

# MIT DER GND ARBEITEN DATENABGLEICH UND -ANREICHERUNG MITHILFE VON OPEN REFINE

Verena Mack (LABW)

# Agenda

- Einführung
  - Die Gurs Datenbank
- Was ist OpenRefine?
  - lobid-gnd und andere Reconciliation-APIs
- Datenabgleich und -anreicherung
  - Datenabgleich und seine Problemstellen
  - Datenanreicherung

# Datenabgleich und -anreicherung am Beispiel der Gurs Datenbank



The screenshot shows the website interface for the Gurs database. At the top, there is a navigation bar with a search icon, a 'MENÜ' dropdown, and links for 'Startseite LEO-BW', 'Über', 'Partner', 'Hilfe', and 'Anmelden'. The main content area features a large search bar with the text 'Entdeckungsreise starten' and a magnifying glass icon. Below the search bar, a prominent heading reads 'Gurs 1940. Die Deportation der jüdischen Bevölkerung aus Baden, der Pfalz und dem Saarland'. A paragraph of text follows, explaining that on October 22, 1940, approximately 6,500 people of Jewish faith or origin were deported to the Gurs internment camp in France. It notes that for the 80th anniversary in 2020, the state archive has created a database to remember this crime and honor the individuals. The interface includes several interactive tiles: 'Kartensuche' (Map search) with a map of the region; 'Einführung' (Introduction) with a photo of the camp; 'Lagersystem Gurs' (Gurs camp system) with a photo of people in the camp; 'Personenindex A-Z' (Alphabetical list of people) with a photo of a document; 'Zugänge' (Access); 'Projekte' (Projects); and 'Kontakt' (Contact).

## Nutzung von Normdaten im Projekt

- Verlinkung der Personen mit ihren Wohnorten zur Zeit der Deportation für einen kartenbasierten Sucheinstieg

Kartenauswahl ^ Vollbild - Karte groß Drucken PDF Legende v Hilfe

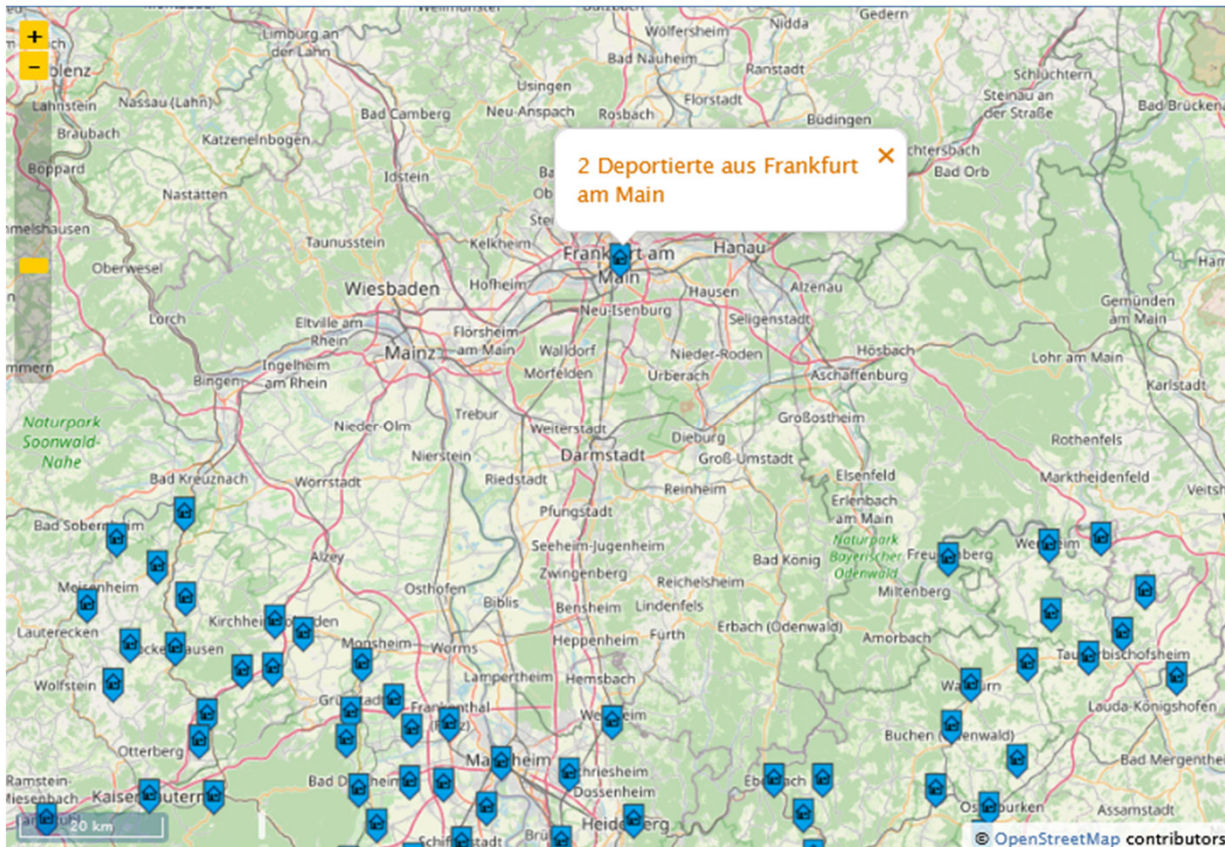
**THEMENKARTEN** ^

- Verfolgungsorte
- Deportationsziele

**HINTERGRUNDKARTEN** ^

- Basiskarte (Openstreetmap)

© OpenStreetMap contributors.



