# AP3 Verfahren für die Anbindung neuer Datenquellen





- Poster Architekturskizze
  - Quelldaten
  - Normalisierter Datenspeicher
  - Zieldaten
- Poster Workflow Daten
  - GND4C RDF API
- Poster Architekturskizze
  - Werkzeugkiste

# **Hauptaufgabe AP3**





- Implementierung und Evaluation eines prototypischen
   Matching- und Mapping-Verfahrens für die Entitäten der Fallbeispiele.
- Neue GND-relevante Datensätze und Übereinstimmungen zur GND finden und zukünftig in die GND einspielen, bzw. existierende GND-Datensätze sollen angereichert werden.

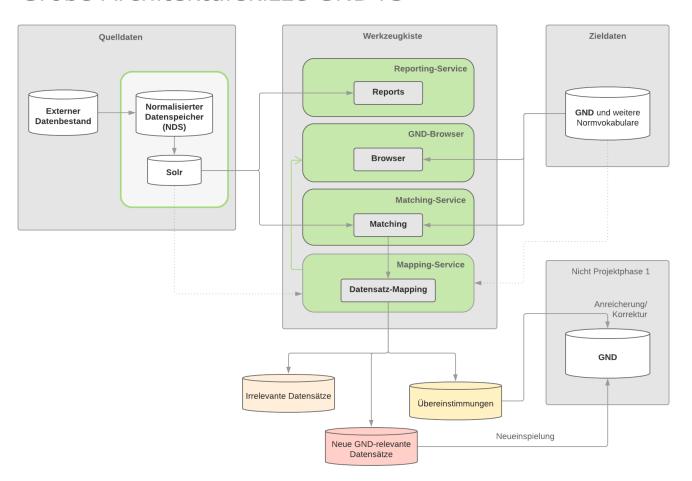
# AP3 Bedeutung "Matching" und "Mapping"



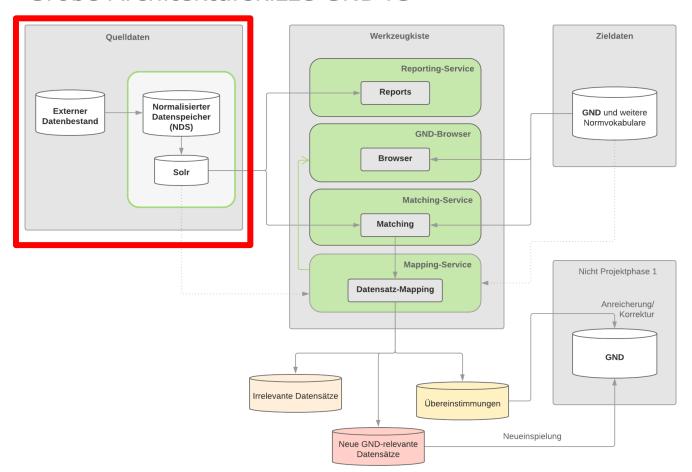


- Matching: automatischer Abgleich
  - Unterschiedliche Vokabulare oder Datenschemata werden auf übereinstimmende oder ähnliche Zeichenketten durchsucht
  - Spezielle Algorithmen und Konfiguration pro Entitätstyp
  - Matches sind Kandidaten f
    ür den Mapping-Prozess
- Mapping: intellektueller Abgleich
  - semantisch übereinstimmende Begriffe/Entitäten festlegen
  - Art der Übereinstimmung kann qualifiziert werden:
    - z.B. exakt oder annähernd übereinstimmend

#### Grobe Architekturskizze GND4C



#### Grobe Architekturskizze GND4C



# Lieferdaten Fallbeispiele





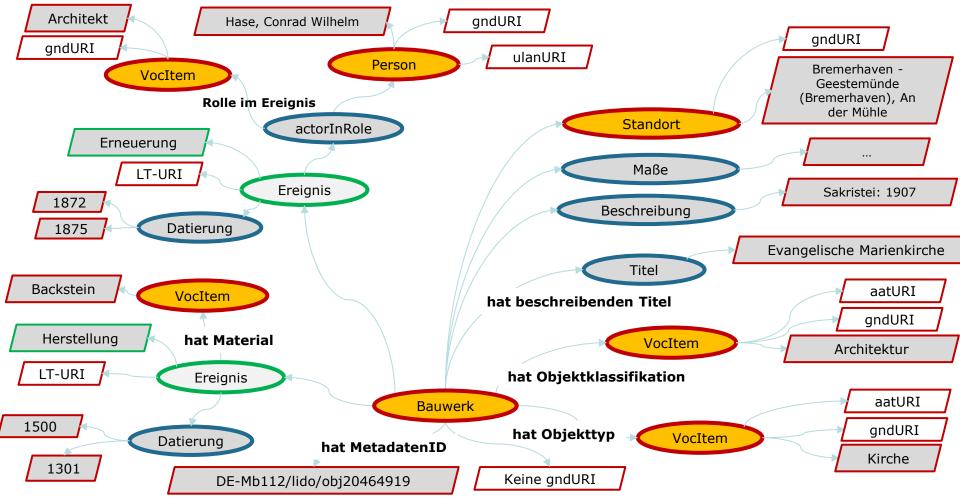
	Quelldaten GND4C					
Entität Datenlieferung		Anzahl DS	Quellformat lst	Quallformat Wunsch 2019		
Akteur	AKL, Imdas	40.462	CSV	XML (XSD-basiert)		
Sachgruppe	Deutsches Historische Museum DDK-Bildarchiv Foto Marburg	30.445 10.478	SKOS	SKOS		
Geographika	Hessisches Landesarchiv, Landesarchiv Baden-Württemberg,	208.338	EAD/CSV	?		
3.5 <b>3</b> .54	DDK-Bildarchiv Foto Marburg	50.631	HIDA	·		
Bauwerke	Landesamt für Denkmalpflege Bremen	1.791	LIDO	LIDO		
Daawono	DDK-Bildarchiv Foto Marburg	4.391	2.50	2.50		

