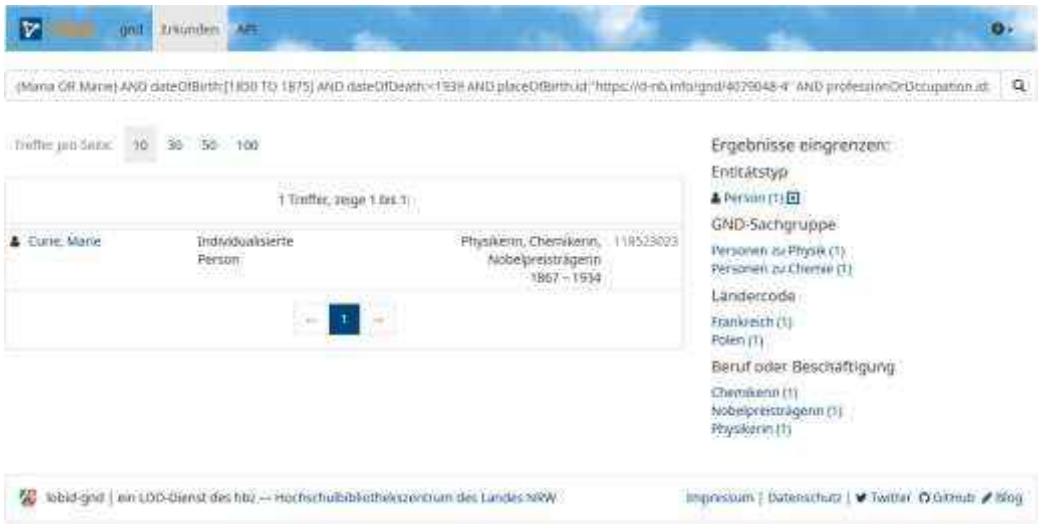


Suchtechnologien am Beispiel GND

Landesarchiv Baden-Württemberg
Zentrale Dienste

Suche in GND via lobid

```
(Maria OR Marie) AND
dateOfBirth:[1850 TO 1875] AND
dateOfDeath:<1939 AND
placeOfBirth.id:
  "https://d-nb.info/gnd/4079048-4" AND  # Warschau
professionOrOccupation.id:(
  "https://d-nb.info/gnd/4219058-7" OR     # Chemikerin
  "https://d-nb.info/gnd/4337175-9" OR     # Physikerin
  "https://d-nb.info/gnd/4202451-1")       # Naturwissenschaftlerin
```



Bildschirmfoto von
https://lobid.org/gnd/search?q=%28Maria+OR+Marie%29+AND+dateOfBirth%3A%5B1850+TO+1875%5D+AND+dateOfDeath%3A%3C1939+AND+placeOfBirth.id%3A%22https%3A%2F%2Fd-nb.info%2Fgnd%2F4079048-4%22+AND+professionOrOccupation.id%3A%28%22https%3A%2F%2Fd-nb.info%2Fgnd%2F4219058-7%22+OR+%22https%3A%2F%2Fd-nb.info%2Fgnd%2F4337175-9%22+OR+%22https%3A%2F%2Fd-nb.info%2Fgnd%2F4202451-1%22%29 (aufgerufen am 27. Mai 2022).

Suchtechnologien am Beispiel GND

Landesarchiv Baden-Württemberg
Zentrale Dienste

Ungenaue Namen mit OpenRefine abgleichen

https://fdmlab.landesarchiv-bw.de/workshop/openrefine-fortgeschrittene/15-erweiterter-gnd-abgleich-mit-lobid/

Name	Geb.datum	Geburtsort
Winlried Kretschnnann	1948-05-17	Spaichingen
Winfried Hemmann	1952	Rottenburg am Neckar
Manfred Lucka	1961-03-13	Garching a. d. Alz
Thehla Wallcer	1969	Dülmen
Tleresla Bauen	1965-04-06	Zweibrücken
Tlenesa Schoqqer	1961	Füssen
Danvai Bayas	1983	Heidelberg
Tlnomas Strolai	1960-03-17	Heilbronn
Nicoie Hollmeister-Kraul	1972	Balingen
Peler Haula	1960-12-24	Walldürn
Marlon Genlgez	1971	Haslach im Kinzigtal
Nicoie Rasayi	1965-05-20	Hongkong
Baaloara Bosch	1958-09-05	Stuttgart