Mendel, Hedwig

Geburtsdatum/-ort 10.11.1885; Karlsruhe

Partner Landesarchiv Baden-Württemberg



Falkenstein, Frieda

Geburtsdatum/-ort 30.11.1895; Neckarzimmern

Beruf/Funktion ■ Hausgehilfin

Partner Landesarchiv Baden-Württemberg

## Personen aus Familien mit Anfangsbuchstaben M

Maas, Dora (1887–1970)  
 Maas, Fanny Selma (1893–?)  
 Maas, Friedrich (1874–1941)  
 Maas, Hedwig (1889–?)  
 Maas, Hermann (1880–1942)  
 Maas, Herta Martha Luise (1922–2005)  
 Maas, Isidor (1876–?); Bäcker  
 Maas, Johanna (1881–1952)  
 Maas, Klara (1883–?)  
 Maas, Klara Klothilde (1877–?)  
 Maas, Melanie (1878–?)  
 Maas, Moses Moritz (1875–1958)  
 Maas, Rolf Ludwig (1878–1940); Ingenieur, Versicherungsvertreter  
 Maas, Stefanie (1900–1948)  
 Machol, Benno (1875–1940)  
 Machol, Karl Benno (1881–1941)  
 Magath, Gertrud Selma (1897–?)  
 Magath, Hermann (1881–?); Kaufmann  
 Magath, Klaus (1927–2003)  
 Magsamer, Hugo (1897–?)  
 Mahler, Karoline (1867–1943)  
 Mai, Babette (1868–1941)  
 Mai, David (1858–1940)

Detailseite

### Mendel, Hedwig

<b>Geburtsname:</b>	Maier
<b>Geburtsdatum/–ort:</b>	10.11.1885; <a href="#">Karlsruhe</a>
<b>Geschlecht:</b>	weiblich
<b>Kurzbiografie:</b>	überlebte den Holocaust, war während der Deportation nur zu Besuch in Karlsruhe, emigrierte später nach Monaco.

#### Deportation

**Wohnort bei Deportation:** [Frankfurt am Main](#) [Klosestraße 33]

**Deportationsziele:**

- Gurs [22.10.1940]
- Auschwitz [10.08.1942]

#### Literatur + Links

**Weiterführende Links:**

- [Gedenkbuch Yad Vashem](#)
- [Hessisches Hauptstaatsarchiv Wiesbaden Bestand 676 Nr. 5784](#)

## Nutzung von Normdaten im Projekt

- Erweiterung der genutzten Normdaten denkbar
- z.B. für weiteren Sucheinstieg über die Berufe
  
- Dies soll beispielhaft an einem Ausschnitt der Gurs-Daten hier gezeigt werden



## Was ist OpenRefine?

- „Excel on steroids!“
  - Programm zur Bereinigung, Umwandlung und Anreicherung von tabellarischen Daten

## Reconciliation-APIs für OpenRefine

- Ein Reconciliation-Service, erstellt auf Basis von Text und ggf. zusätzlichen Informationen eine gewichtete Liste mit Kandidaten
- In OpenRefine kann jede Reconciliation-API eingebunden werden, die der [reconciliation API v0.1](#) folgt
  - Darunter: Wikidata, Geo-names(subsets), ORCID, WW2 thesaurus, Getty vocabularies, usw.
    - Auch: lobid-gnd

## Lobid-gnd



# lobid

- lobid ist ein Service des hbz — Hochschulbibliothekszenrum des Landes NRW
- Bietet UI und reconciliation-API zur GND auf JSON-LD Basis
  - > auch maschinell nutzbar
- Basierend auf aufbereiteten GND-RDF-Daten von der DNB via [OAI-PMH](#)

## Datenabgleich mit der GND

- Vor dem Abgleich: Daten aufräumen!
  - Schreibfehler korrigieren
  - Feldbelegungen vereinheitlichen
  - Schreibweisen vereinheitlichen
    - v.a. von Berufsangaben
  - Datumsangaben vereinheitlichen

# Datenabgleich in OpenRefine

126 rows

Show as: rows records    Show: 5 10 25 50 100 500 1000 rows

« first < previous 1

Geburtsort-GND-ID	Beruf	Religion	Wohnort (Deportation)	Wohnort-GND-ID (Deportation)	Todesdatum (exakt)	Sterbeort	Sterbeort-GND-ID
4026463-4		Facet		4104199-9	08.05.1942	Toulouse (Internierungslager, Frankreich)	4060518-8
4018118-2		Text filter	aden	4004169-4	26.04.2000	Salt Lake City (USA)	4118295-9
4095962-4	Zahnarzt	Edit cells	aden	4004169-4	03.11.1974	Straßburg (Frankreich)	4057878-1
4739647-7	Kunstmaler	Edit column	aden	4004169-4	04.09.1942	Noé (Internierungslager, Frankreich)	4444429-1
4043115-0	Arzt	Transpose	aden	4004169-4	21.11.1940	Gurs	4243797-0
4062311-7	Haushälterin	Sort...	israelitisch	Bretten			
4309185-4		View	israelitisch	Bruchsa	08.05.1945	Auschwitz	4003697-2
4008726-8		Reconcile	israelitisch	Bühl	25.02.1942	Gurs	4243797-0
4014586-4	Händlerin	Start reconciling...	israelitisch	Emmenc	10.10.1942	Johannesburg (Südafrika)	4028698-8
4018272-1	Auszubildender	Facets	israelitisch	Emmenc	00.00.1952	Emmendingen	4014586-4
4835832-0	Landgerichtsrat	Actions	israelitisch	Frankenthal	10.08.1942	Auschwitz	4003697-2
4070734-9	Händler	Copy reconciliation data...	israelitisch	Freiburg	22.10.1940	Freiburg	4018272-1
4037372-1	Kinderärztin	Use values as identifiers...	israelitisch	Freiburg	14.08.1942	Auschwitz	4003697-2
4018272-1	Rechtsanwalt, Politiker	Add entity identifiers column...	israelitisch	Freiburg	00.00.1960	Freiburg	4018272-1

urs GND Permalink

### Reconcile column "Religion"

[Access service API](#)

Reconcile each cell to an entity of one of these types:

- Normdatenressource  
AuthorityResource
- Körperschaft  
CorporateBody
- Individualisierte Person  
DifferentiatedPerson

Also use relevant details from other columns:

