Automatisierte Verfahren zur Datenanreicherung

Angela Vorndran
Team Datenmanagement, Deutsche Nationalbibliothek

DEUTSCHE NATIONAL BIBLIOTHEK

Hintergrund

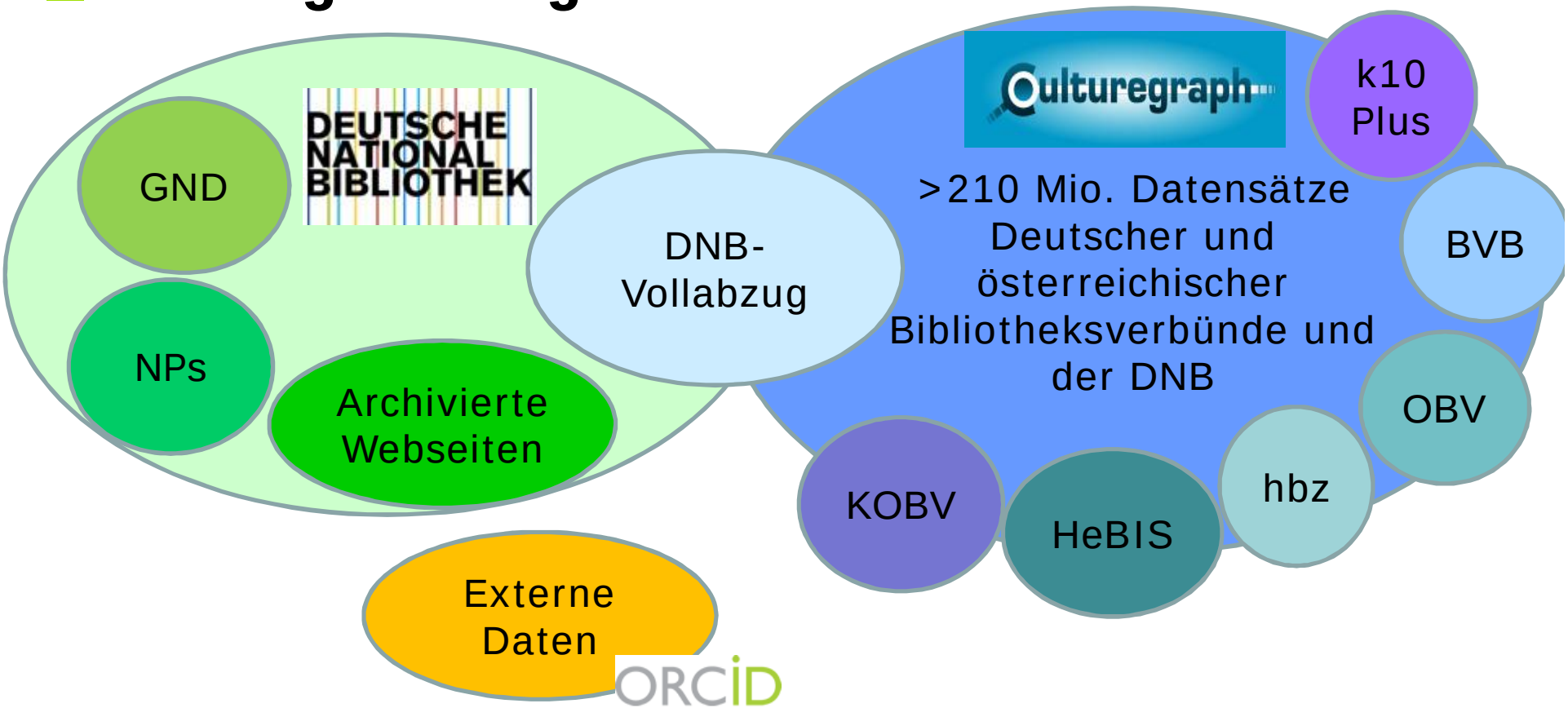
- Im Bestand der DNB befinden sich über 43 Millionen Medien und mittlerweile mehr als 10 Millionen Netzpublikationen
- Im Schnitt kommen jeden Tag 5.800 neue NPs dazu
- Maschinelle Lösungen sind unvermeidbar, um die Metadaten v.a. der Netzpublikationen möglichst auch mit GND-Verknüpfungen zu erschließen und ihre Qualität damit zu verbessern
- Ein Lösungsansatz: Fremddatenübernahme