Analyse Fallbeispiele: Beispiel Lieferdaten Bauwerke





# Vereinfachte LIDO-Teilstruktur Evangelische Marienkirche

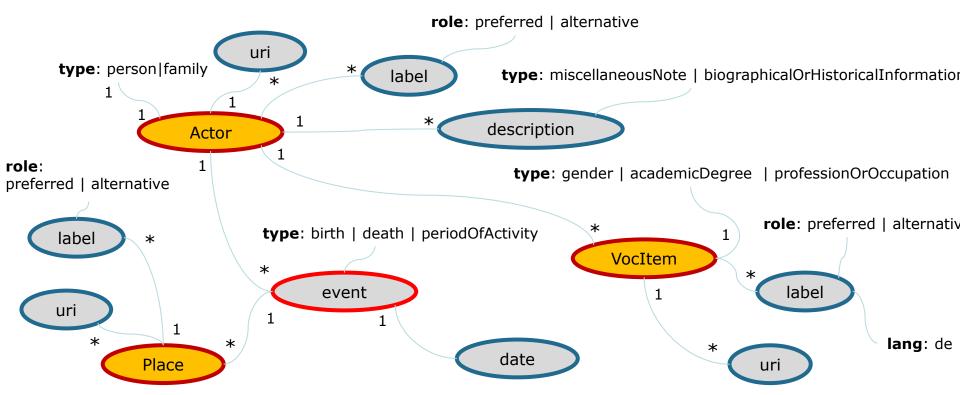


Analyse Fallbeispiele → einheitliches Modell für den normalisierten Datenspeicher für alle 4 Fallbeispiele Modell NDS für Personen:



Analyse Fallbeispiele → einheitliches Modell für den normalisierten Datenspeicher für alle 4 Fallbeispiele Modell NDS für Personen:





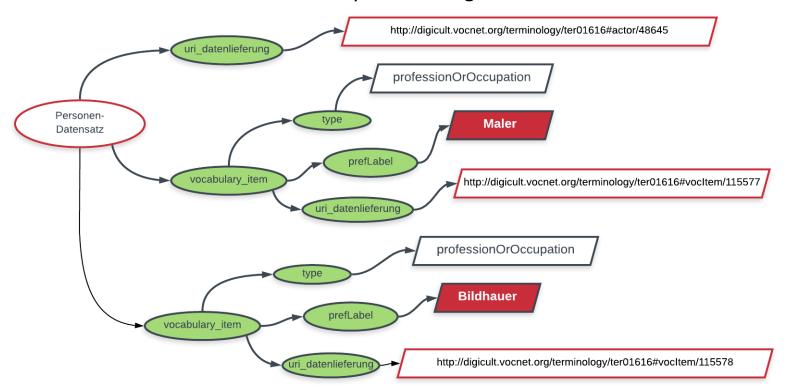
- Eindeutiger URI für Datenlieferung.
   Z.B. <a href="http://digicult.vocnet.org/terminology/ter01616">http://digicult.vocnet.org/terminology/ter01616</a> für Datenlieferung AKL
- Datenkorrektur Quelldaten über regelbasierte Algorithmen notwendig.
  - Z.B. "1762.10.18 / (vor) 1762.10.18" nach "../1762-10-18"
  - "13. April 1695" nach "1695-04-13"
  - "Nachname, Vorname (YYYY) nach "Nachname, Vorname"