Column	Include?	As property
ID	<input type="checkbox"/>	<input type="text"/>
ID Freigegeben	<input type="checkbox"/>	<input type="text"/>
Geschlecht (m/w)	<input type="checkbox"/>	<input type="text"/>
akad. Grad/Titel	<input type="checkbox"/>	<input type="text"/>
Namen	<input type="checkbox"/>	<input type="text"/>
Vornamen	<input type="checkbox"/>	<input type="text"/>
Geburtsname	<input type="checkbox"/>	<input type="text"/>
Name GND-Nummer	<input type="checkbox"/>	<input type="text"/>
Geburtsdatum (exakt)	<input type="checkbox"/>	<input type="text"/>
Geburtsort	<input type="checkbox"/>	<input type="text"/>
Geburtsort-GND-ID	<input type="checkbox"/>	<input type="text"/>
Beruf	<input type="checkbox"/>	<input type="text"/>

Reconcile against type:

Reconcile against no particular type

Auto-match candidates with high confidence

Maximum number of candidates to return

## Automatisches Matching für einen Teil der Daten

Der Abgleich basiert auf einem internen  
Scoring-System

The screenshot shows a data management interface with a facet filter on the left and a table of 126 records on the right. The facet filter for 'Beruf' (Occupation) is active, showing 3 choices: (blank) 42, matched 67, and none 34. Below this, a 'best candidate's score' histogram shows a distribution of scores from 23 to 80. The table on the right lists records with columns for 'Geburtsort' (Place of Birth), 'Geburtsort-GND-ID', and 'Beruf' (Occupation). The 'Beruf' column contains suggestions for each record, such as 'Zahnarzt' for Ingenheim and 'Haushälterin' for Vallendar. Each suggestion includes a 'Choose new match' link and a 'Create new item' checkbox.

Geburtsort	Geburtsort-GND-ID	Beruf
Idar-Oberstein	4026463-4	
Frankfurt am Main	4018118-2	
Ingenheim	4095962-4	Zahnarzt
Heinebach	4739647-7	Kunstmaler
Odessa (Russisches Reich heute Ukraine)	4043115-0	Arzt
Vallendar	4062311-7	Haushälterin
Flehingen	4309185-4	
Bühl	4008726-8	
Emmendingen	4014586-4	Händlerin
Freiburg	4018272-1	Auszubildender
Osann bei Trier	4835832-0	Landgerichtsrat
Elberfeld	4070734-9	Händler
Mannheim	4037372-1	Kinderärztin

# Nicht alles ist sinnvoll abzugleichen!

Search for match		
israelitisch	Freibu	
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item		
Search for match		
<b>edit</b> evangelisch	Freibu	
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Evangelisch-reformiertes Glaubensbekenntnis (28)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Evangelisch-lutherisches Glaubensbekenntnis (28)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Evangelisch-reformiertes Bekenntnis (28)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Evangelisch-lutherisches Bekenntnis (28)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Evangelisch-Theologische Fakultät (28)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Neuendettelsau / Evangelisch-Lutherisches Diakoniewerk / Jubiläum <2004> (21)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Vereinigte Evangelisch-Lutherische Kirche Deutschlands / Jubiläum <1998> (21)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Evangelisch-Lutherische Kirche in Oldenburg / Jubiläum <1999> (21)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Reformierte Kirche (18)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Lutherische Freikirche (18)		
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item		
Search for match		
evangelisch	Freibu	
<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Evangelisch-reformiertes Glaubensbekenntnis (28)		

?



Auf was soll hier gematcht werden?

[Sachbegriff (GND)]		Verwendung:   s
Sachbegriff:	<b>Evangelische Kirche</b>	
	Hierarchisch untergeordnete Sachbegriffe?	
K10plusPPN:	106335138	
GND-Nummer:	4015833-0	
	Link zu diesem Datensatz in der GND: <a href="http://d-nb.info/gnd/4015833-0">http://d-nb.info/gnd/4015833-0</a>	
Typ:	Allgemeinbegriff (saz)	
Quelle:	Au=DB; Beispiel in RSWK 3. Aufl.	
Verwendungshinweis:	Bis März 2010: nicht permutierend, wenn nur zur Unterscheidung bzw. Präzisierung der konfessionellen Richtung benutzt, d. h. wenn Konfession bzw. Religionsgemeinschaft nicht Hauptgegenstand der Darstellung ist; ist die Evangelische Kirche Hauptgegenstand der Darstellung, permutiert das SW (seit April 2010 permutiert "Evangelische Kirche" immer)	
GND-Systematik:	3.6b [Evangelische Kirchen]	
DDC-Notation:	284	
Synonym:	Protestant Church (AF) (IxTheo: alternative Ansetzung in eng) Igreja evangélica (AF) (IxTheo: alternative Ansetzung in por) Chiesa evangelica (AF) (IxTheo: alternative Ansetzung in ita) Église protestante (AF) (IxTheo: alternative Ansetzung in fre) Iglesia evangélica (AF) (IxTheo: alternative Ansetzung in spa) Evangelical Church (VW) (IxTheo: alternative Verweisung in eng) 新教教会 (AF) (IxTheo: alternative Ansetzung in chi) 新教教會 (AF) (IxTheo: alternative Ansetzung in chi) Евангелическая церковь (AF) (IxTheo: alternative Ansetzung in rus)	
Thematischer Bezug:	Protestantismus [Verwandter Begriff, allgemein]	
In anderen Normdaten:	Protestant churches (DLC) Églises protestantes	
	Suche nach Eintrag "Evangelische Kirche" in WIKIPEDIA?	
	Titelsuche mit Normsatz "Evangelische Kirche" <b>im BVB?</b>	
	Titelsuche mit Normsatz "Evangelische Kirche" <b>im OBV?</b>	



