

Übersicht von Personen- Identifiern in Wikidata

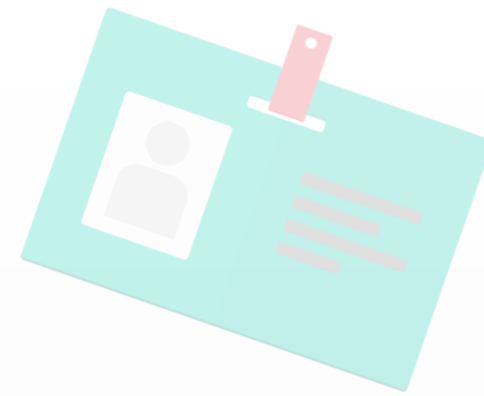
Prüfungsleistung im Modul
Semantische Datenintegration
SoSe 2017

Friederike Kühn – Matr.-Nr.: 1378387



Gliederung

- Wiederholung Identität
- Identifier in Wikidata
- Personen-Identifier in Wikidata
- Analysen



Identifizierung in Wikidata



- Items in Wikidata haben Identifier
- es gibt aber auch definierte Properties, die der Integration externer IDs dienen
 - z.B. Property **P227** GND ID
Identifier from an international authority file of names, subjects, and organizations
(vgl.: <https://www.wikidata.org/wiki/Property:P227>)
- das Objekt im Tripel ist dann eine URL, die zum Datensatz der ID im externen System führt
- Nutzen: Dinge eindeutig identifizieren, Mapping, Datenintegration