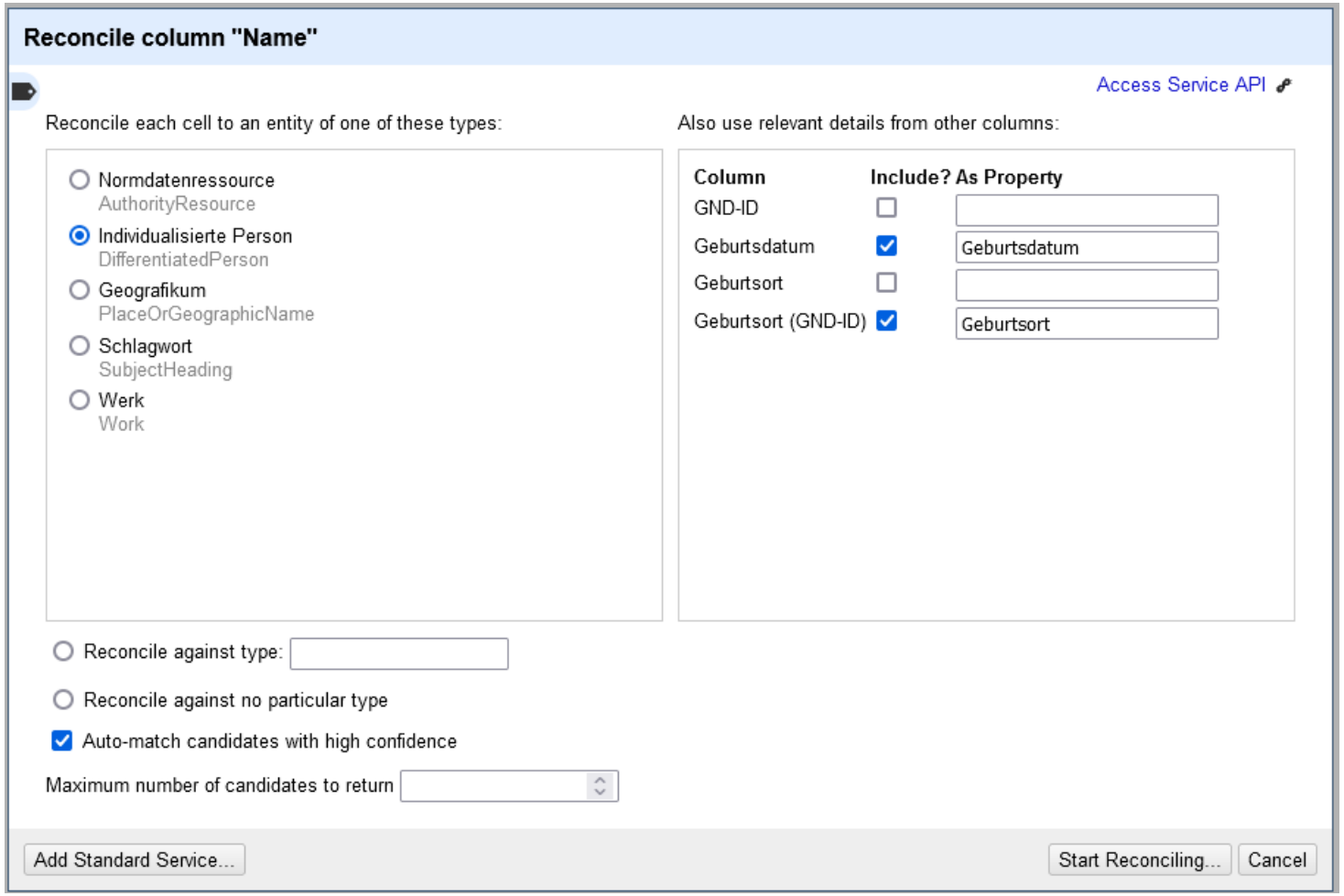
Kabinett Kretschmann III

Suchtechnologien am Beispiel GND

Landesarchiv Baden-Württemberg
Zentrale Dienste

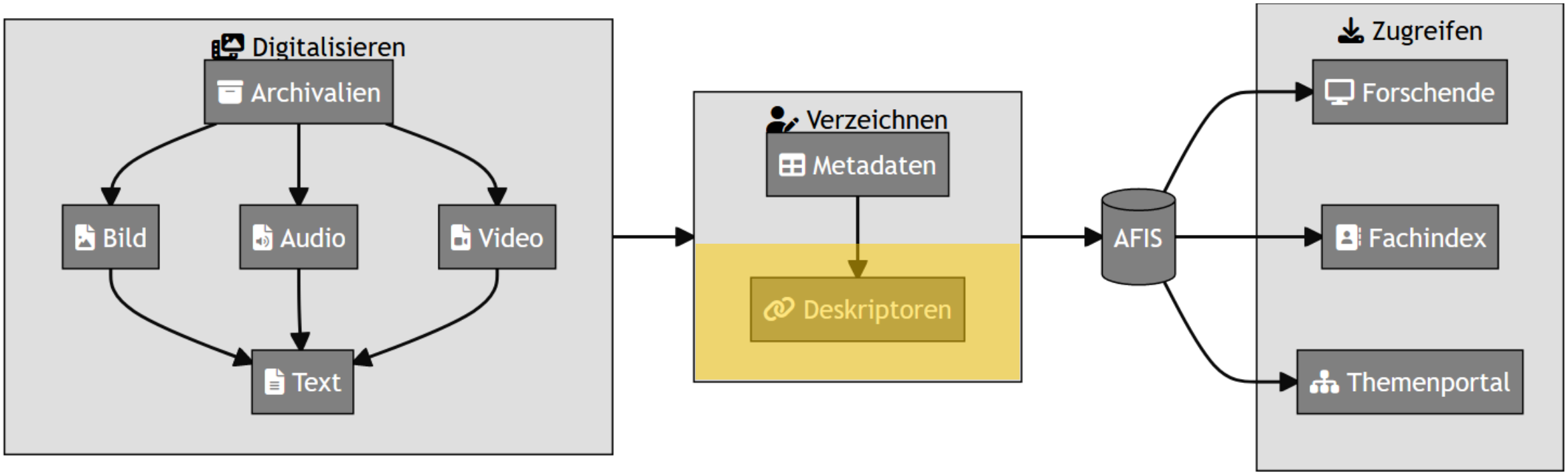
Ungenaue Namen mit OpenRefine abgleichen

https://fdmlab.landesarchiv-bw.de/workshop/openrefine-fortgeschrittene/15-erweiterter-gnd-abgleich-mit-lobid/