DEUTSCHE NATIONAL BIBLIOTHEK

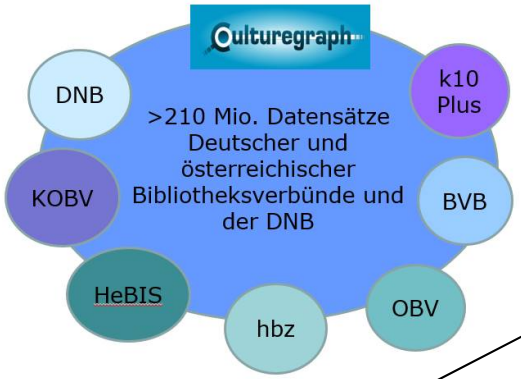
Gliederung

1. Werkbündelung
2. Personen mit Normdatenverknüpfungen anreichern
3. Datenabgleich mit externen Informationsquellen zur Datenanreicherung, hier: ORCiD

Datengrundlagen



Werkbündelung



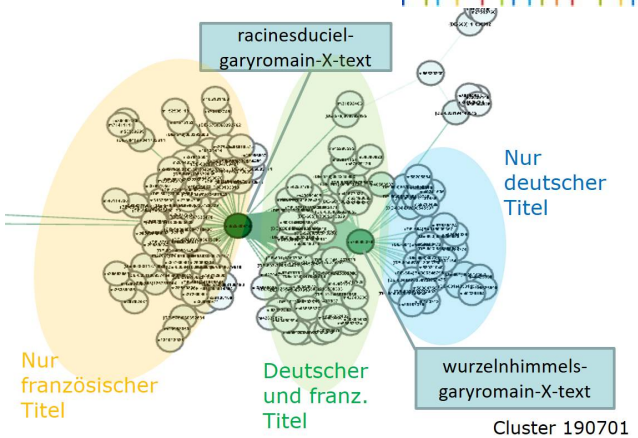
Datengrundlage

Für jeden Datensatz Schlüssel für den Abgleich erstellen

oclc961837268-X-meeralsversprechen-text

9783525253243-X-meeralsversprechen-text

meeralsversprechenbedeutungufunktionvonseeherrschaftbeithukydides-kopphans-X-text



Breitensuche vereinigt Datensätze mit mind. einem identischen Schlüsseln in einem Bündel

DEUTSCHE NATIONAL BIBLIOTHEK

Titel im Culturegraph-Bestand abgleichen

- Abgleich der Publikationen durch Schlüssel
- Nach definierten Regeln werden relevante Datenelemente aus dem MARC-Format extrahiert und kombiniert
- Datensätze mit gleichem Schlüssel werden zu Bündeln zusammengefasst
- Potenzial zur Übertragung von
  - Sacherschließungsdaten
  - Anreicherung durch Normdatenverknüpfungen

  von einem Bündelmitglied auf ein anderes, das darüber nicht verfügt

DEUTSCHE NATIONAL BIBLIOTHEK

Hauptsachtitel	Das Meer Als Versprechen
Zusatz	Bedeutung Und Funktion Von Seeherrschaft Bei Thukydides
Inhalt	Inhaltsverzeichnis Inhaltstext Inhaltstext
Person	aut Kopp, Hans 1127785095
Körperschaft	pbl Vandenhoeck & Ruprecht
Umfang	303 Seiten
Erscheinungsjahr	2017
Material	book
Erscheinungsort	Göttingen
Herausgeber	Vandenhoeck & Ruprecht
Schlagwort	655 Hochschulschrift 4113937-9 650 Seeherrschaft 4077302-4 600 Thucydides 4138063-0
Klassifikation	DDC 355 DDC 880 DDC 930
Standard-Identifier	OCLC 961837268 EKI DNB1115738097
Verlags-Identifier	GTIN14 9783525253243 ISBN13 9783525253243