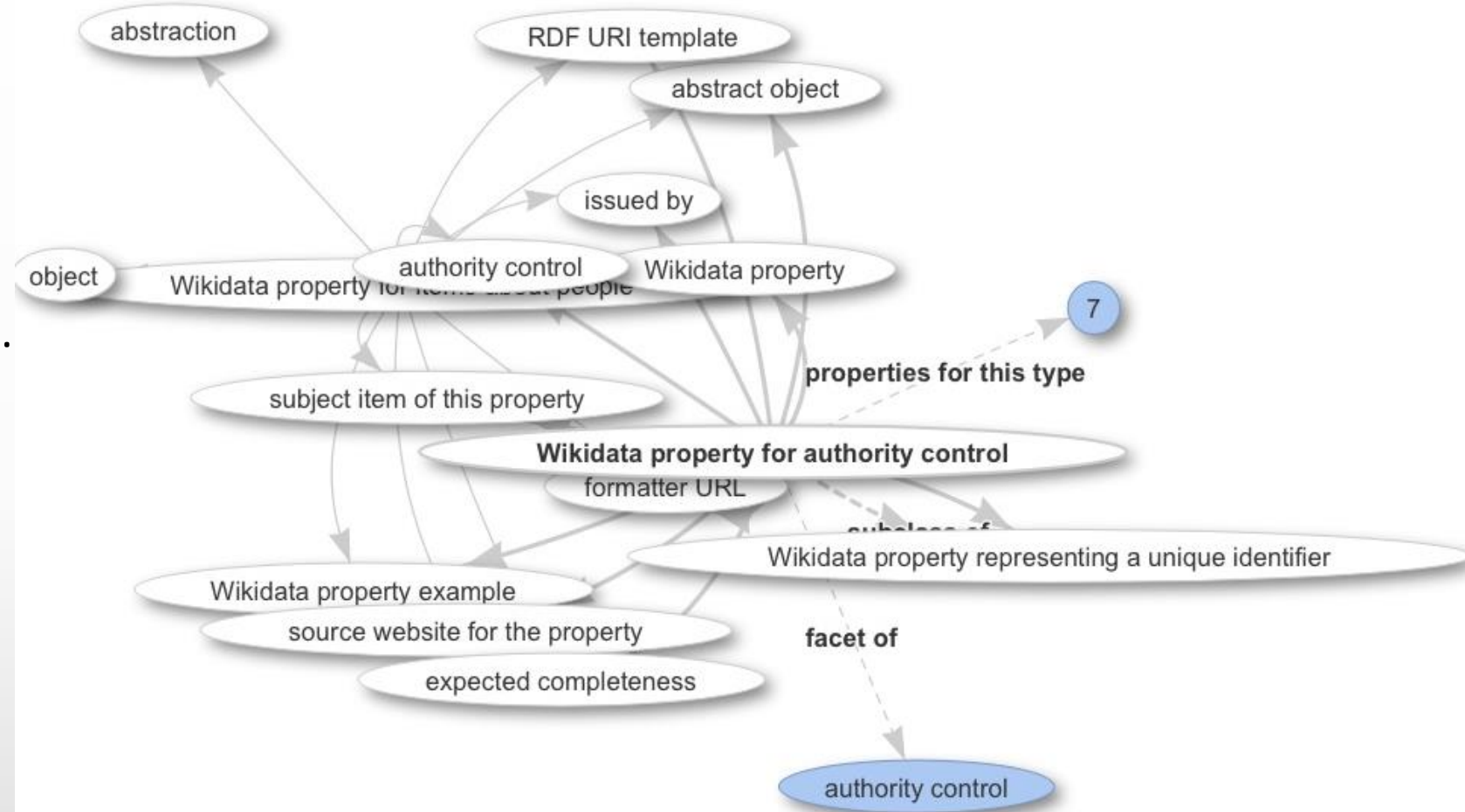
Personen-Identifier in Wikidata

- Wikidata property for authority control for people
= Wikidata-Eigenschaft für Personen-Normdaten (Q19595382)
- [Übersicht über alle externen Personen-Ids](#) (572, Stand: 28.05.2017)
- [SPARQL](#):

```
SELECT ?prop ?propLabel ?propDescription
WHERE{
  ?prop wdt:P31 wd:Q19595382 .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
ORDER by ?propLabel
```

→ **622** (Stand: 28.05.2017)

(Stand: 27.05.17)



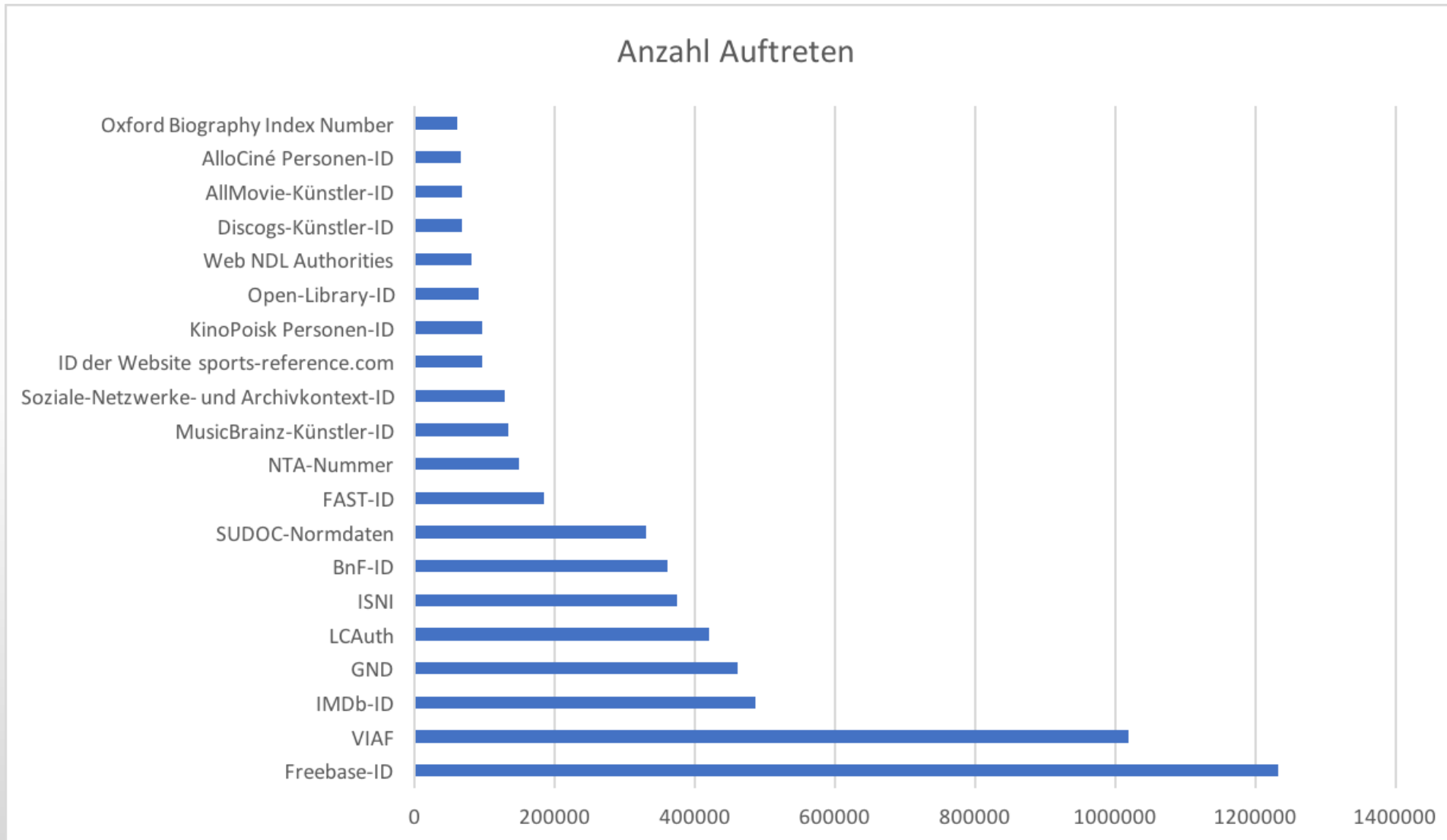
Beispiel VIAF ID (P214)

- [P214](#)
- [SPARQL](#):

```
SELECT ?a ?aLabel ?b
WHERE
{
  wd:P214 ?a ?b.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" .
  }}
```

