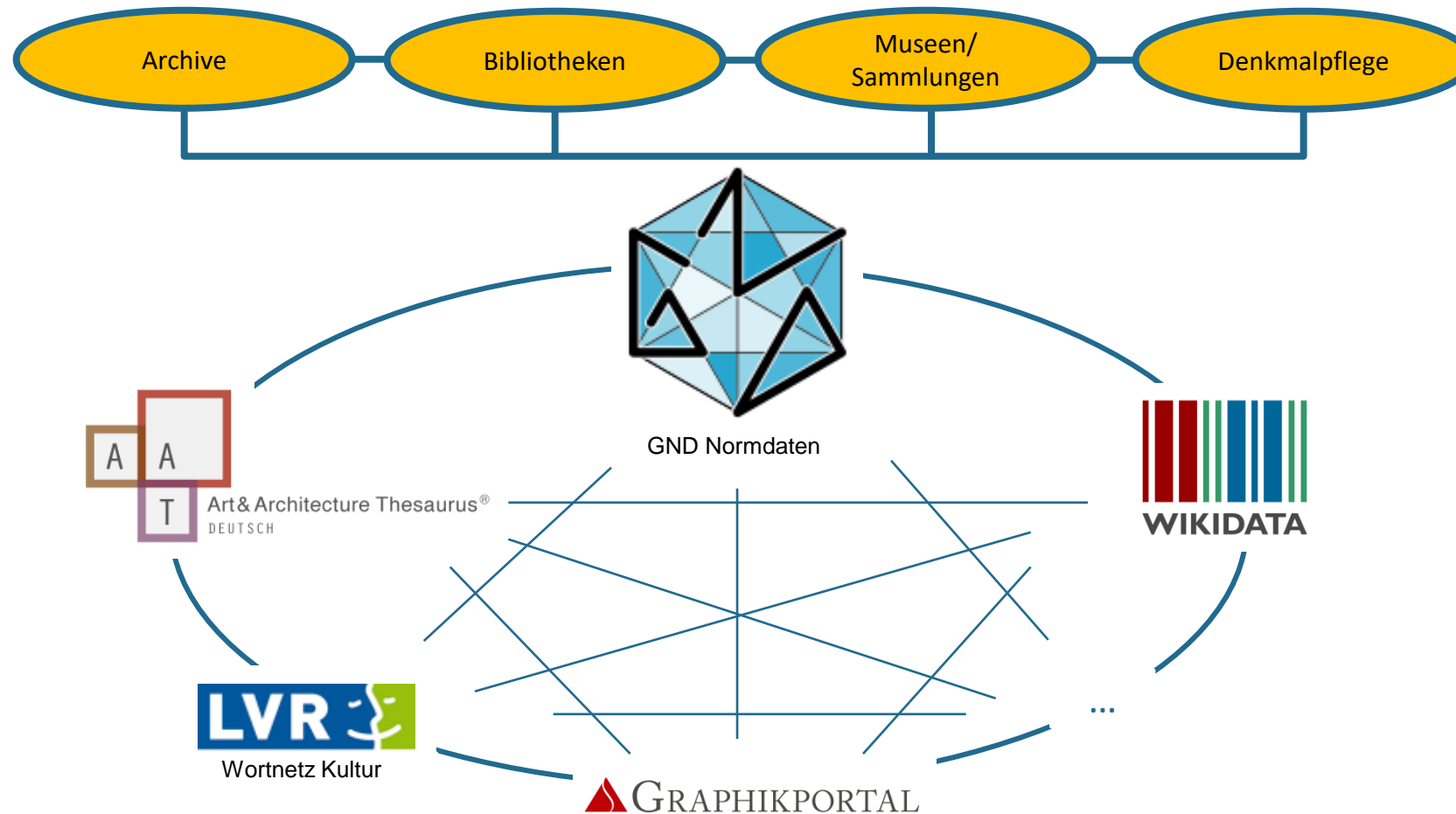


- GND4C – GND für Kulturdaten
- GND4C Werkzeugkiste
 - Matching/Mapping
 - Grober Workflow/Datenworkflow
 - Features
 - Oberfläche Version 1
- GND4C RDF API