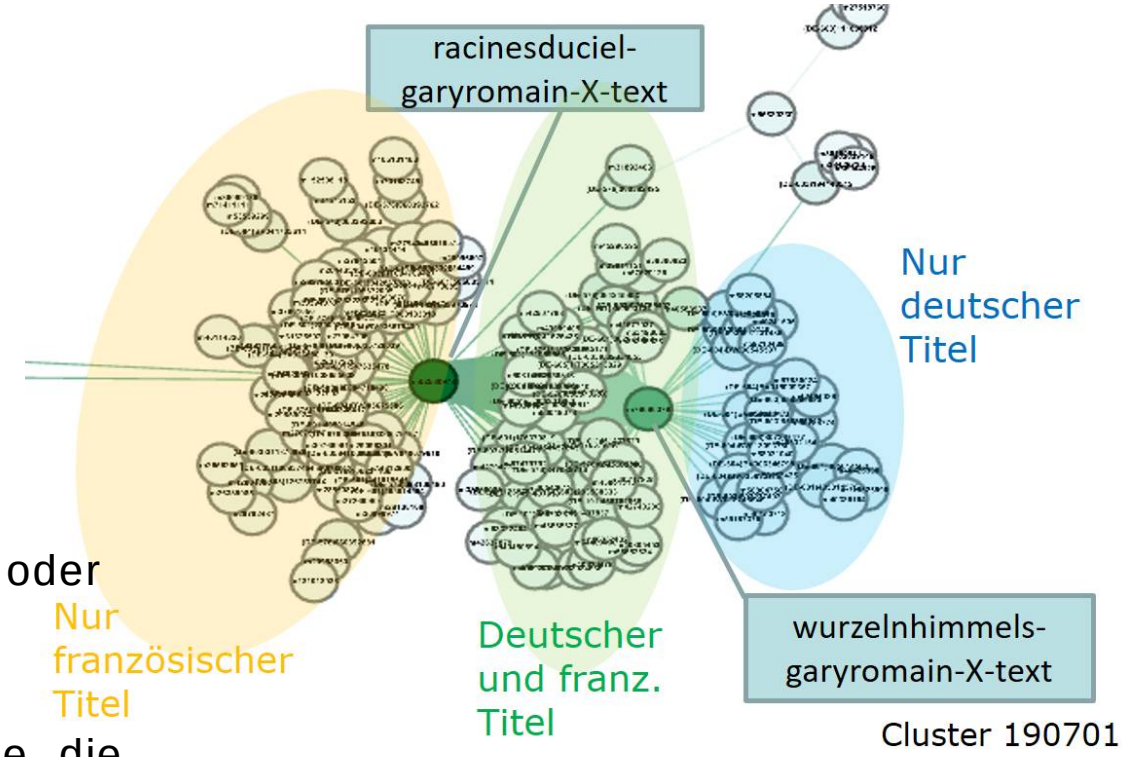
Titel+Titelzusatz-ErstellerIn-Band-Publikationstyp:
meeralsversprechenbedeutungufunktionvonseeherrschaftbeithukydides-kopphans-X-text

Kontrollnummer-Band-Titel-Publikationstyp:
oclc961837268-X-meeralsversprechen-text

Kontrollnummer-Band-Titel-Publikationstyp:
DNB1115738097-X-meeralsversprechen-text