# Auftretende Probleme: Mehrfachbelegung von Zellen

**OpenRefine** vortrag beispiel Gurs GND [Permalink](#)

Facet / Filter Undo / Redo 1 / 1

Refresh Reset all Remove all

**126 rows** Show as: rows records Show: 5 10 25 50 100 500 1000

**Beruf** change 61 choices Sort by: name count Cluster

- Haushälterin 1
- Hilfsarbeiter 1
- Jurist und Dichter 1
- Kantor, Chorleiter, Religionslehrer 1 edit include
- Kanzleigehilfe 1
- Kaufmann, Politiker 1
- Kaufmännischer Gehilfe 1
- Kinderarzt 1
- Kinderärztin 1
- Kinderärztin und Lehrerin 1
- Kontoristin 1

Geburtsort-GND-ID	Beruf	Religion	Wohn
4026463-4		israelitisch	Appenweik
4018118-2		israelitisch	Baden-Ba
4095962-4	Zahnarzt	israelitisch	Baden-Ba
4739647-7	Kunstmaler	israelitisch	Baden-Ba
4043115-0	Arzt	israelitisch	Baden-Ba
4062311-7	Haushälterin	israelitisch	Bretten
4309185-4		israelitisch	Bruchsal
4008726-8		israelitisch	Bühl
4014586-4	Händlerin	israelitisch	Emmending
4018272-1	Auszubildender	israelitisch	Emmending
4835832-0	Landgerichtsrat	israelitisch	Frankenth
4070734-9	Händler	israelitisch	Freiburg
4037372-1	Kinderärztin	israelitisch	Freiburg
4018272-1	Rechtsanwalt, Politiker	israelitisch	Freiburg

# Fehlerbehebung: Zellen aufteilen

Show: 5 10 25 50 100 500 1000 rows

sort-GND-ID	Beruf	Religion	Wohnort (Deportation)
	Facet	israelitisch	Appenweiler
	Text filter		
	Edit cells	Transform...	
	Edit column	Common transforms	
	Transpose	Fill down	
	Sort...	Blank down	
	View	Split multi-valued cells...	
	Reconcile	Join multi-valued cells...	
	Haushälterin	Cluster and edit...	
		Replace...	
		israelitisch	Bühl
	Händlerin	israelitisch	Emmendingen

### Split multi-valued cells

**How to split multi-valued cells**

by separator  
 Separator   regular expression

by field lengths

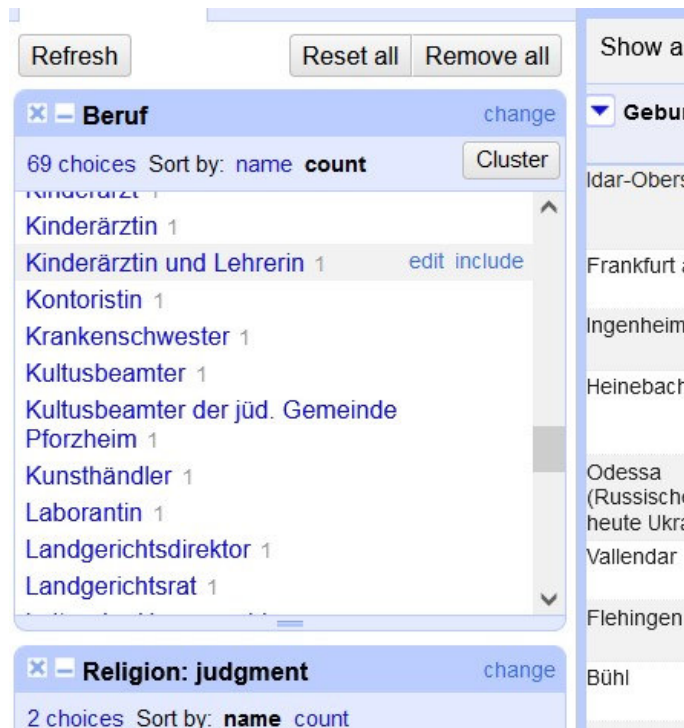
List of integers separated by commas, e.g., 5, 7, 15

by transition from lowercase to uppercase  
 Reverse splitting order  
 [11Abc, Def22]

by transition from numbers to letters  
 Reverse splitting order  
 [11, AbcDef22]

OK Cancel

# Achtung! Uneinheitliche Daten machen Probleme



Refresh    Reset all    Remove all    Show as:

**Beruf** change

69 choices Sort by: name count Cluster

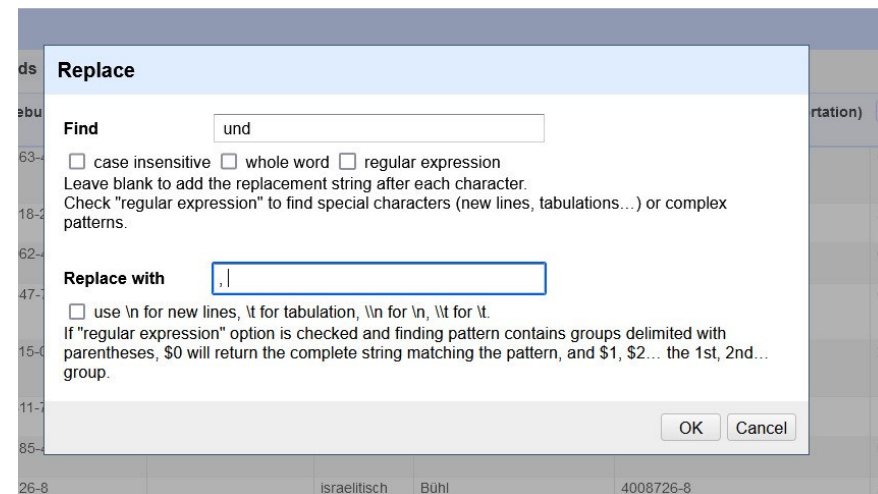
- Kinderärztin 1
- Kinderärztin und Lehrerin 1 edit include
- Kontoristin 1
- Krankenschwester 1
- Kultusbeamter 1
- Kultusbeamter der jüd. Gemeinde Pforzheim 1
- Kunsthändler 1
- Laborantin 1
- Landgerichtsdirektor 1
- Landgerichtsrat 1

**Religion: judgment** change

2 choices Sort by: name count

**Gebur**

- Idar-Obers
- Frankfurt a
- Ingenheim
- Heinebach
- Odessa (Russische heute Ukra
- Vallendar
- Flehing
- Bühl



**Replace**

Find

case insensitive     whole word     regular expression

Leave blank to add the replacement string after each character.  
Check "regular expression" to find special characters (new lines, tabulations...) or complex patterns.