GND als Brücke zwischen den Kultursparten



Erweiterung der GND und Vernetzung der GND mit anderen Normvokabularen

Verlässlichkeit Verlässlichkeit Verlässlichkeit Verlässlichkeit Verlässlichkeit Verlässlichkeit Verlässlichkeit Verlässlichkeit



BRÜCKEN IM NETZWERK DER KULTUR
UND WISSENSCHAFT

www.dnb.de/gnd



- Zwei Projektphasen geplant:
 - Phase 1:
 - Mai 2018 – Okt 2020
 - Konzeption, Vorbereitung, teilweiser Pilotbetrieb
 - Umsetzung Werkzeugkiste: April 2019 - Ende 2020
 - Phase 2:
 - Voraussichtlich Beginn 2021 – Mitte 2023
 - Produktivnahme und Erweiterung





- Implementierung und Evaluation eines prototypischen **Matching**- und **Mapping**-Verfahrens für Entitätsdatensätze (Personen, Geografika, Sachbegriffe und Bauwerke).
- Neue GND-relevante Datensätze und Übereinstimmungen zur GND finden und zukünftig in die GND einspielen, bzw. anreichern von existierenden GND-Datensätzen.
- GND-URIs an Erfassungssysteme liefern.



– Matching: automatischer Abgleich

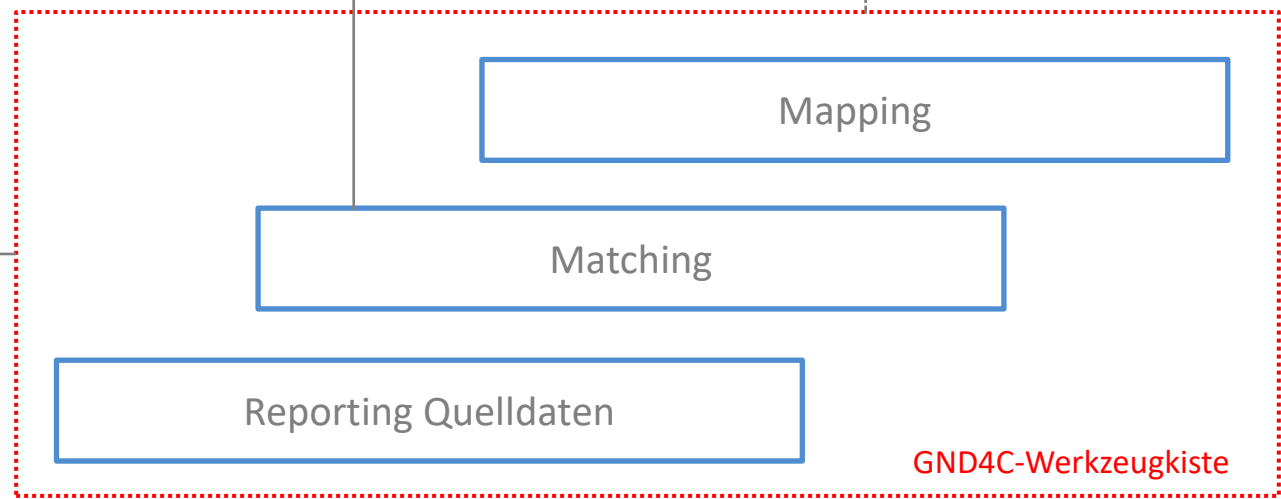
- Unterschiedliche Vokabulare oder Datenschemata werden auf übereinstimmende oder ähnliche Zeichenketten durchsucht
- Spezielle Algorithmen und Konfiguration pro Entitätstyp
- Matches sind Kandidaten für den Mapping-Prozess

– Mapping: intellektueller Abgleich

- semantisch übereinstimmende Begriffe/Entitäten festlegen
- Art der Übereinstimmung kann qualifiziert werden



APIs Normdatensysteme:
u.a. GND4C-RDF (GND),
lobid (GND), Wikidata,
GeoNames, OSM, AAT,
GOV, TGN, ...