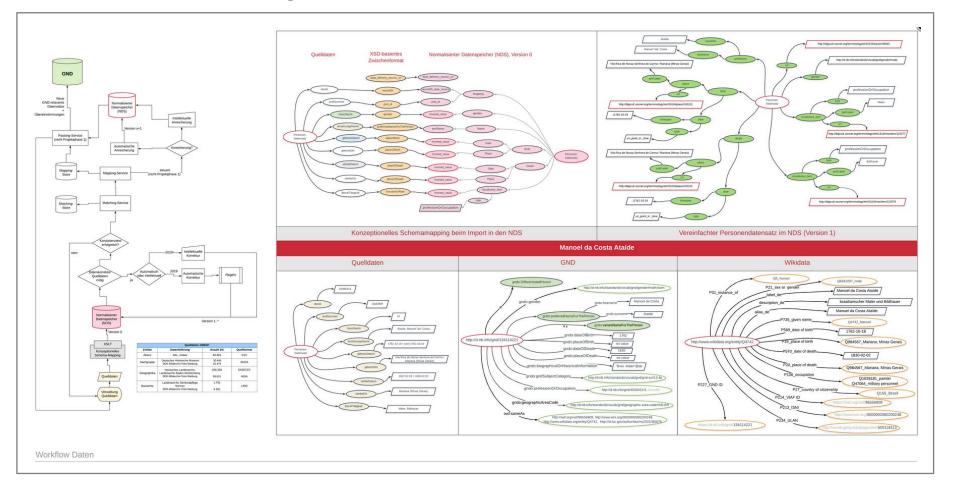


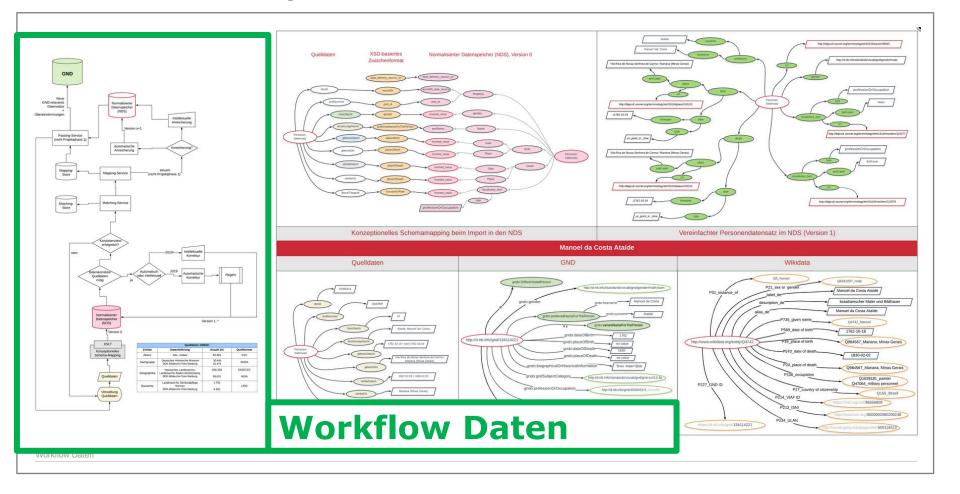
Normalisierung der Quelldaten
 Beispiel Lieferdaten Beruf/Tätigkeit in CSV: "Maler, Bildhauer"



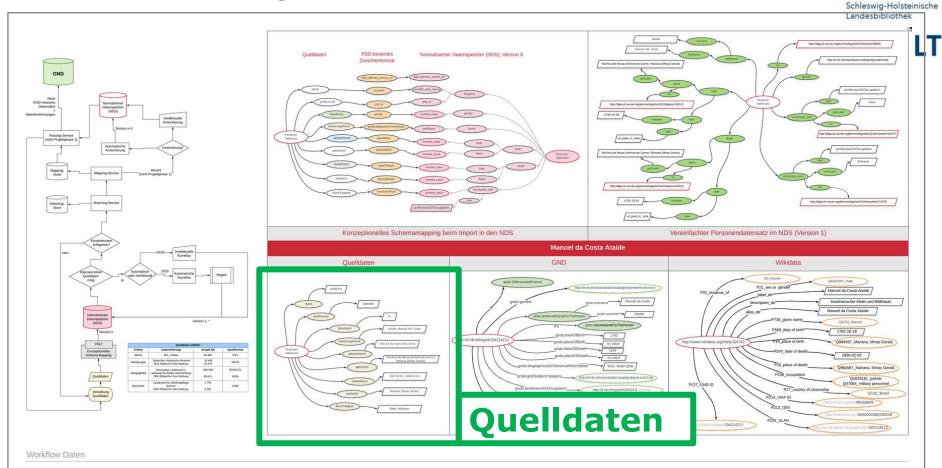
URIs für Werte im Prozess des kompletten Abgleichverfahrens