Replace with

use \n for new lines, \t for tabulation, \n for \n, \t for \t.

If "regular expression" option is checked and finding pattern contains groups delimited with parentheses, \$0 will return the complete string matching the pattern, and \$1, \$2... the 1st, 2nd... group.

OK Cancel

# Auftretende Probleme: Nicht alles, was für den Menschen lesbar ist, versteht die Maschine

		israelitisch	Friese
Weinhändler		israelitisch	Genge
		israelitisch	Germe
Buch-	<a href="#">edit</a>	israelitisch	Heidel
Kunsthändler	<a href="#">Edit this cell</a>		
Lehrer		israelitisch	Heidel
Publizistin		israelitisch	Heidel

-> Abkürzungen sollten in der Datenbereinigung aufgelöst werden

## Auftretende Probleme: kein genaues Match verfügbar

- Oft sind die Normdatensätze der Berufe allgemeiner als die Beschreibung in den archivischen Daten
  - e.g. „Gehilfe“ statt „Kaufmännischer Gehilfe“

	Search for match	
Universitätsprofessor	Universitätsprofessor <a href="#">edit</a> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hochschullehrer (23) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	israe
Arzt	Arzt Choose new match	
Kaufmännischer Gehilfe	Gehilfe Choose new match	israe
Leiter der Herrenmühle	Leiter	israe

- Auch können ähnliche Bezeichnungen auf den gleichen Normdatensatz verweisen

Dichterin	<a href="#">Search for match</a> Dichterin is <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Schriftstellerin (38) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
Schriftstellerin	<a href="#">Search for match</a> Schriftstellerin <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
Kultusbeamter der	<a href="#">Search for match</a> Kultusbeamter der iüd. Gemeinde is

## Auftretende Probleme: Normdatensätze zu spezifisch

- In manchen Tätigkeitsbeschreibungen in der GND findet sich die Nationalität bzw. Religionszugehörigkeit wieder, um diese Informationen im Datenmodell unterzubringen -> manchmal zu kleingliedrig

Schüler zum Zeitpunkt der Deportation	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Schüler zum Zeitpunkt der Deportation <a href="#">edit</a> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Zeitpunkt (77) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Deportation (73) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Deportation <Motiv> (61) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Schüler (47) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Schüler schreiben - Schzeichnen (44) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Marokkanischer Schüler (38) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jugoslawischer Schüler <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Kroatischer Schüler (38) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Serbischer Schüler (38) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <u>Jüdischer Schüler</u> (38) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item <input type="text"/> Search for match	israelitisch	Mannheim	4037372-1
später Lehrer	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> später Lehrer <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Lehrer (50) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Polnischer Lehrer (40)			

Jüdischer Schüler (7504662-3)

Schlagwort senso stricto

Open... Export ▾ Help

Extensions Wikidata

« first < previous 1 next > last »

▼ Wohnort (Deportation) ▼ Wohnort-GND-ID (Deportation) ▼ Todesdatum (exakt) ▼

### Search for match

Search for "Kunstmaler"

Match other cells with same content  
 Match this cell only

**Match** **New Item** **D**

Maler

Select an item from the list:

- Maler** 121780791  
1957- | Poln. Malerin in Frankreich
- Maler** 1229178791
- Maler** 4037215-7
- Maler und Lackierer, Fachrichtung Maler** 4555227-7  
Ausbildungsberuf
- Maler, Johannes** 119751402  
Beitr. im VD-16
- Pix, Maler** 12785567X
- Maler, Wilhelm** 124737595  
1902-1976 | Komponist, Deutschland
- Maler, Christian** 129149705  
1578-1652 | Sohn von Valentin Maler; Medailleur un
- Maler, Sabina** 132542447