Normalisierter Datenspeicher (NDS)

Quelldaten

Personen, Geographika,
Bauwerke, Sachbegriffe

Erfassungssysteme

Grober Workflow

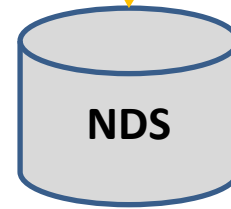
Lieferdaten Fallbeispiele

APIs Normdatensysteme:

u.a. GND4C-RDF (GND),
lobid (GND), Wikidata,
GeoNames, OSM, AAT,
GOV, TGN, ...



2021



Normalisierter Datenspeicher

Schema-Mapping > XSLT

XML GND4C

CSV Niedersachsen

SKOS

LIDO XML

CSV Muisis

CSV AKL

EAD/CSV

HIDA DDK

LIDO XML (DHM)

LIDO XML

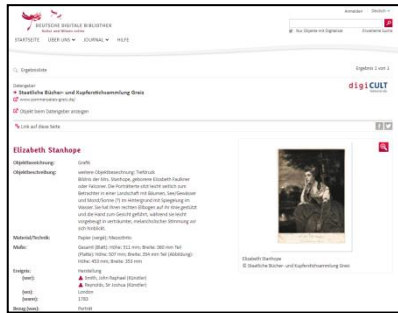
Personen

Geografika

Sachbegriffe

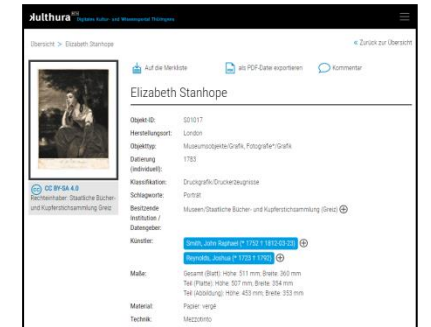
Bauwerke

Fallbeispiele GND4C



Datensatz nur im
DDB-Testsystem verfügbar

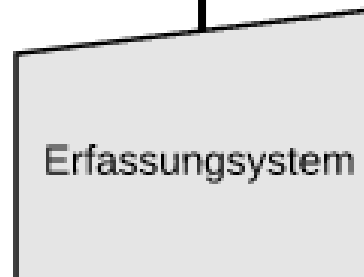
Person mit GND-URI



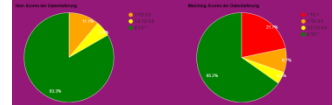
https://kuwitueringen.de/de/suche/p/1/h/1/nc/1.html?tx_jomuseo_pi1009%5BjoDetailView%5D=DE-MUS-871812%2Fflido%2Fdc00001163



LIDO-Datensatz ohne GND-URI zu Person



- Konfigurierbares und mehrstufiges Matchingverfahren auf unterschiedliche Normdatensysteme
- Matchingdurchläufe werden als Matchingereignisse gespeichert
- Unterschiedliche Suchstrategien pro Entitätstyp und Datenset
- Regeln und Berechnung für Konfidenzwerte als Scoring. Darstellung über Farbampel
- Reporting von Quelldaten zur Unterstützung der Qualitätssicherung
- Solr-basierte Volltextsuche über Quelldaten und Matchingereignisse
- Erweiterte Algorithmen über Fuzzy-Search, Stemming, Levenshtein-Distanz
- Unterschiedliche Wertelisten für „Art der Übereinstimmung“ beim Mapping
- Versionierung im normalisierten Datenspeicher (NDS)



Wählen Sie eine Datenlieferung für den Analysetest aus

Datenauswahl		Reporting	Matchingereignisse	Anzahl	Entität	URI ISIL	Anmerkung
Nr	Datenlieferung	Import NDS	URI Datenlieferung	DS		Museum	
12	Personen (Numismatischer Bestand) - Landesmuseum Württemberg		http://digicult.vocnet.org/terminology/ter01873		Person		Numismatische Datensätze
11	Sachbegriffe - DDK	2019-09-16 21:07:25	http://digicult.vocnet.org/terminology/ter01798	10869	Sachbegriff		
10	Geografika Niedersachsen - Niedersächsisches Landesarchiv	2019-08-14 09:05:10	http://digicult.vocnet.org/terminology/ter01788	17044	Geografika		
9	Testdatenset Personen - digiCULT	2019-10-13 13:20:06	http://digicult.vocnet.org/terminology/ter01731	7	Person		
8	Personen, imdas pro-Datenbank - Staatliches Museum für Naturkunde Karlsruhe	2019-03-19 16:17:08	http://digicult.vocnet.org/terminology/ter01621	23	Person		
7	Personen, imdas pro-Datenbank - Linden-Museum Stuttgart, Staatliches Museum für Völkerkunde	2019-03-19 16:35:38	http://digicult.vocnet.org/terminology/ter01617	334	Person		
6	Testdatenset Personen - AKL	2019-03-19 16:19:20	http://digicult.vocnet.org/terminology/ter01616	12683	Person		
5	Personen, imdas pro-Datenbank - Stiftung Schloss Friedenstein Gotha	2019-03-19 16:35:38	http://digicult.vocnet.org/terminology/ter01618	6054	Person		