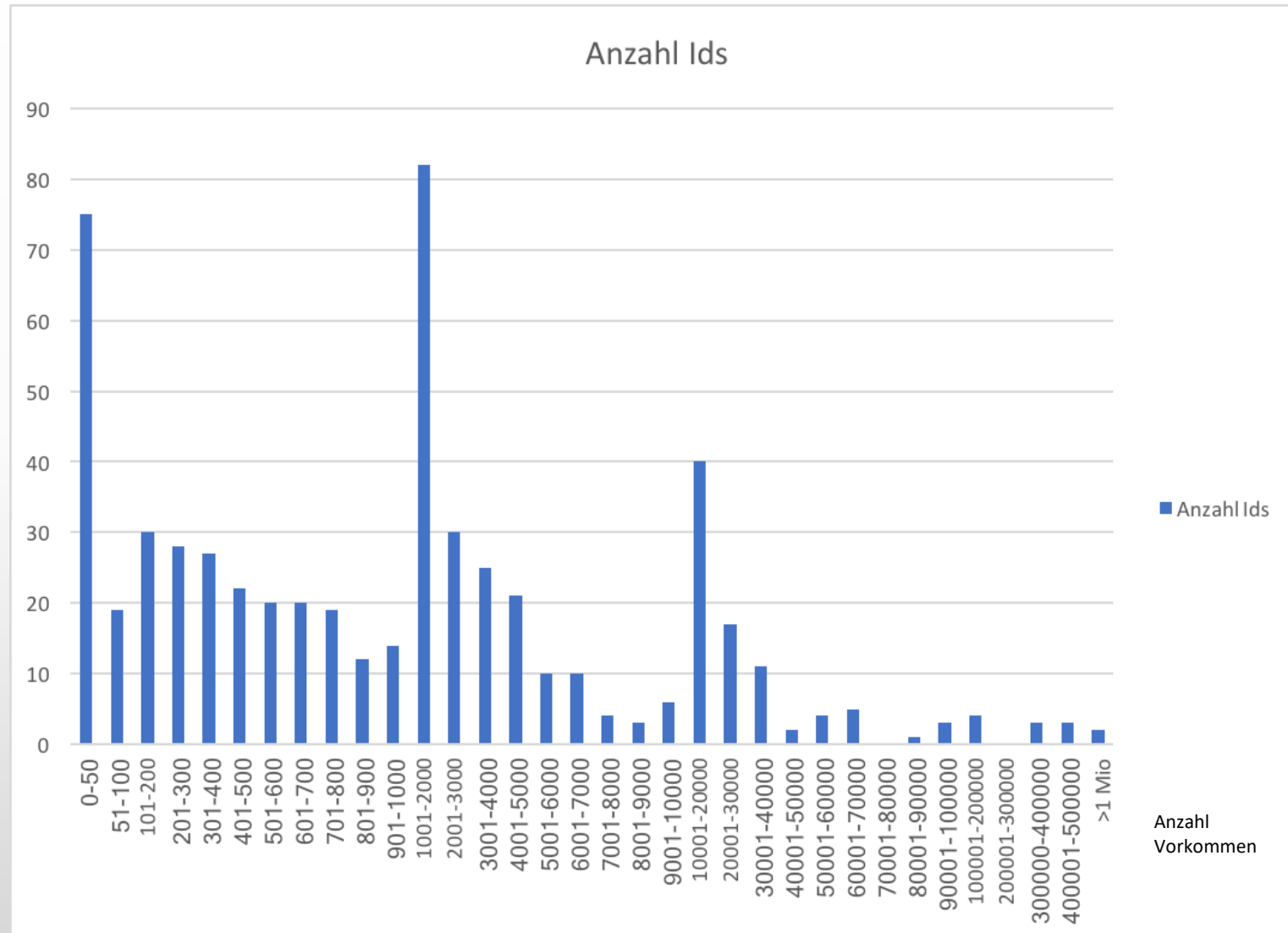
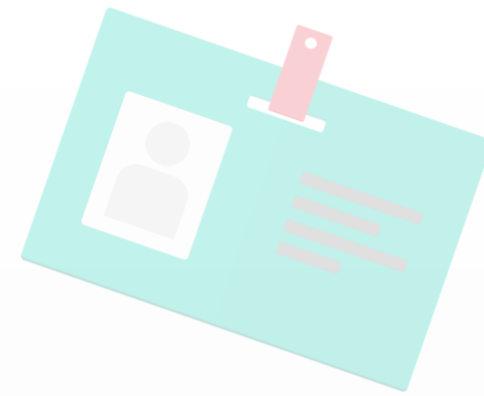