Maler (1229178791)  
Schauspieler  
Individualisierte Person

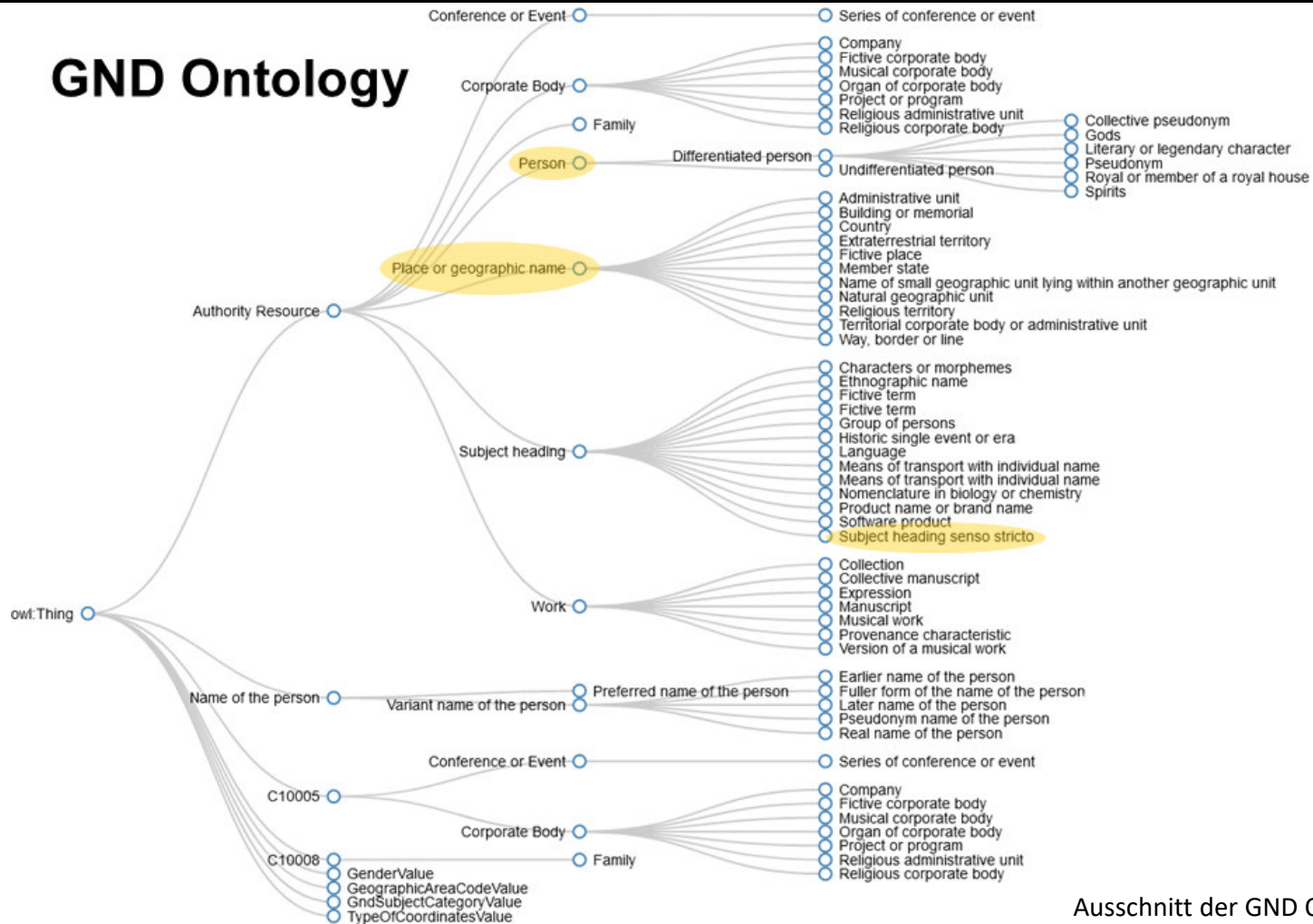
Kunstmaler	Kunstmaler <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match					
Arzt	Arzt Choose new match					
Haushälterin	Haushälterin <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Pfarrhaushälterin (35) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Hausgehilfin (35) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match					
Händlerin	Händlerin Choose new match	israelitisch	Emmendingen	4014586-4	10.10.1942	Joha (Süd
Auszubildender	Auszubildender	israelitisch	Emmendingen	4014586-4	00.00.1952	Emm



## Auftretende Probleme: false positives

- Bei automatischer Zuweisung des Matches mit dem besten Scoring können falsche Zuweisungen entstehen
- Tritt v.a. auf, wenn sehr allgemein gegen die GND abgeglichen wird (z.B. gegen alle Sachschlagworte, statt nur gegen Berufe)
  - Kenntnisse der GND-Ontologie verbessern den Abgleich

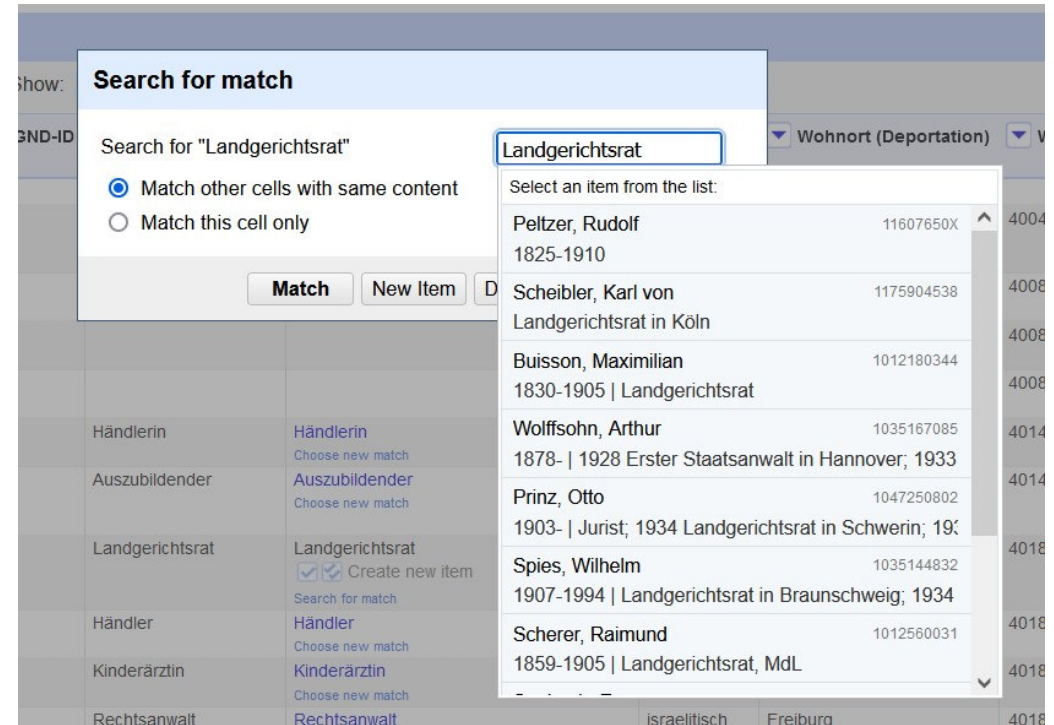
Rechtsanwalt	Rechtsanwalt Choose new match		israe
Politiker	Politiker Choose new match		
Senatspräsident am Oberlandesgericht	Oberlandesgericht Choose new match	edit	israe
Rechtsanwalt	Rechtsanwalt Choose new match		israe
Zahnarzt	Zahnarzt Choose new match		israe



Ausschnitt der GND Ontologie Abbildung: FDMLab

# Auftretende Probleme: Nicht alle Begriffe bereits in der GND vorhanden

- Teilweise aber trotzdem in den Beschreibungen als Literal zu finden
- -> Gefahr von false positives bei breit angelegten Suchen



**Search for match**

Search for "Landgerichtsrat"

Match other cells with same content  
 Match this cell only

**Match** **New Item**

Select an item from the list:

Name	GND-ID	Wohnort (Deportation)
Peltzer, Rudolf 1825-1910	11607650X	4004
Scheibler, Karl von Landgerichtsrat in Köln	1175904538	4008
Buisson, Maximilian 1830-1905   Landgerichtsrat	1012180344	4008
Wolffsohn, Arthur 1878-   1928 Erster Staatsanwalt in Hannover; 1933	1035167085	4014
Prinz, Otto 1903-   Jurist; 1934 Landgerichtsrat in Schwerin; 1934	1047250802	4018
Spies, Wilhelm 1907-1994   Landgerichtsrat in Braunschweig; 1934	1035144832	4018
Scherer, Raimund 1859-1905   Landgerichtsrat, MdL	1012560031	4018

## Anreicherung von Daten

- Nach dem Abgleich der Daten mit der GND (oder anderen Normdateien) können durch diese Verknüpfungen weitere Daten nachgeladen werden

500 1000 records

Beruf GND	Religion	Wohnort (Deportation)	Wohnort-Gl
Facet	israelitisch	Bretten	4008223-4
Text filter	israelitisch	Bruchsal	4008388-3
Edit cells			008726-8
Edit column	Split into several columns...		014586-4
Transpose	Join columns...		014586-4
Sort...	Add column based on this column...		018114-5
View	Add column by fetching URLs...		018272-1
Reconcile	Add columns from reconciled values...		018272-1
Händler	Rename this column...		018272-1
Kinderärztin	Remove this column		018272-1
Rechtsanwalt	Move column to beginning		018272-1
Politiker	Move column to end		
Arzt	Move column left		
	Move column right		018272-1

### Add columns from reconciled column Beruf GND

Add Property      Preview      Reset

Oberbegriff mehrgliedrig

Suggested Properties

- In Beziehung stehende Dewey-Dezimalklassifikation mit Determiniertheitsgrad 2
- In Beziehung stehende Dewey-Dezimalklassifikation mit Determiniertheitsgrad 3
- In Beziehung stehende Dewey-Dezimalklassifikation mit Determiniertheitsgrad 4
- In Beziehung stehende Familie
- In Beziehung stehende Konferenz oder Veranstaltung
- In Beziehung stehende Körperschaft
- In Beziehung stehende Person

Beruf GND	GND-Nummer <a href="#">remove</a> <a href="#">configure</a>	Oberbegriff allgemein <a href="#">remove</a> <a href="#">configure</a>	Oberbegriff generisch <a href="#">remove</a> <a href="#">configure</a>
<not reconciled>			
Zahnarzt	4067303-0	Facharzt	
Maler	4037215-7		Künstler
Arzt	4003157-3		Heilberuf
Hausgehilfin	4159247-5		
<not reconciled>			
Handlerin	4158695-5		
Auszubildender	4035113-0		
Richter	4049923-6	Jurist	
Handler	4221401-4		
Kinderärztin	4315464-5	Fachärztin	
Rechtsanwalt	4048755-6		Jurist
Politiker	4046517-2		
Arzt	4003157-3		Heilberuf
Kinderärztin	4315464-5	Fachärztin	

OK Cancel

1000 records

Beruf GND	GND-Nummer	Oberbegriff	Oberbegriff GND ID	
				israelit
Zoologin <small>Choose new match</small>	7537890-5	Biologin	4380756-2	israelit
Arzt <small>Choose new match</small>	4003157-3	Heilberuf	4024024-1	israelit
Schriftsteller <small>Choose new match</small>	4053309-8	Autor	4003982-1	
Jurist <small>Choose new match</small>	4029050-5			israelit
Schriftsteller <small>Choose new match</small>	4053309-8	Autor	4003982-1	
Hochschullehrer <small>Choose new match</small>	4025243-7			israelit
Arzt <small>Choose new match</small>	4003157-3	Heilberuf	4024024-1	
Gehilfe <small>Choose new match</small>	4156321-9			israelit
Kaufmann <small>Choose new match</small>	4128272-3	Kaufmännischer Beruf	4030067-5	israelit
Stadtratsmitglied <small>Choose new match</small>	4134536-8	Mitglied	4114609-8	
		Kommunalpolitiker	4396366-3	
Hochschullehrer <small>Choose new match</small>	4025243-7			israelit
Arzt	4003157-3	Heilberuf	4024024-1	israelit

Neue Spalten

**Achtung:**  
beim Zusammenführen von Spalten verwirft OpenRefine die hinterlegten Reconciliation der vorherigen Spalten

- Alle im Projekt vorhandenen IDs können zum Nachladen der Daten verwendet werden

100 500 1000 records « first

Wohnort-GND-ID (Deportation)	Todesdatum (exakt)	Sterbeort	Sterbeort-GND-ID
	16.05.1960	Schweiz	4053881-3
	19.03.1945	Berkeley (Kalifornien, USA)	4005720-3
	08.05.1945	Auschwitz	4003697-2
		Winterthur (Schweiz)	4066337-1
4023996-2		Heidelberg	4023996-2
4023996-2			
4023996-2			

- Facet ▶
- Text filter
- Edit cells ▶
- Edit column ▶
- Transpose ▶
- Sort...
- View ▶
- Reconcile ▶
  - Start reconciling...
  - Facets ▶
  - Actions ▶
  - Copy reconciliation data...
  - Use values as identifiers...
  - Add entity identifiers... Mark cells as reconciled with their values as identifiers
  - column...