BCP: > Datenlieferung: Testdatenset Personen 0

Wählen Sie die Entität für den Analysetest aus

AKTEUR

GEOGRAFIKUM

SACHBEGRIFF



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

Auswahl der Zielsysteme

(1.) GND4C RDF API

1. Lobid





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

Matchingevent definieren

test_as_1

Event speichern



BCP: > Datenlieferung: Testdatenset Personen 0 > Zielentität: Personen > Zielsystem: Lobid > Matchingevent: test_as_1








Limit

1
5
Alle Datensätze (7)



BCP: > Matchingevent: test_as_1

Matchingevent

Matching		Reporting			
ID NDS	Vorname	Nachname	Score/Loops	Normdatensystem	Details
79106	Joshua	Reynolds	8.1875 / 1	 GND	anzeigen (1)
79107	John Raphael	Smith	2.1875 / 1	 GND	anzeigen (1)
79108	Adolphe Charles	Adam	2.1875 / 1	 GND	anzeigen (1)
79109	Giovanni	Abondio	0.875 / 4	 GND	anzeigen (1)
79110		Johann Friedrich I, Sachsen, Kurfürst	5.2 / 1	 GND	anzeigen (1)
79111		Johann Georg von Brandenburg	0 / 2	 GND	anzeigen (2)
79112	Alwine	Stapff	1.6375 / 5	 Wikidata	anzeigen (1)

BCP: > Matchingevent: test_as_1

Matchingevent

Matching		Reporting			
ID NDS	Vorname	Nachname	Score/Loops	Normdatensystem	Details
79106	Joshua	Reynolds	8.1875 / 1	 GND	anzeigen (1)
79107	John Raphael	Smith	2.1875 / 1	 GND	anzeigen (1)
79108	Adolphe Charles	Adam	2.1875 / 1	 GND	anzeigen (1)
79109	Giovanni	Abondio	0.875 / 4	 GND	anzeigen (1)
79110		Johann Friedrich I, Sachsen, Kurfürst	5.2 / 1	 GND	anzeigen (1)
79111		Johann Georg von Brandenburg	0 / 2	 GND	anzeigen (2)
79112	Alwine	Stapff	1.6375 / 5	 Wikidata	anzeigen (1)

BCP: > Matchingevent: test_as_1

Matchingevent

Matching		Reporting			
ID NDS	Vorname	Nachname	Score/Loops	Normdatensystem	Details
79106	Joshua	Reynolds	8.1875 / 1	■ GND	anzeigen (1)

Quelldaten

ID im Quellsystem: 79106
 Quellsystem: NDS
 Entitätstyp:
 Vorname: Joshua Benennung:
 Nachname: Reynolds
 Geburtsdatum: 1723
 Sterbedatum: 1792

Vollständigkeit der Quelldaten

Score: 8.1875

Anmerkungen und Kommentare(0)

[Einen neuen Kommentar verfassen](#)

Zieldaten(1)

URI ins Zielsystem: http://d-nb.info/gnd/118744771
 Zielsystem: GND
 Entitätstyp:
 Vorname: Joshua Benennung:
 Nachname: Reynolds
 Geburtsdatum: 1723
 Sterbedatum: 1792
 Entityfacts: [+](#)