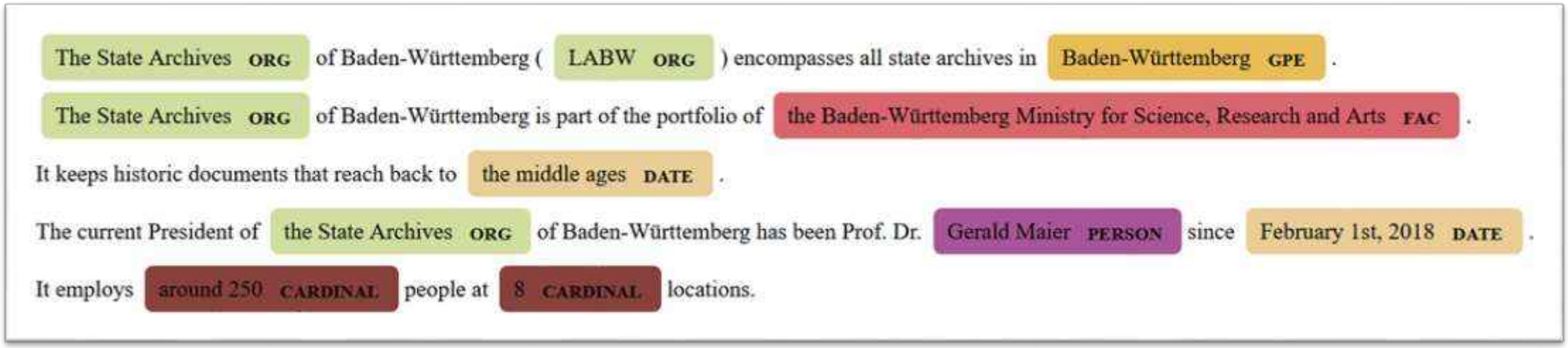
Workflow Digitalisierung

Landesarchiv Baden-Württemberg
Zentrale Dienste



Named-entity recognition: Englisch

Landesarchiv Baden-Württemberg
Zentrale Dienste



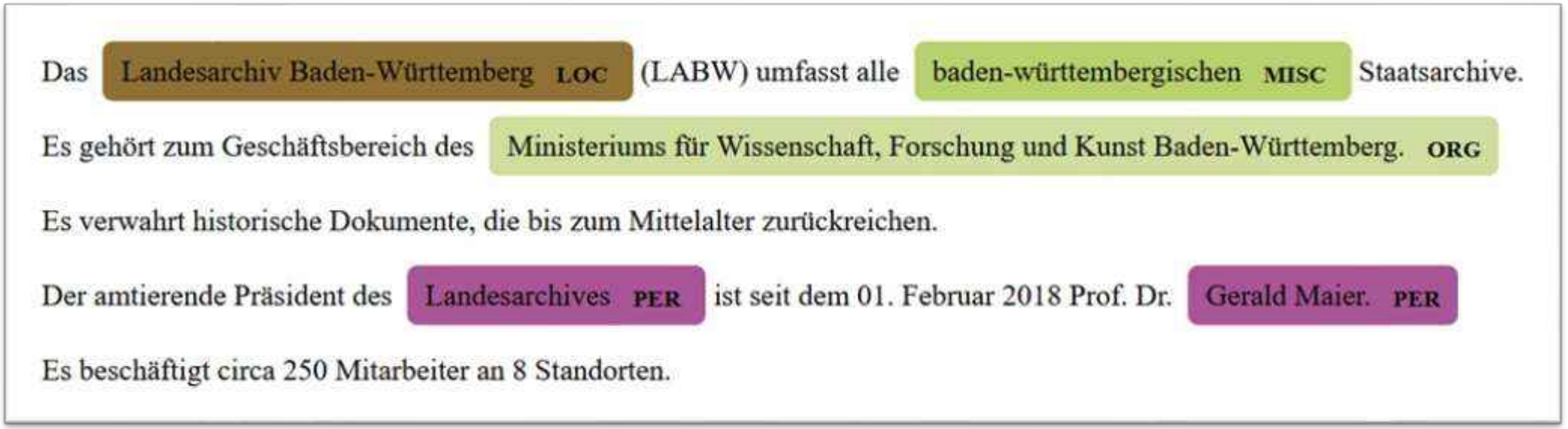
Das Model *en_core_web_lg* hat einen **f-score** von **0,85** und unterscheidet **18** Arten von **Entitäten**: CARDINAL, DATE, EVENT, FAC, GPE, LANGUAGE, LAW, LOC, MONEY, NORP, ORDINAL, ORG, PERCENT, PERSON, PRODUCT, QUANTITY, TIME und WORK_OF_ART.

Quellen:

- Erstellt mit spaCy 3.3 en_core_web_lg (https://spacy.io/models/en#en_core_web_lg).
- Wikipedia Autoren, „Landesarchiv Baden-Württemberg“, Wikipedia – Die freie Enzyklopädie, https://de.wikipedia.org/w/index.php?title=Landesarchiv_Baden-W%C3%BCrttemberg&oldid=223510351 (aufgerufen am 20. Juli 2022).

Named-entity recognition: Deutsch

Landesarchiv Baden-Württemberg
Zentrale Dienste



Das Model *de_core_news_lg* hat einen **f-score** von **0,85** und unterscheidet **4** Arten von **Entitäten**: LOC, MISC, ORG, PER.

Quellen:

- Erstellt mit spaCy 3.3 de_core_news_lg (https://spacy.io/models/de#de_core_news_lg).
- Wikipedia Autoren, „Landesarchiv Baden-Württemberg“, Wikipedia – Die freie Enzyklopädie, https://de.wikipedia.org/w/index.php?title=Landesarchiv_Baden-W%C3%BCrttemberg&oldid=223510351 (aufgerufen am 20. Juli 2022).