ISBN-Band-Titel-Publikationstyp:
9783525253243-X-meeralsversprechen-text

Werkbündel

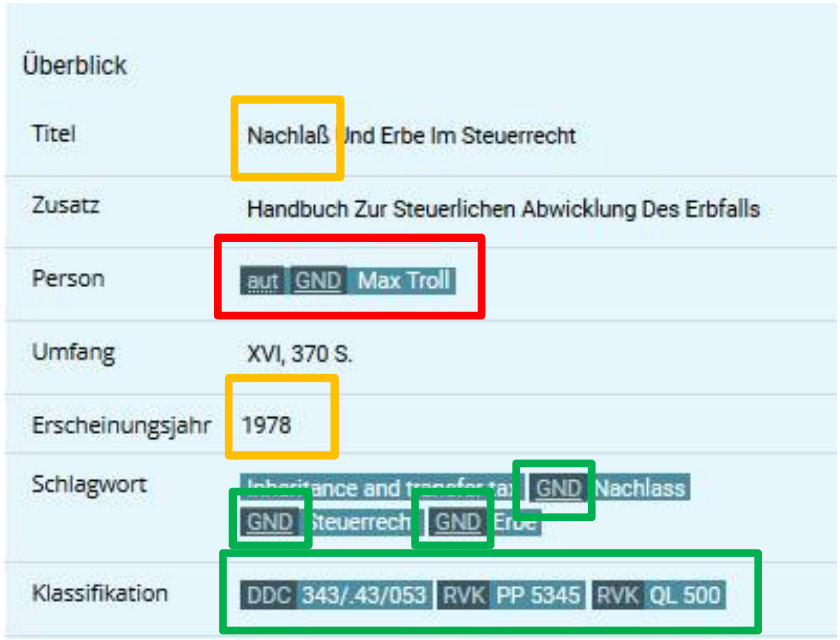


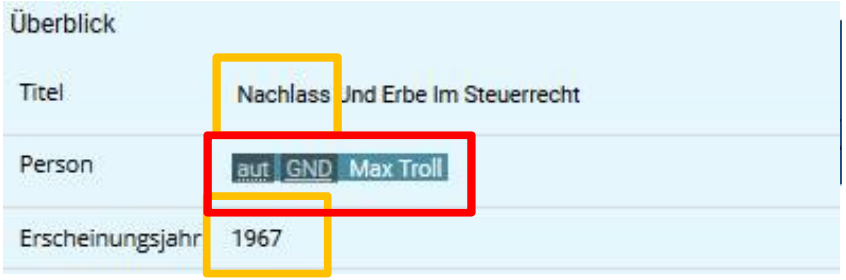
- Für jeden Datensatz liegen ein oder mehrere Schlüssel vor
- Bündel vereinen alle Datensätze, die mindestens einen gemeinsamen Schlüssel mit einem anderen Mitglied besitzen

DEUTSCHE NATIONAL BIBLIOTHEK

Statistik Werkbündel (Stand 6/2022)

- 132.880.199 Datensätze wurden verarbeitet (nicht behandelt: unselbständige Teile von mehrbändigen Werken, fortlaufende Ressourcen etc.)
- 21.295.838 Bündel mit mehr als einem Mitglied
- Darin enthalten 101.948.360 Datensätze
- Durchschnittliche Bündelgröße (wenn mehr als ein Mitglied): 4,8


Überblick
Titel
Nachlaß Und Erbe Im Steuerrecht
Zusatz
Handbuch Zur Steuerlichen Abwicklung Des Erbfalls
Person
aut GND Max Troll
Umfang
XVI, 370 S.
Erscheinungsjahr
1978
Schlagwort
GND Nachlass
GND GND
Klassifikation
DDC 343/.43/053 RVK PP 5345 RVK QL 500


Überblick
Titel
Nachlass Und Erbe Im Steuerrecht
Person
aut GND Max Troll
Erscheinungsjahr
1967


Überblick
Titel
NACHLASS UND ERBE IM STEUERRECHT. HANDBUCH ZUR STEUERLICHEN ABWICKLUNG DES ERBFALLS. NACHTRAG NACH D. STANDE VOM 1. JANUAR 1970. VON MAX TROLL. MUENCHEN (U.A.): BECK 1967-1970. 27, XII, 364 S.
Person
aut GND Max Troll
Erscheinungsjahr
1967


Überblick
Titel
Nachlass Und Erbe Im Steuerrecht
Zusatz
Handbuch Zur Steuerl. Abwicklung D. Erbfalls
Person
aut GND Max Troll
Umfang
XVI, 370 S.
Erscheinungsjahr
1978
Schlagwort
Erbfall Steuerrecht Steuer Nachlaß Erbrecht