Vollständigkeit der Zieldaten

Score: 8.1875

Matching

ID NDS	Benennung	Bundesland	Score/Loops	Normdatensystem	Details
181189	Aaltukerei	Niedersachsen	0 / 0	OSM	anzeigen (1)
181190	Aarbach	Niedersachsen	0 / 0	OSM	anzeigen (1)
181191	Aashausen	Niedersachsen	0 / 0	geonames	anzeigen (1)
181192	Abbau	Niedersachsen	0 / 0	geonames	anzeigen (1)
181193	Abbecke	Niedersachsen	0 / 0	geonames	anzeigen (1)
181194	Abbecker Straße	Niedersachsen	0 / 0	OSM	anzeigen (1)
181195	Abbehausen	Niedersachsen	0 / 0	GND	anzeigen (1)
181196	Abbehauser Altendeich	Niedersachsen	0 / 0	geonames	anzeigen (1)
181197	Abbehauser Hörne	Niedersachsen	0 / 0	geonames	anzeigen (1)
181198	Abbehauser Wehl	Niedersachsen	0 / 0	geonames	anzeigen (1)



Matching

ID NDS	Benennung	Bundesland	Score/Loops	Normdatensystem	Details
181189	Aaltukerei	Niedersachsen	0 / 0	OSM	anzeigen (1)
181190	Aarbach	Niedersachsen	0 / 0	OSM	anzeigen (1)
181191	Aashausen	Niedersachsen	0 / 0	geonames	anzeigen (1)
181192	Abbau	Niedersachsen	0 / 0	geonames	anzeigen (1)
181193	Abbecke	Niedersachsen	0 / 0	geonames	anzeigen (1)
181194	Abbecker Straße	Niedersachsen	0 / 0	OSM	anzeigen (1)
181195	Abbehausen	Niedersachsen	0 / 0	GND	anzeigen (1)

Quelldaten

ID im Quellsystem: 181195
Quellsystem: NDS
Entitätstyp: Geografika
Benennung: Abbehausen
Bundesland: Niedersachsen
Kreis: Wesermarsch
Gemeinde: Nordenham
Latitude: 53.484071700286
Longitude: 8.4355385576146

Vollständigkeit der Quelldaten

Score: 0

Zieldaten(1)

URI ins Zielsystem: http://d-nb.info/gnd/4773939-3
Zielsystem: GND
Entitätstyp: Geografika
Subtyp: TerritorialCorporateBodyOrAdministrativeUnit, PlaceOrGeographicName, AuthorityResource
Benennung: Abbehausen
Bundesland: Niedersachsen
Latitude: 53.48431
Longitude: 8.438449
Entfernung (Luftlinie) zwischen den Geodaten von Quell- und Zielsystem: 0.32412624571717 km
Entityfacts: ⊕

Vollständigkeit der Zieldaten

Score: 0

181196	Abbehauser Altendeich	Niedersachsen	0 / 0	geonames	anzeigen (1)
181197	Abbehauser Hörne	Niedersachsen	0 / 0	geonames	anzeigen (1)
181198	Abbehauser Wehl	Niedersachsen	0 / 0	geonames	anzeigen (1)