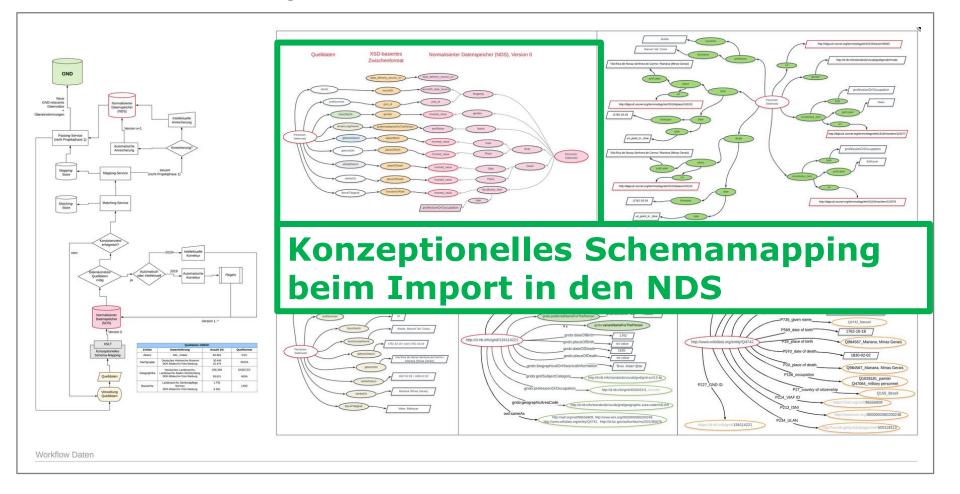


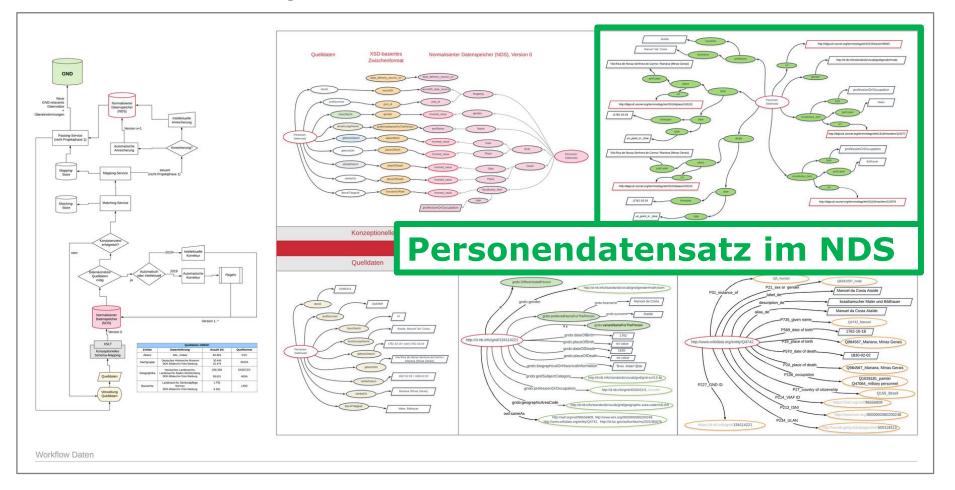


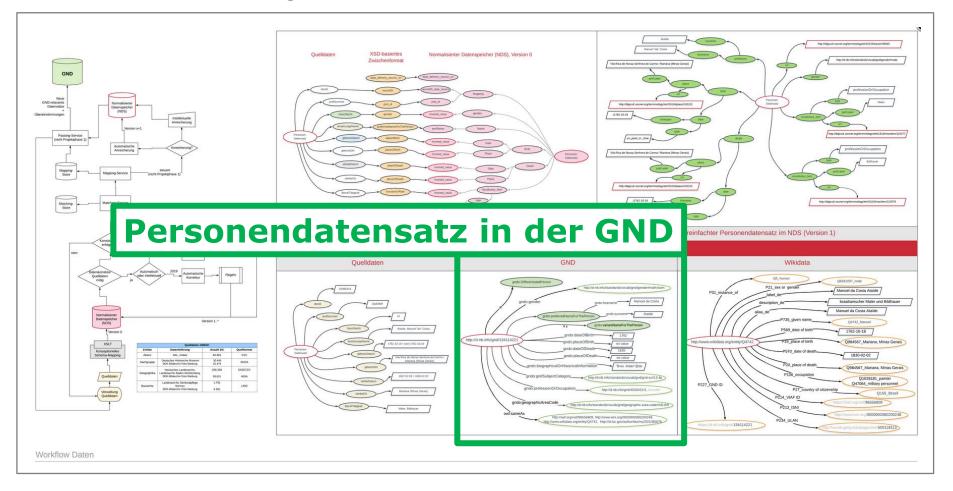


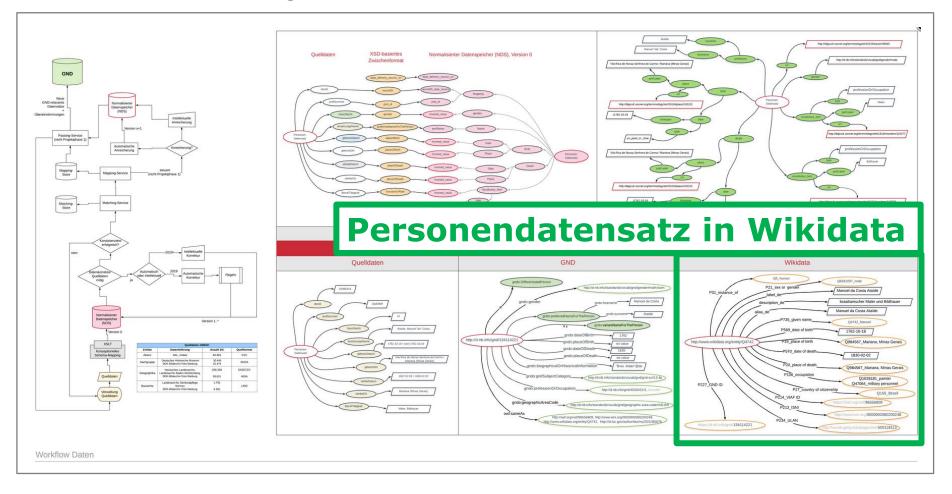




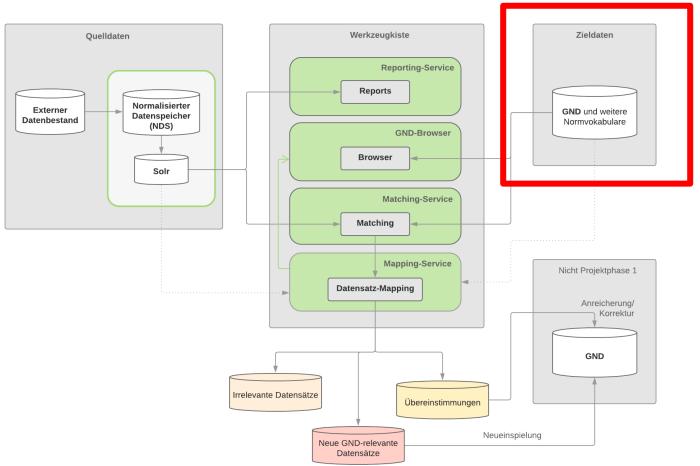


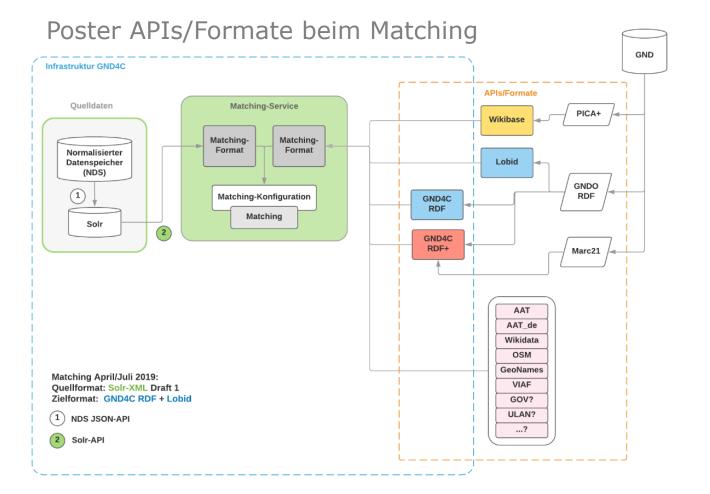