Mögliche Anreicherungen

Link zu diesem Datensatz	http://d-nb.info/820108170
Titel	Tage und Nächte steigen aus dem Strom : e. Donaufahrt Georg Lentz
Person(en)	Buchheim, Lothar-Günther (Verfasser)
Ausgabe	1. Aufl.
Verlag	[München] : Goldmann
Zeitliche Einordnung	Erscheinungsdatum: 1981
Umfang/Format	269 S. ; 18 cm
ISBN/Einband/Preis	978-3-442-06343-7 kart. : DM 6.80 3-442-06343-4 kart. : DM 6.80
Beziehungen	Ein Goldmann-Taschenbuch ; 6343
Anmerkungen	Lizenz d. Langen-Müller-Verl., München, Wien
Schlagwörter	Donau
Sachgruppe(n)	61 Geographie, Heimat- und Länderkunde, Reisen

Schlagwortfolge	Donau Bootsfahrt Reisebericht 1938 Buchheim, Lothar-Günther Autobiographie
Einzelschlag-wörter (GND)	4049275-8 Reise 4015701-5 Europa 4263903-7 Flusswandern
Schlagwörter aus anderen Thesauri	Donau Reisebeschreibungen Danube River Description and travel Autobiographie 1941 Donau – Bootsfahrt – Reisebericht Buchheim, Lothar-Günther
Klassifikation	BKL 21.30 BKL 74.05 SFB ERD 242 RVK GN 9999 BKL 18.10 BKL 17.97

IE-Übernahme aus Werkbündeln

- Ziel: Werkbündel nutzen, um inhalts- und formalerschließende Metadaten von anderen Bündelmitgliedern zu übernehmen
- Schlagwörter und Klassifikation retrospektiv in Fremddatenfelder übernehmen wenn keine intellektuelle Erschließung vorhanden ist
- Ausgewählte Thesauri und Klassifikationssysteme
- Für Neuzugänge alles übernehmen

Gliederung

1. Werkbündelung
2. Personen mit Normdatenverknüpfungen anreichern
3. Datenabgleich mit externen Informationsquellen zur Datenanreicherung, hier: ORCiD

Übernahme von Verknüpfungen aus Culturegraph-Werkbündeln

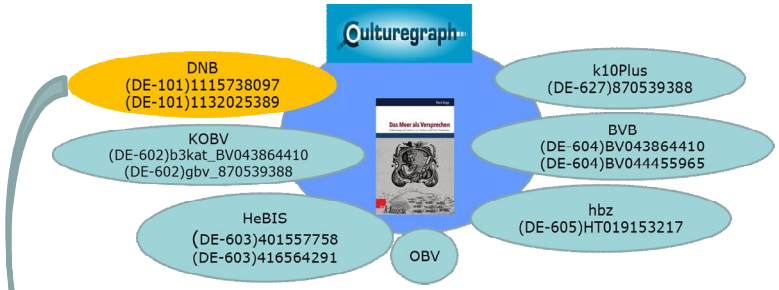
Hauptsachtitel	Das Meer Als Versprechen
Zusatz	Bedeutung Und [illegible]ion Von Seeherrschaft I
Person	aut Kopp, Hans
Körperschaft	-
Umfang	1 Online-Ressource (303 Seiten)
Erscheinungsjahr	2017
Material	file
Erscheinungsort	Göttingen
Herausgeber	Vandenhoeck & Ruprecht
Standard-Identifier	DOI 10.13109/9783666253249 EKI 1002269444 OCLC vandP1096134
Verlags-Identifier	-