Named-entity recognition: Deutsch im Archiv

Landesarchiv Baden-Württemberg
Zentrale Dienste

Besitzhinweis auf Mobiliar (hauptsächlich gekauft bei der Möbelfirma LOC Winfried Mueller, Stuttgart PER),

1 vollständigen Silberkasten und mindestens 4 Silber-Becher.

Zahlung von 11100 fl. samt rückständigen neunjährigen Zinsen in Höhe von 4999 fl. oder Einsetzung in das adelige Gut

Werenwag.

Gen. Forderung stammt aus Streit über Erbe des im November 1595 verst. K[C]aspar (von) Laubenberg. MISC

Entzug des ordentlichen Richters und Verstoß gegen Haigerlocher Evokationsprivileg PER durch Verschleppung

kläg. Sohns gen. "junger Lenz" nach Sigmaringen, wo gegen ihn Kriminalprozeß PER angestrengt wurde.

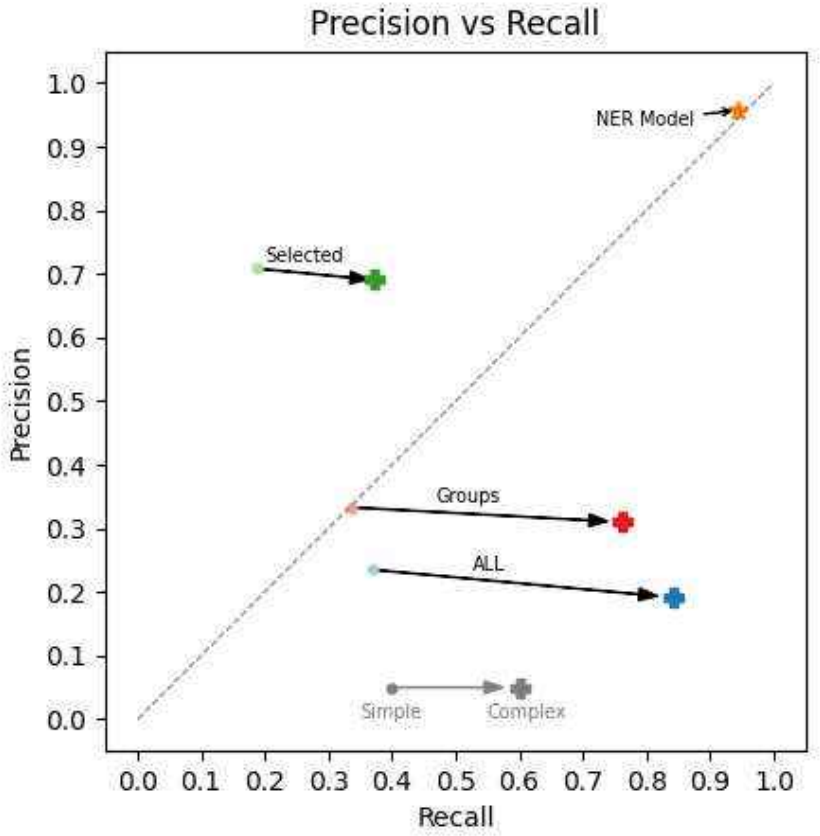
Quellen:
- Erstellt mit spaCy 3.3 de_core_news_lg (https://spacy.io/models/de#de_core_news_lg).

Named-entity recognition: Eigene Modelle?

Landesarchiv Baden-Württemberg
Zentrale Dienste

NER Modelle für Objekte

- Erkennung von Objekten
- Beispiel: Ludwigsburger Porzellantasse
- Starthilfe mit Sachbegriffen aus der GND
- Mit GND-Daten: **f-score 0,44-0,48**
- Aktives Lernverfahren
- Eigenes Modell: **f-score 0,95**



Landesarchiv
Baden-Württemberg
Zentrale Dienste

04

KI-Infrastruktur für Archive?