Die 20 am häufigsten verwendeten IDs



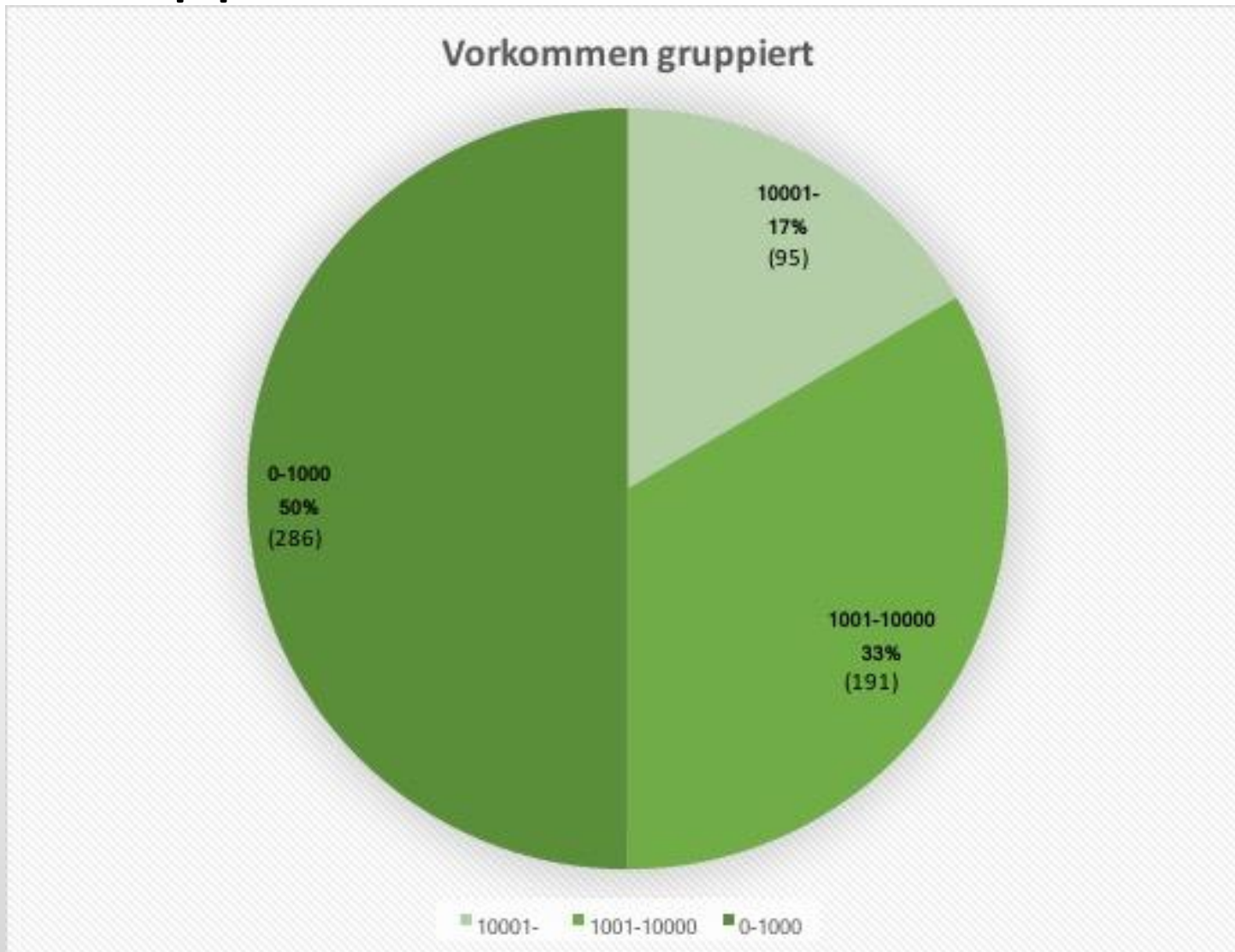
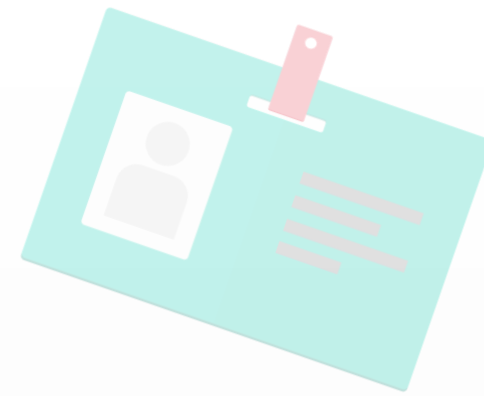
Auf der Basis von
https://www.wikidata.org/wiki/Wikidata:List_of_properties/Person/Authority_control. (Stand: 15.05.2017 / Zahlen: Stand: 28.05.17)

Gruppierte Anzahlen des Vorkommens



Auf der Basis von
https://www.wikidata.org/wiki/Wikidata:List_of_properties/Person/Authority_control. (Stand:
15.05.2017/ Zahlen:
Stand: 28.05.17)

Gruppierte Anzahlen des Vorkommens II



Auf der Basis von
https://www.wikidata.org/wiki/Wikidata:List_of_properties/Person/Authority_control. (Stand: 15.05.2017/ Zahlen: Stand: 28.05.17)

Datensätze über Menschen

Menschen in Wikidata:

[SPARQL:](#)

```
SELECT (COUNT (DISTINCT ?per)
AS ?count)
WHERE
{
?per wdt:P31 wd:Q5.
}
```

→**3.475.229**

(Stand: 04.06.2017, 10:05)

Summe aller Anzahlen des
Vorkommens:

→**8.372.067**

(Stand: 28.05.2017)

Anzahl Menschen, die keinen
Identifizierer haben

?