how as: rows records Show: 5 10 25 50 100 500 1000 records « first < prev

on	Geburtsort	Geburtsort-GND-ID	Geburtsort Ländercode	Wohnort (Deportation)	Wohnort-GND-ID (Deportation)	Wohnort Ländercode	Todes
	Idar-Oberstein	4026463-4 <a href="#">Choose new match</a>	Rheinland-Pfalz	Appenweiler	4104199-9 <a href="#">Choose new match</a>	Baden-Württemberg	08.05.194
	Frankfurt am Main	4018118-2 <a href="#">Choose new match</a>	Hessen	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	26.04.200
	Ingenheim	4095962-4 <a href="#">Choose new match</a>	Rheinland-Pfalz	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	03.11.197
	Heinebach	4739647-7 <a href="#">Choose new match</a>	Hessen	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	04.09.194
	Odessa (Russisches Reich heute Ukraine)	4043115-0 <a href="#">Choose new match</a>	Ukraine	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	21.11.194
	Vallendar	4062311-7 <a href="#">Choose new match</a>	Rheinland-Pfalz	Bretten	4008223-4 <a href="#">Choose new match</a>	Baden-Württemberg	
	Fiehingen	4309185-4 <a href="#">Choose new match</a>	Baden-Württemberg	Bruchsal	4008388-3 <a href="#">Choose new match</a>	Baden-Württemberg	08.05.194
	Bühl	4008726-8 <a href="#">Choose new match</a>	Baden-Württemberg	Bühl	4008726-8 <a href="#">Choose new match</a>	Baden-Württemberg	25.02.194
	Emmendingen	4014586-4 <a href="#">Choose new match</a>	Baden-Württemberg	Emmendingen	4014586-4 <a href="#">Choose new match</a>	Baden-Württemberg	10.10.194
	Freiburg	4018272-1 <a href="#">Choose new match</a>	Baden-Württemberg	Emmendingen	4014586-4 <a href="#">Choose new match</a>	Baden-Württemberg	00.00.195
	Osann bei Trier	4835832-0 <a href="#">Choose new match</a>	Rheinland-Pfalz	Frankenthal	4018114-5 <a href="#">Choose new match</a>	Rheinland-Pfalz	10.08.194
	Elberfeld	4070734-9 <a href="#">Choose new match</a>	Deutschland	Freiburg	4018272-1 <a href="#">Choose new match</a>	Baden-Württemberg	22.10.194
	Mannheim	4037372-1	Deutschland, Deutsches Reich Baden-Württemberg	Freiburg	4018272-1	Baden-Württemberg	14.08.194

Beispiel: Nachladen der Ländercodes der Geburts- und Wohnorte



how as: rows records Show: 5 10 25 50 100 500 1000 records

on	Geburtsort	Geburtsort-GND-ID	Geburtsort Ländercode	Wohnort (Deportation)	Wohnort-GND-ID (Deportation)	Wohnort Ländercode	Todes
	Idar-Oberstein	4026463-4 <a href="#">Choose new match</a>	Rheinland-Pfalz	Appenweiler	4104199-9 <a href="#">Choose new match</a>	Baden-Württemberg	08.05.194
	Frankfurt am Main	4018118-2 <a href="#">Choose new match</a>	Hessen	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	26.04.200
	Ingenheim	4095962-4 <a href="#">Choose new match</a>	Rheinland-Pfalz	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	03.11.197
	Heinebach	4739647-7 <a href="#">Choose new match</a>	Hessen	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	04.09.194
	Odessa (Russisches Reich heute Ukraine)	4043115-0 <a href="#">Choose new match</a>	Ukraine	Baden-Baden	4004169-4 <a href="#">Choose new match</a>	Baden-Württemberg	21.11.194
	Vallendar	4062311-7 <a href="#">Choose new match</a>	Rheinland-Pfalz	Bretten	4008223-4 <a href="#">Choose new match</a>	Baden-Württemberg	
	Fiehingen	4309185-4 <a href="#">Choose new match</a>	Baden-Württemberg	Bruchsal	4008388-3 <a href="#">Choose new match</a>	Baden-Württemberg	08.05.194
	Bühl	4008726-8 <a href="#">Choose new match</a>	Baden-Württemberg	Bühl	4008726-8 <a href="#">Choose new match</a>	Baden-Württemberg	25.02.194
	Emmendingen	4014586-4 <a href="#">Choose new match</a>	Baden-Württemberg	Emmendingen	4014586-4 <a href="#">Choose new match</a>	Baden-Württemberg	10.10.194
	Freiburg	4018272-1 <a href="#">Choose new match</a>	Baden-Württemberg	Emmendingen	4014586-4 <a href="#">Choose new match</a>	Baden-Württemberg	00.00.195
	Osann bei Trier	4825002-0 <a href="#">Choose new match</a>	Rheinland-Pfalz	Frankenthal	4018114-5 <a href="#">Choose new match</a>	Rheinland-Pfalz	10.08.194
	Elberfeld	4070734-9 <a href="#">Choose new match</a>	Deutschland	Freiburg	4018272-1 <a href="#">Choose new match</a>	Baden-Württemberg	22.10.194
	Mannheim	4037372-1 <a href="#">Choose new match</a>	Deutschland, Deutsches Reich Baden-Württemberg	Freiburg	4018272-1 <a href="#">Choose new match</a>	Baden-Württemberg	14.08.194



Für Orte in Deutschland sind als „Ländercode“ die Bundesländer hinterlegt. Aber nicht für historische Orte

Beispiel: Nachladen der Ländercodes der Geburts- und Wohnorte

## Fazit

- OpenRefine kann als Tool zur Unterstützung des Datenabgleichs und der Anreicherung aus und mit Normdateien wie der GND genutzt werden
- Als völlig automatisierter Prozess ohne manuelle (Nach)Bearbeitung ist dies aber nicht möglich
- Der Aufwand und die Vollständigkeit des Ergebnisses ist je nach spezifischem Abgleich dabei stark schwankend

## Weiterführende Links

- Website OpenRefine, <https://openrefine.org/>
- Übersicht zu den Reconciliation services von OpenRefine, <https://reconciliation-api.github.io/testbench/>
- Workshops des FDMLab@LABW, <https://fdmlab.landesarchiv-bw.de/workshops/>
- Workshop des hbz, <https://slides.lobid.org/2021-kim-reconcile/>

# Vielen Dank für Ihre Aufmerksamkeit!

Ihre Fragen gerne jetzt im Anschluss...

...oder via [gnd4c@leo-bw.de](mailto:gnd4c@leo-bw.de)

<https://www.leo-bw.de/gnd-agentur>