Anreicherung

[illegible]titel	Das Meer Als Versprechen
[illegible]z	Bedeutu[illegible]d Funktion Von Seeherrschaft Bei Thukydides
[illegible]	Inhaltsverzeichni[illegible]haltstext Inhaltstext
Person	aut Kopp, Hans 1127785095
Körperschaft	-
Umfang	303 Seiten
Erscheinungsjahr	2017
Material	book
Erscheinungsort	Bristol, CT
Herausgeber	Vandenhoeck & Ruprecht
Schlagwort	655 Hochschulschrift 4113937-9 650 Seeherrschaft 4077302-4 600 Thucydides 4138063-0
Klassifikation	DDC 938.05 RVK EH 26175 RVK NH 5850

DEUTSCHE NATIONAL BIBLIOTHEK

Ablauf beim Abgleich von Personen



Personen mit Tp-ID
Personen ohne ID

Prüfen, welche Personen in DNB-Titeldatensätzen keine Tp-Verknüpfungen enthalten

metafacture

1127785095

1127785095

Für diese Personen im selben Bündel namensgleiche Personen mit Tp-Verknüpfungen suchen

python

Hauptsachtitel	Das Meer Als Versprechen
Zusatz	Bedeutung Und Funktion Von Seeherrschaft Bei Thukydides
Person	aut Kopp, Hans 1127785095
Körperschaft	-
Umfang	1 Online-Ressource (303 Seiten)
Erscheinungsjahr	2017
Material	file
Erscheinungsort	Göttingen
Herausgeber	Vandenhoeck & Ruprecht
Standard-Identifier	DOI 10.13109/9783666253249 EKI 1002269444 OCLC vandP1096134
Verlags-Identifier	-

Tp-ID im Titeldatensatz ergänzen

Verknüpfung erstellen

CBS/ Elvis

Erfolgte Normdatenanreicherung

Ausgabe

1021053295, (DE-101)962759643, Petersen§ Holger, 100
Tp-ID, IDN Titeldatensatz, Name, Feld

- 7,5 Mio. DNB-Personen in Werkbündeln ohne Tp-Verknüpfung
- Ca. 1,4 Mio. Tp-Anreicherungen in 1,2 Mio. Datensätzen ermittelt (19 %)
- Im Sommer 2021 Anreicherungen vorgenommen

DEUTSCHE NATIONAL BIBLIOTHEK

Manifestationsbündelung

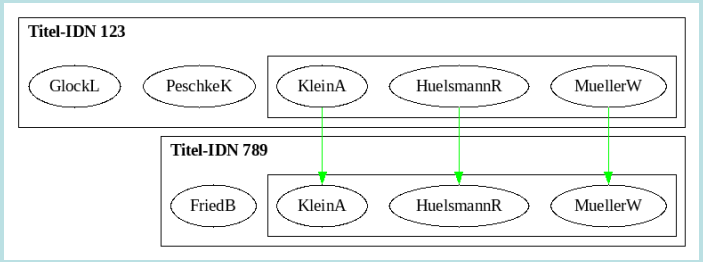
- Testen der Übernahme von Metadatenelementen von anderen Manifestationen (=gleiche Ausgabe eines Werkes)
- Exaktere Übereinstimmung einer Publikation notwendig als bei Werken
- Übernahme von Autor*innen und ggf. beteiligten Personen
- Auffinden im Bestand fehlender Publikationen

DEUTSCHE NATIONAL BIBLIOTHEK

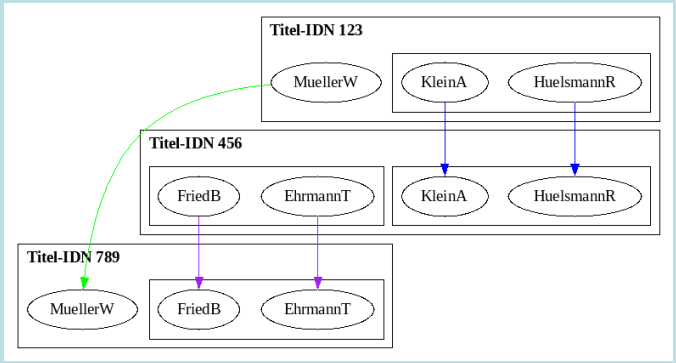
Nutzung von Ko-Autor*innen-Beziehungen

- Wissenschaftskommunikation erfolgt in den meisten Fällen kollaborativ
- Forschende aus derselben Domäne (oder auch interdisziplinär), aber u.U. aus unterschiedlichen Institutionen arbeiten über längere Zeiträume immer wieder zusammen und veröffentlichen gemeinsam