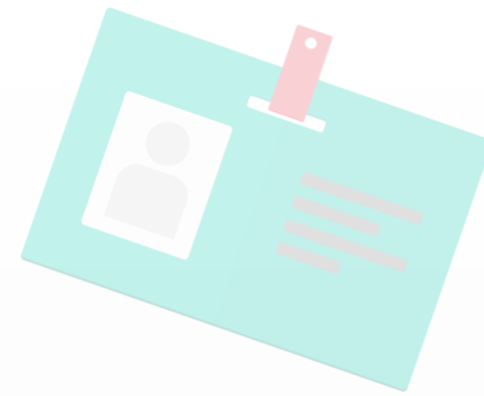
Übersichten nach Art I

- [SPARQL](#)

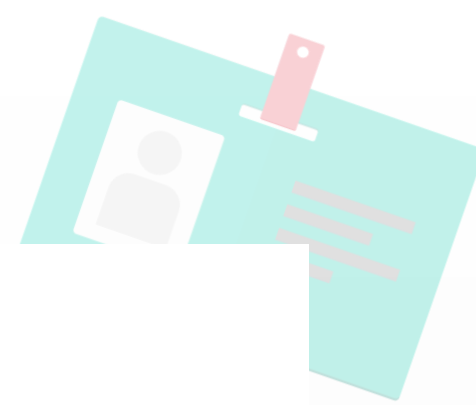
```
SELECT ?prop ?propLabel ?propDescription ?kos ?kosLabel ?kostype ?kostypeLabel
WHERE
{
  ?prop wdt:P31 wd:Q19595382 .

  OPTIONAL {?prop wdt:P1629 ?kos.
            ?kos wdt:P31 ?kostype}.

  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en"
  }
}
ORDER by ?kostypeLabel
```

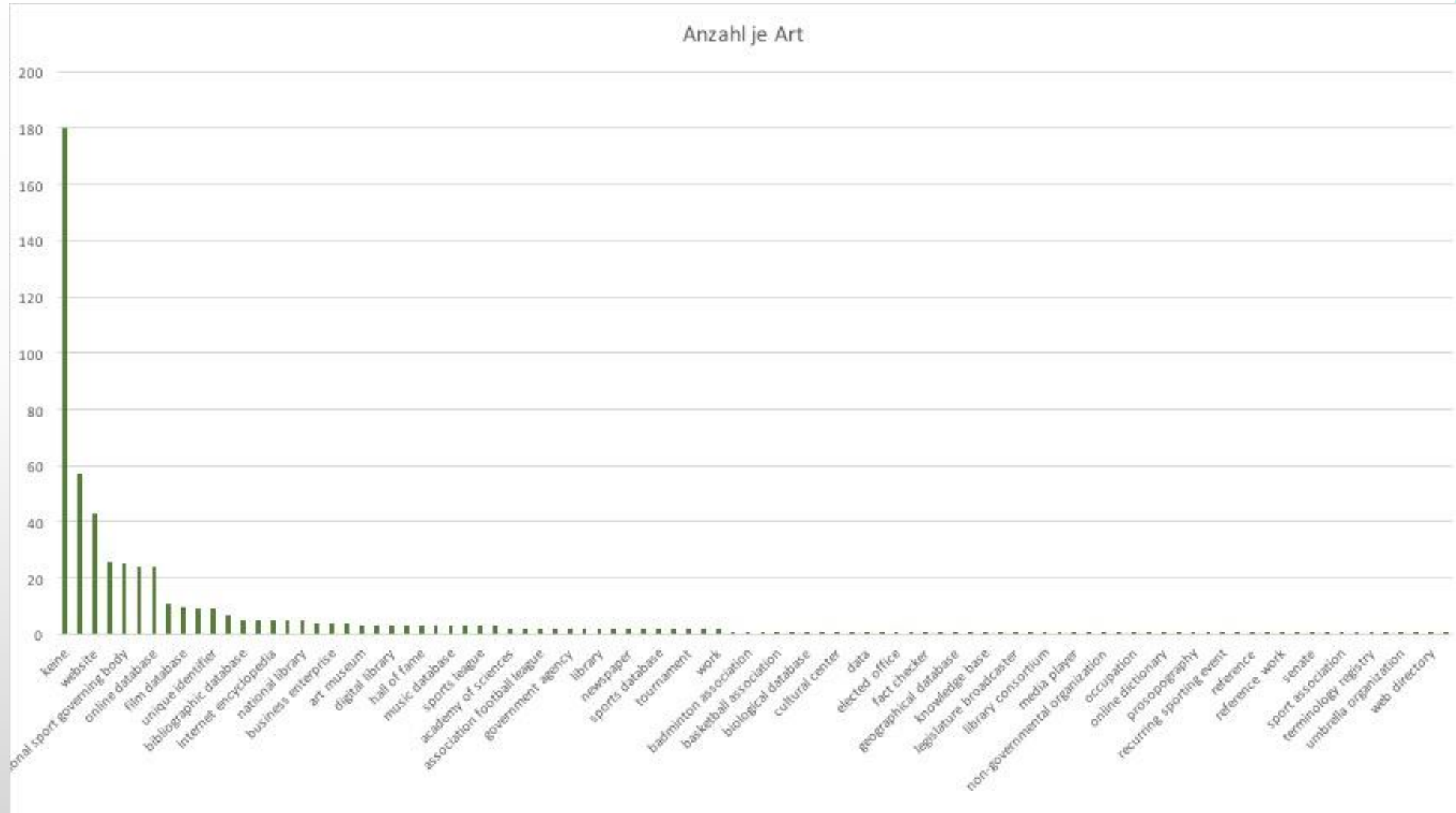
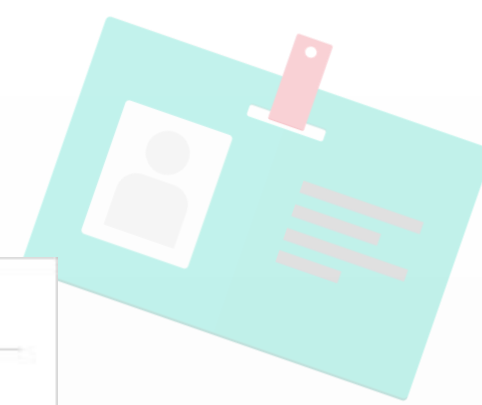