#### Grobe Architekturskizze GND4C

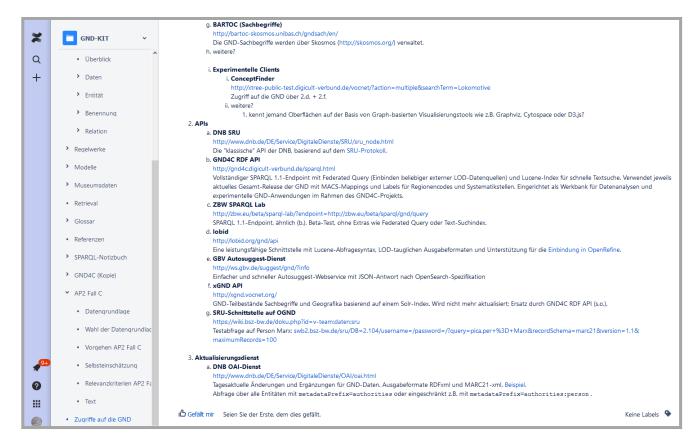




## Zugriffe auf die GND: Weboberflächen und APIs



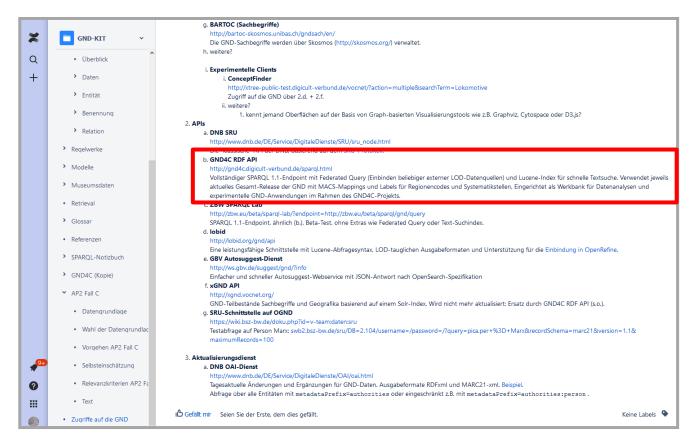




## Zugriffe auf die GND: Weboberflächen und APIs







# **GND4C-RDF API**



digiCULT



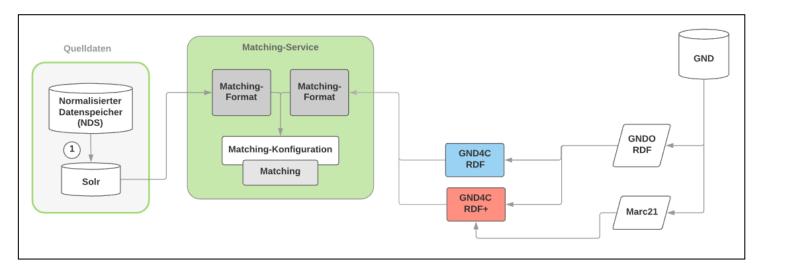
GND4C RDF-API. SPARQL-Endpoint:

221.600.210 Triples in Fuseki3 (RDF-Dump Stand Feb. 2019)

- Incl. MACS-Mappings
- In Umsetzung
  - Umstellung auf OAI

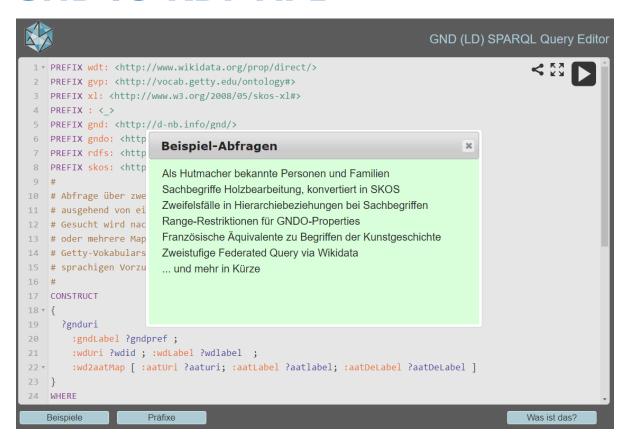


Erweiterung um einzelne Marc21-Elemente





## **GND4C-RDF API**



SH \*\*

Schleswig-Holsteinische
Landesbibliothek

digi CULT



http://gnd4c.digicult-verbund.de/spargl.html

# Homonymzusatz für "Bogen < Waffe>"





GND			
Link zu diesem Datensatz	http://d-nb.info/gnd/41462	11-7	
Sachbegriff	Bogen <waffe></waffe>		
Quelle	B 1996 3.		
Oberbegriffe	Fernwaffe Bogen <gerät></gerät>		
Thematischer Bezug	Verwandter Begriff: Pfeil		
DDC-Notation	623.441 688.7920285 739.73 799.20285	Vorzugebezeichnung in der I preferredNameForTheSubjed	<u> </u>
Systematik	8.4 Militär	· · ·	
Тур	Allgemeinbegriff (saz)		
Andere Normdaten	LCSH: Bow and arrow RAMEAU: Arcs		

# Homonymzusatz für "Bogen < Waffe>"





GND			
Link zu diesem Datensatz	http://d-nb.info/gnd/41462	11-7	
Sachbegriff	Bogen <waffe></waffe>		
Quelle	B 1996 3.		
Oberbegriffe	Fernwaffe Bogen <gerät></gerät>		
Thematischer Bezug	Verwandter Begriff: Pfeil		
DDC-Notation	623.441 688.7920285 739.73 799.20285	"Bogen <waffe>" in Marc21 <datafield ind<br="" tag="150"><subfield <="" code="a" th=""><th>&gt;Bogen</th></subfield></datafield></waffe>	>Bogen
Systematik	8.4 Militär	<subfield code="g">Waffe</subfield>	
Тур	Allgemeinbegriff (saz)		
Andere Normdaten	LCSH: Bow and arrow RAMEAU: Arcs		

# Homonymzusatz für "Bogen < Waffe>"





GND		
Link zu diesem Datensatz	http://d-nb.info/gnd/4146211-7	
Sachbegriff	Bogen <waffe></waffe>	
Quelle	B 1996 3.	

Verbesserte Abfragemöglichkeiten über Homonymzusätze beim Matching durch Hinzufügen von zusätzlichem Namespace aus SKOS-XL und vocnet: gnd:4146211-7 a gndo:SubjectHeadingSensoStricto; skosxl:prefLabel [ skosxl:literalValue "Bogen < Waffe>"; vocnet:hasAsLabel "Bogen"; vocnet:hasQualifier "Waffe" ];

RAMEAU: Arcs

# **GND4C RDF-API**

SH 🗱 Schleswig-Holsteinische Landesbibliothek

digiCULT

- Beispiele:
  - SPARAL-Notizbuch Detlev Balzer:

https://digicult.atlassian.net/wiki/spaces/GNDKIT/pages/480149789/SPAR QL-Notizbuch

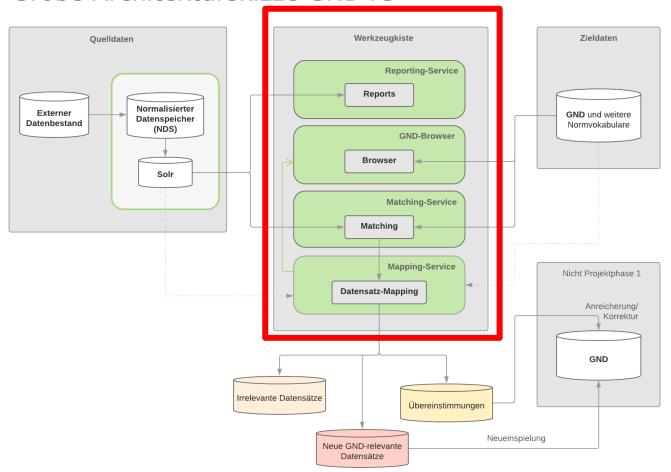
 Damerau-Levenshtein-Distanz unter Verwendung von Lucene-Suchindex