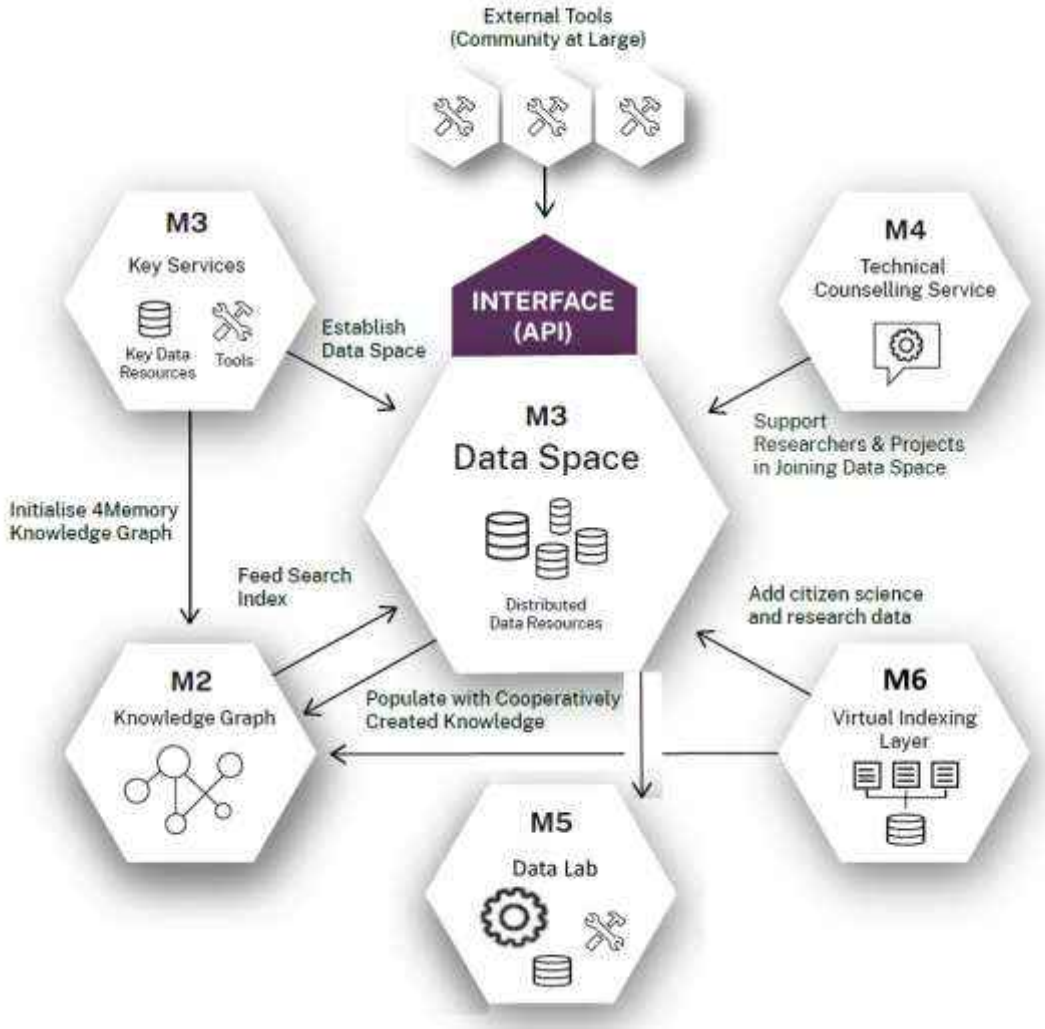
NFDI4Memory

Landesarchiv Baden-Württemberg
Zentrale Dienste

Task Area 3: Data Services

- **M5: Data Lab**
- High Performance Cluster
- Passende (online) Werkzeuge

https://4memory.de/



Quellen:
- Daniel Fähle, Harald Sack, "Perspektive „digitaler Werkzeugkasten“ für historische Forschung mit Archivgut" (Präsentation), 53. Deutscher Historikertag, München, 06. Oktober 2021.

Landesarchiv
Baden-Württemberg
Zentrale Dienste

Vielen Dank!

Benjamin Rosemann
Landesarchiv Baden-Württemberg
Zentrale Dienste
Projekt FDMLab@LABW
URL: https://fdmlab.landesarchiv-bw.de
Tel.: +49 711 335075-512
E-Mail: benjamin.rosemann@la-bw.de
https://orcid.org/0000-0002-0780-3979

www.landesarchiv-bw.de