Übersichten nach Art II



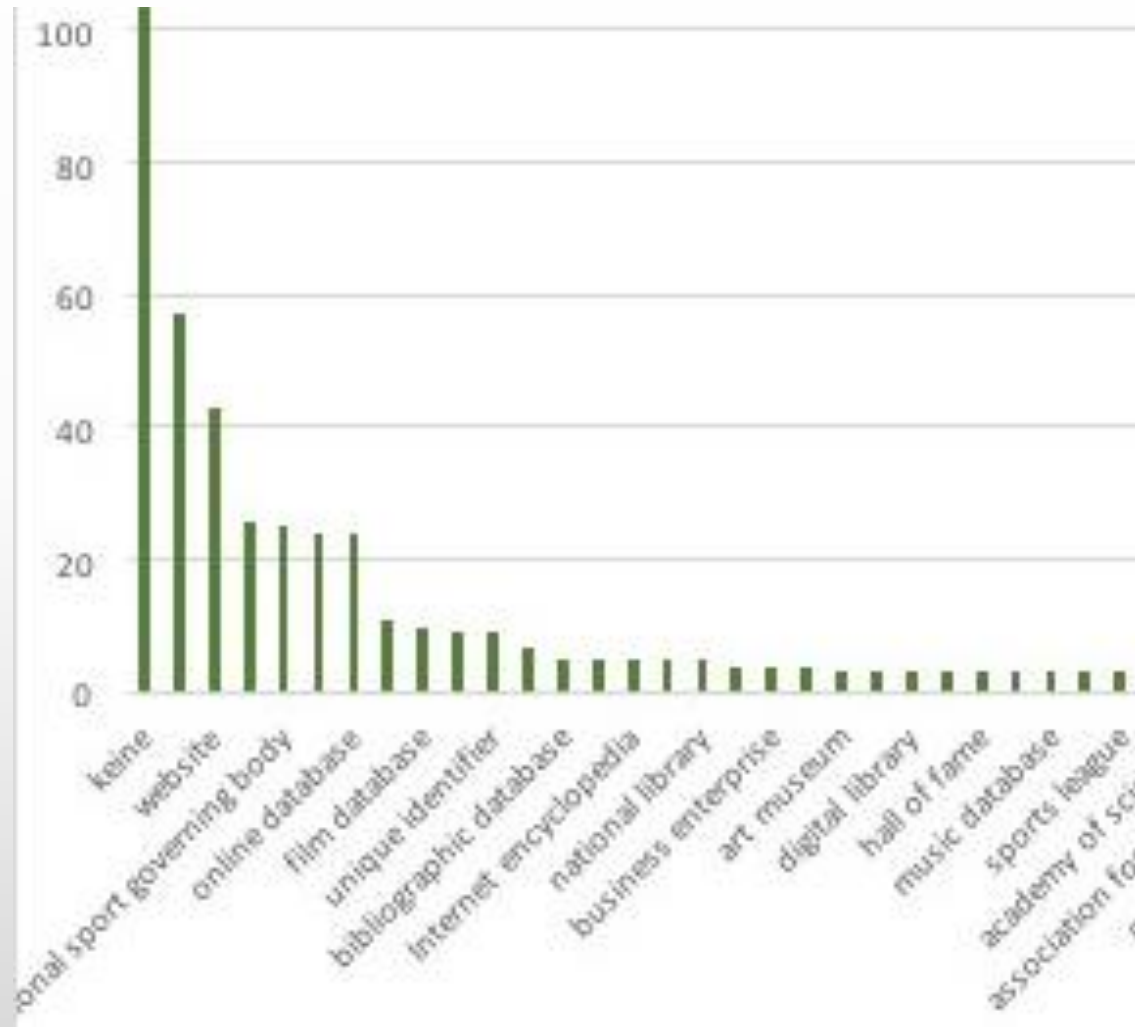
Wortwolke: erstellt mit wordart.com

Übersichten nach Art III



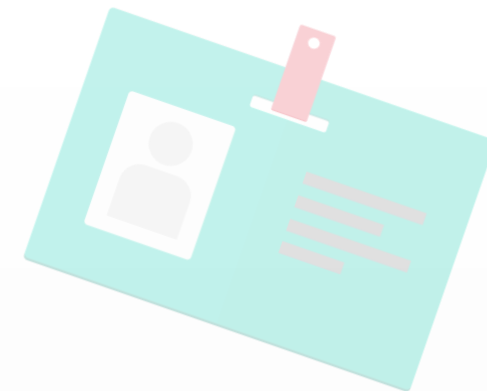
Auf der Basis von https://www.wikidata.org/wiki/Wikidata:List_of_properties/Person/Authority_control. Ergänzt um Art, Gruppierung nach Art. (Stand: 28.05.2017)

Übersichten nach Art IV



Vergrößerung
eines Ausschnitts
der
vorhergehenden
Folie

Anzahlen von ID-Kombinationen ermitteln



- [SPARQL:](#)

```
SELECT (COUNT (?per) AS ?count)
WHERE
{
  ?per wdt:P31 wd:Q5.
  ?per wdt:P227 ?GND_ID.
  ?per wdt:P214 ?VIAF.
}
```

→ **393.485**

GND: 395.201

VIAF: 895.508

(Stand: 05.06.2017)

Anzahlen von ID-Kombinationen ermitteln



- SPARQL:

```
SELECT (COUNT (?per) AS ?count)
WHERE
{
  ?per wdt:P31 wd:Q5.
  ?per wdt:P227 ?GND_ID.
  ?per wdt:P214 ?VIAF.
}
```

→ **393.485**

GND: 395.201

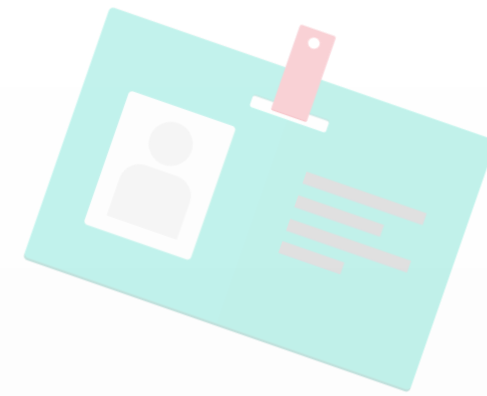
VIAF: 895.508

(Stand: 05.06.2017)

- PYTHON

```
import itertools
from SPARQLWrapper import SPARQLWrapper, JSON
if __name__ == '__main__':
    sparql = SPARQLWrapper("https://query.wikidata.org/sparql")
    sparql.setReturnFormat(JSON)

    idList = ["wdt:P1017", "wdt:P1255", "wdt:P1315", "wdt:P1463"]
    idCombs = [t for t in itertools.combinations(idList, 2)]
    for idx, idy in idCombs:
        sparql.setQuery("""
        SELECT (COUNT (?item) AS ?count)
        WHERE
        {
          ?item wdt:P31 wd:Q5.
          ?item "" + idx + "" ?x .
          ?item "" + idy + "" ?y .
        }
        """)
        ret = sparql.query().convert()