Kutepow, Kutepov, uktepov, ... findet Kutepov

 Namen sämtlicher der GND bekannten Königinnen und Könige von Polen: <a href="https://digicult.atlassian.net/wiki/spaces/GNDKIT/pages/497778691/Text-Retrievalkomponente">https://digicult.atlassian.net/wiki/spaces/GNDKIT/pages/497778691/Text-Retrievalkomponente</a>

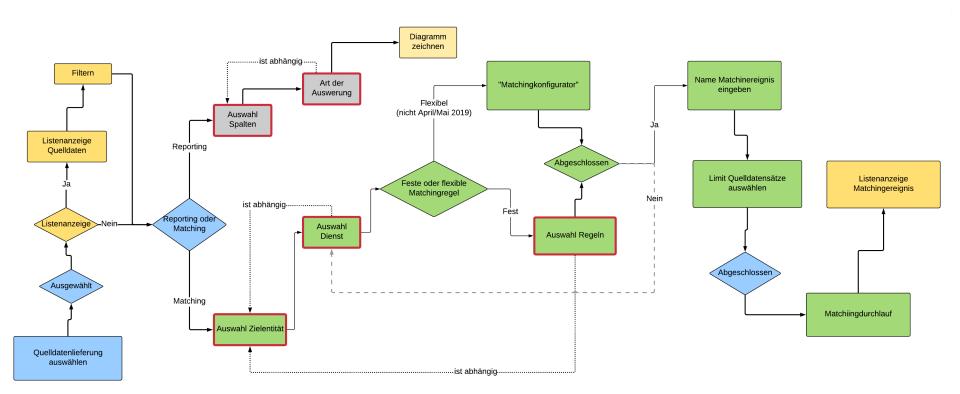


#### Grobe Architekturskizze GND4C



# Werkzeugkiste (Umsetzung seit 15.03.2019)







#### Wählen Sie eine Datenlieferung für den Analysetest aus

	Matching	Reporting	Matchingereignisse			
Nr	Datenlieferung	Import NDS	URI Datenlieferung	Anzahl DS	URI ISIL Museum	Anmerkung
7	Testdatenset Persone	en 05.05.2019	http://digicult.vocnet.org/terminology/ter017	31 7		
1	imdas pro-Datenbank Staatliches Museum f Naturkunde Karlsruhe	für	http://digicult.vocnet.org/terminology/ter016	21 23	noch nicht vergeben	
6	AKL	19.03.2019	http://digicult.vocnet.org/terminology/ter016	16 12.683		Datenlieferung 17.05.2018
5	imdas pro-Datenbank Museum Stuttgart, Staatliches Museum f Völkerkunde		http://digicult.vocnet.org/terminology/ter016	17 334	http://ld.zdb- services.de/resource/organisations/DE- MUS-129413	
4	imdas pro-Datenbank Schloss Friedenstein (		http://digicult.vocnet.org/terminology/ter016	18 6.054	http://ld.zdb- services.de/resource/organisations/DE- G16	
3	imdas pro-Datenbank Kunsthalle Karlsruhe	: / Staatliche 19.03.2019	http://digicult.vocnet.org/terminology/ter016	19 10.021	http://ld.zdb- services.de/resource/organisations/DE- MUS-072515	keine recordID  → temporäre in Quelldaten angelegt
2	imdas pro-Datenbank Städtische Museen Fr		http://digicult.vocnet.org/terminology/ter016	20 11.347	http://ld.zdb- services.de/resource/organisations/DE- MUS-048413	



#### Wählen Sie eine Datenlieferung für den Analysetest aus

Matching		Rep	orting	Matchingereignisse			
Nr	Datenlieferung		Import NDS	URI Datenlieferung	Anzahl DS	URI ISIL Museum	Anmerkung
7	Testdatenset Persone	:n	05.05.2019	http://digicult.vocnet.org/terminology/ter01731	7		
	imdas pro-Datenbank Staatliches Museum f Naturkunde Karlsruhe	ür	19.03.2019	http://digicult.vocnet.org/terminology/ter01621	23	noch nicht vergeben	
6	AKL		19.03.2019	http://digicult.vocnet.org/terminology/ter01616	12.683		Datenlieferung 17.05.2018
5	imdas pro-Datenbank Museum Stuttgart, Staatliches Museum f Völkerkunde		19.03.2019	http://digicult.vocnet.org/terminology/ter01617	334	http://ld.zdb- services.de/resource/organisations/DE- MUS-129413	
4	imdas pro-Datenbank Schloss Friedenstein (		19.03.2019	http://digicult.vocnet.org/terminology/ter01618	6.054	http://ld.zdb- services.de/resource/organisations/DE- G16	
3	imdas pro-Datenbank Kunsthalle Karlsruhe	/ Staatliche	19.03.2019	http://digicult.vocnet.org/terminology/ter01619	10.021	http://ld.zdb- services.de/resource/organisations/DE- MUS-072515	keine recordID  → temporäre in Quelldaten angelegt
2	imdas pro-Datenbank Städtische Museen Fr		19.03.2019	http://digicult.vocnet.org/terminology/ter01620	11.347	http://ld.zdb- services.de/resource/organisations/DE- MUS-048413	