GND (LD) SPARQL Query Editor

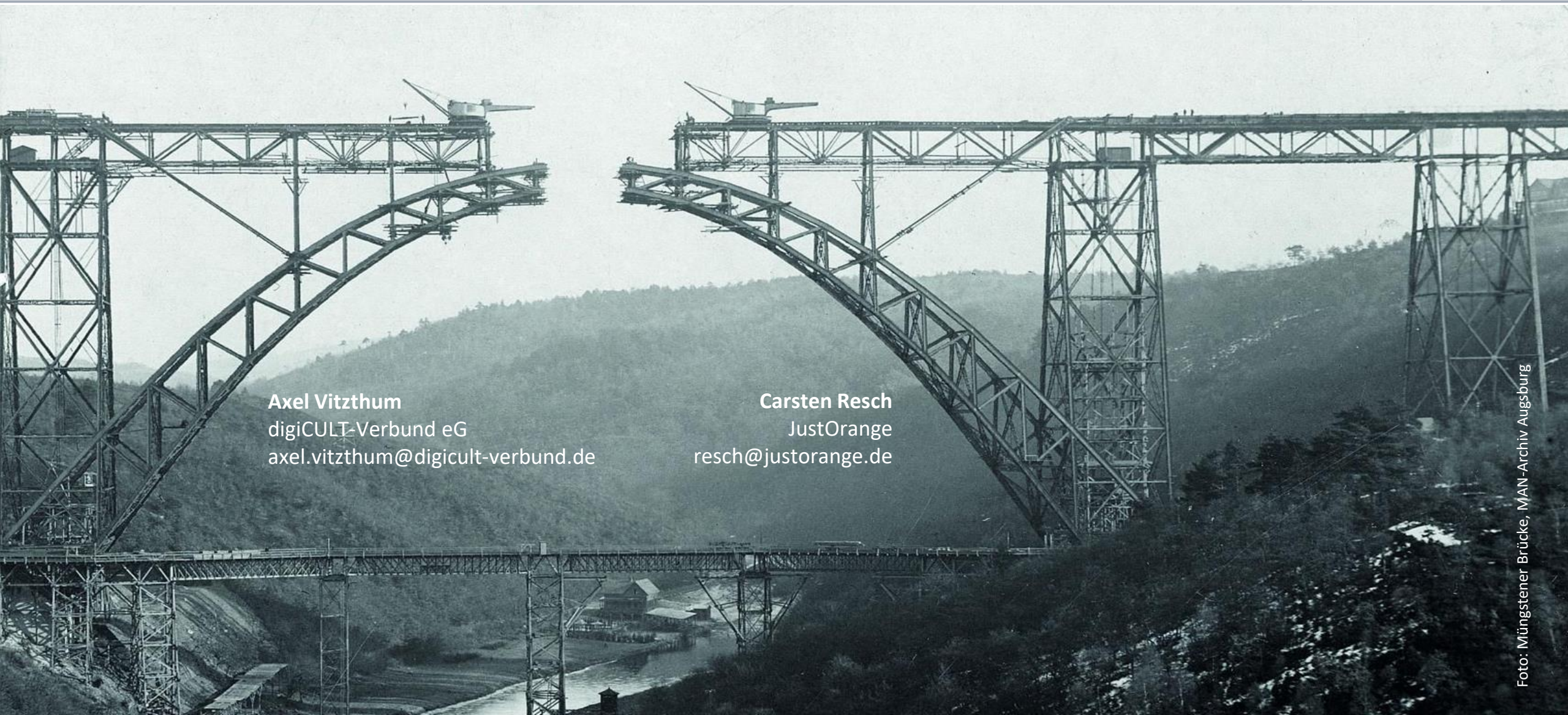
```
1 PREFIX wdt: <http://www.wikidata.org/prop/direct/>
2 PREFIX gvp: <http://vocab.getty.edu/ontology#>
3 PREFIX xl: <http://www.w3.org/2008/05/skos-xl#>
4 PREFIX : <_>
5 PREFIX gnd: <http://d-nb.info/gnd/>
6 PREFIX gndo: <http://d-nb.info/gnd/ontology#>
7 PREFIX rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
8 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
9 #
10 # Abfrage über zweistufige Federated Query via Wikidata
11 # ausgehend von einer GND-Instanz
12 # Gesucht wird nach GND-Instanzen, die
13 # oder mehrere Mapping-Instanzen zu einer
14 # Getty-Vokabular-Instanz haben
15 # sprachigen Vorzug haben
16 #
17 CONSTRUCT
18 {
19   ?gnduri
20     :gndLabel ?gndpref ;
21     :wdUri ?wdid ; :wdLabel ?wdlabel ;
22     :wd2aatMap [ :aatUri ?aaturi; :aatLabel ?aatlabel; :aatDeLabel ?aatDeLabel ]
23 }
24 WHERE
```

Beispiel-Abfragen

- Als Hutmacher bekannte Personen und Familien
- Sachbegriffe Holzbearbeitung, konvertiert in SKOS
- Zweifelsfälle in Hierarchiebeziehungen bei Sachbegriffen
- Range-Restriktionen für GND-Properties
- Französische Äquivalente zu Begriffen der Kunstgeschichte
- Zweistufige Federated Query via Wikidata
- ... und mehr in Kürze

Beispiele Präfixe Was ist das?

GND4C: under construction...



Axel Vitzthum
digiCULT-Verbund eG
axel.vitzthum@digicult-verbund.de

Carsten Resch
JustOrange
resch@justorange.de

Foto: Münstener Brücke, MAN-Archiv Augsburg