        print("%s times %s and %s"
              % (ret["results"]["bindings"][0]["count"]["value"], idx, idy))
```

IDs ähnlicher Art kombinieren

- [SPARQL:](#)

```
SELECT ?prop ?propLabel ?propDescription ?kos ?kosLabel ?kostype ?kostypeLabel ?y
WHERE
{
  ?prop wdt:P31 wd:Q19595382 .
  ?prop wdt:P1629 ?kos.
  ?kos wdt:P31 ?kostype.
  ?kostype rdfs:label ?y.
  FILTER (lang(?y) = "en")
  FILTER regex (str(?y),"data")
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
ORDER by ?y
```

→ **167** (Stand: 05.06.2017)

Kombinationen: Label der Art enthält „data“



Kombinationen				
67958	P2019	AllMovie artist ID	P345	IMDb ID
58206	P2435	PORT person ID	P345	IMDb ID
50899	P345	IMDb ID	P2168	SFDb person ID
46251	P2435	PORT person ID	P2019	AllMovie artist ID
41300	P2019	AllMovie artist ID	P2168	SFDb person ID
34815	P2168	SFDb person ID	P2435	PORT person ID
34766	P2387	Elonet person ID	P345	IMDb ID
34061	P2639	Filmportal ID	P345	IMDb ID
30112	P2387	Elonet person ID	P2019	AllMovie artist ID
28366	P2387	Elonet person ID	P2168	SFDb person ID
27400	P2387	Elonet person ID	P2435	PORT person ID
22198	P2626	DNF person ID	P345	IMDb ID
21107	P2639	Filmportal ID	P2435	PORT person ID
20501	P2639	Filmportal ID	P2019	AllMovie artist ID
18666	P2626	DNF person ID	P2019	AllMovie artist ID
18555	P2626	DNF person ID	P2168	SFDb person ID
17893	P2639	Filmportal ID	P2168	SFDb person ID
17626	P2626	DNF person ID	P2435	PORT person ID
16474	P1953	Discogs artist ID	P345	IMDb ID
16449	P1728	AllMusic artist ID	P1953	Discogs artist ID