BCP: > Datenlieferung: Testdatenset Personen 0

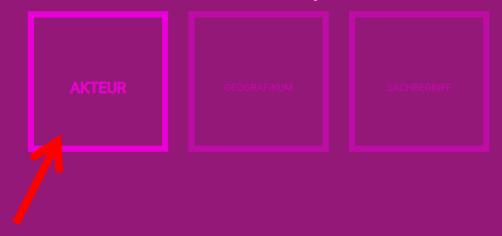
#### Wählen Sie die Entität für den Analysetest aus





BCP: > Datenlieferung: Testdatenset Personen 0

#### Wählen Sie die Entität für den Analysetest aus





BCP: > Datenlieferung: Testdatenset Personen 0 > Zielentität: Personen

#### Auswahl der Zielsysteme

(1.) GND4C RDF API
1. Lobid



BCP: > Datenlieferung: Testdatenset Personen 0 > Zielentität: Personen

## Auswahl der Zielsysteme





BCP: > Datenlieferung: Testdatenset Personen 0 > Zielentität: Personen > Zielsystem: Lobid

#### Matchingevent definieren

Bitte geben Sie Ihrem Event einen Namen

Event speichern



BCP: > Datenlieferung: Testdatenset Personen 0 > Zielentität: Personen > Zielsystem: Lobid

#### Matchingevent definieren





Limit		
	1	
	5	
	Alle Datensätze (7)	



#### Limit





ID NDS	Vorname	Nachname	Score/Loops	Details
79106	Manuel da Costa	Ataíde	6.3375 / 1	anzeigen (1)
79107	Heinz	Becker	8.1875 / 1	anzeigen (1)
79108	Adolphe Charles	Adam	2.4875 / 1	anzeigen (1)
79109	Giovanni	Abondio	0.875 / 4	anzeigen (1)
79110		Johann Friedrich I, Sachsen, Kurfürst	5.8 / 1	anzeigen (1)



ID NDS	Vorname	Nachname	Score/Loops	Details
79106	Manuel da Costa	Ataíde	6.3375 / 1	anzoigen (1)
79107	Heinz	Becker	8.1875 / 1	igen (1)
79108	Adolphe Charles	Adam	2.4875 / 1	anzeigen (1)
79109	Giovanni	Abondio	0.875 / 4	anzeigen (1)
79110		Johann Friedrich I, Sachsen, Kurfürst	5.8 / 1	anzeigen (1)



ID NDS	Vorname	Nachname	Score/Loops	Details
79106	Manuel da Costa	Ataide	6.3375 / 1	anzeigen (1)
Quelldaten		Zieldaten(1)		
	uel da Costa íde	URI ins Zielsystem: http://d-nb.info/gnd/134114221 Zielsystem: GND Vorname: Manoel da Costa Nachname: Ataide Geburtsdatum: 1762 Sterbedatum: 1830 Entityfacts: + Vollständigkeit der Zieldaten Score: 8.1875		
79107	Heinz	Becker	8.1875 / 1	anzeigen (1)
79108	Adolphe Charles	Adam	2.4875 / 1	anzeigen (1)
79109	Giovanni	Abondio	0.875 / 4	anzeigen (1)
79110		Johann Friedrich I, Sachsen, Kurfürst	5.8 / 1	anzeigen (1)



ID NDS	Vorname	Nachname	Score/Loops	Details
79106	Manuel da Costa	Ataíde	6.3375 / 1	anzeigen (1)
Quelldaten		Zieldaten(1)		
	uel da Costa aíde	URI ins Zielsystem: http://d-nb.info/gnd/134114221 Zielsystem: GND Vorname: Manoel da Costa Nachname: Ataíde Geburtsdatum: 1762 Sterbedatum: 1830 Entityfacts: +  Vollständigkeit der Zieldaten Score: 8.1875		
79107	Heinz	Becker	8.1875 / 1	an eigen (1)
79108	Adolphe Charles	Adam	2.4875 / 1	an eigen (1)
79109	Giovanni	Abondio	0.875 / 4	an eigen (1)
79110		Johann Friedrich I, Sachsen, Kurfürst	5.8 / 1	an eigen (1)





Rätsel: wieviel Farben hat unsere erste Score-Ampel?











Rätsel: wieviel Farben hat unsere erste Score-Ampel?







Auflösung: Kleines Sitzungszimmer, 14:45-15:30 Uhr



# digiCULT



<end />