Stand: 05.06.2017

Anzahl gesamt	
486806 (245893)	IMDb ID
67690 (42976)	Discogs artist ID
67191 (67190)	AllMovie artist ID
58133 (58165)	PORT person ID
51241 (51243)	SFDb person ID
46702 (35826)	Filmportal ID
34566 (34681)	Elonet person ID
21067 (21990)	DNF person ID

Stand: 28.05./ 05.06.2017

Kombinationen: Label der Art mit Bibliotheksbezug



- SPARQL:

```
SELECT DISTINCT ?prop #?propLabel ?propDescription ?kos ?kosLabel ?kostype ?kostypeLabel ?y
WHERE
{
  ?prop wdt:P31 wd:Q19595382 .
  ?prop wdt:P1629 ?kos.
  ?kos wdt:P31 ?kostype.
  ?kostype rdfs:label ?y.
  FILTER (regex (str(?y),"bibliot") || regex (str(?y),"librar"))
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
ORDER by ?prop
```

→**13** (Stand: 05.06.2017)

→es fehlen wieder die, die nicht als Personen-Identifizier gekennzeichnet sind

→darum Ergänzung um IDs aus Nationalbibliotheken der Übersicht

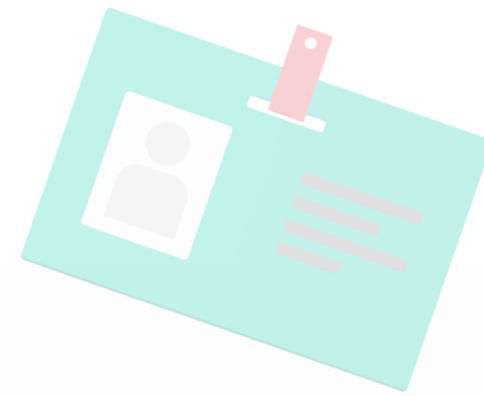
→**39**

Kombinationen: Label der Art mit Bibliotheksbezug



Kombinationen					Gesamt	
240867	wdt:P227	GND	wdt:P244	LC Auth	404204	LC Auth (Library of Congress)
225223	wdt:P268	BnF-ID	wdt:P269	SUDOC-Normdaten	318167	SUDOC-Normdaten (franz. Bibl.verbund)
208134	wdt:P244	LC Auth	wdt:P269	SUDOC-Normdaten	395198	GND (Gemeinsame Normdatei)
191174	wdt:P244	LC Auth	wdt:P268	BnF-ID	301006	BnF-ID (Franz. Nationalbibliothek)
180039	wdt:P227	GND	wdt:P269	SUDOC-Normdaten	152076	NTA-Nummer (Königl. Bibl. d. Niederlande)
162351	wdt:P227	GND	wdt:P268	BnF-ID	99781	Open Library ID
128945	wdt:P244	LC Auth	wdt:P1006	NTA-Nummer	52435	CERL-ID (Zus.Schluss von Forschungsbibl. in Europa)
119301	wdt:P1006	NTA-Nummer	wdt:P269	SUDOC-Normdaten	43159	NLI (Israel. Nationalbibliothek)
118137	wdt:P227	GND	wdt:P1006	NTA-Nummer		
103901	wdt:P268	BnF-ID	wdt:P1006	NTA-Nummer		
79420	wdt:P648	Open Library ID	wdt:P244	LC Auth		
63068	wdt:P648	Open Library ID	wdt:P269	SUDOC-Normdaten		
61322	wdt:P648	Open Library ID	wdt:P227	GND		
53535	wdt:P648	Open Library ID	wdt:P268	BnF-ID		
51940	wdt:P227	GND	wdt:P1871	CERL-ID		
47168	wdt:P648	Open Library ID	wdt:P1006	NTA-Nummer		
38947	wdt:P244	LC Auth	wdt:P949	NLI (Israel)		
34555	wdt:P227	GND	wdt:P949	NLI (Israel)		
33109	wdt:P269	BnF-ID	wdt:P949	NLI (Israel)		
31045	wdt:P244	LC Auth	wdt:P1871	CERL-ID		

Fragen?



Quellen

- Abbildung auf allen Folien:

janjf93. (2016). *Name nameplate shield id press*. pixabay. Lizenziert unter <https://creativecommons.org/publicdomain/zero/1.0/>. Abgerufen am 27.05.2017 von <https://pixabay.com/en/name-nameplate-shield-id-press-1714231/>.