Ko-Autorschaften auswerten



Unmittelbare Ko-Autorschaft



Mittelbare Ko-Autorschaft

- Bei selben Namen können vorhandene GND-Verknüpfungen übertragen werden
- Bei 2 identischen Ko-Autoren Anreicherungen bei 276.000 Personen im DNB-Bestand möglich

Gliederung

1. Werkbündelung
2. Personen mit Normdatenverknüpfungen anreichern
3. Datenabgleich mit externen Informationsquellen zur Datenanreicherung, hier: ORCiD

DEUTSCHE NATIONAL BIBLIOTHEK

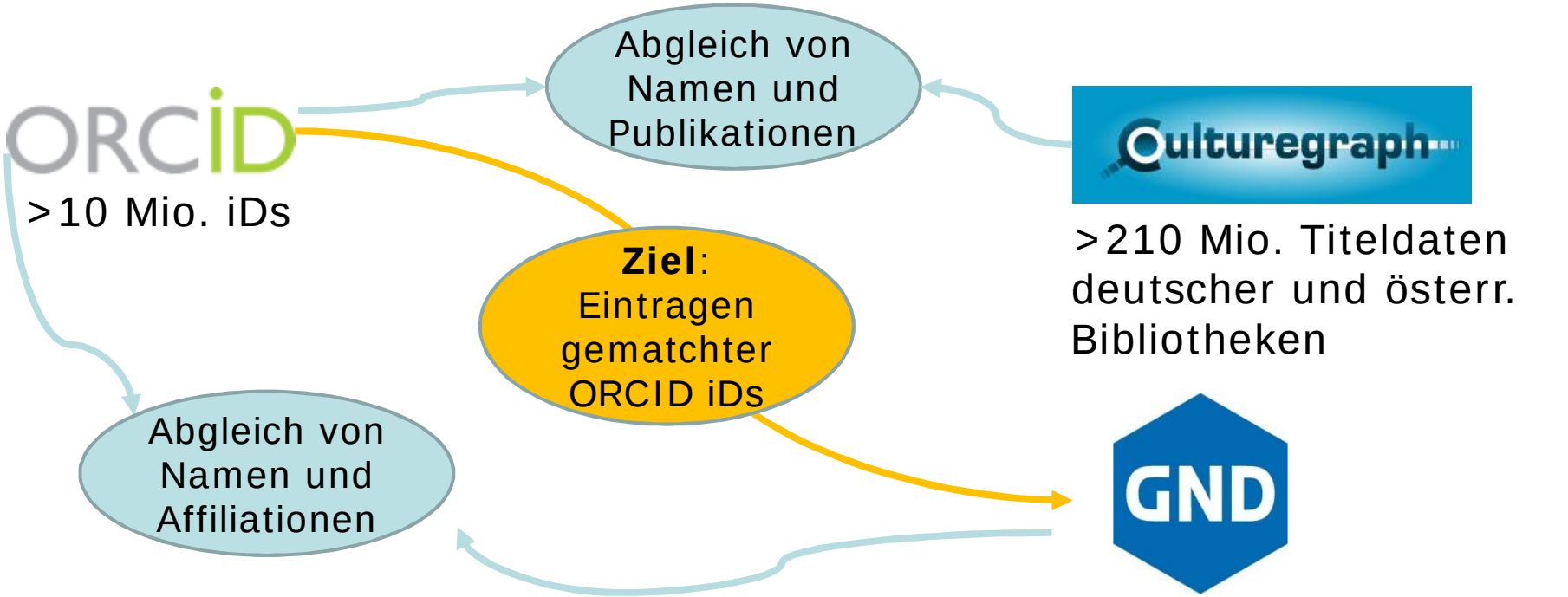
ORCID DE

Förderung der Open Researcher and Contributor ID in Deutschland

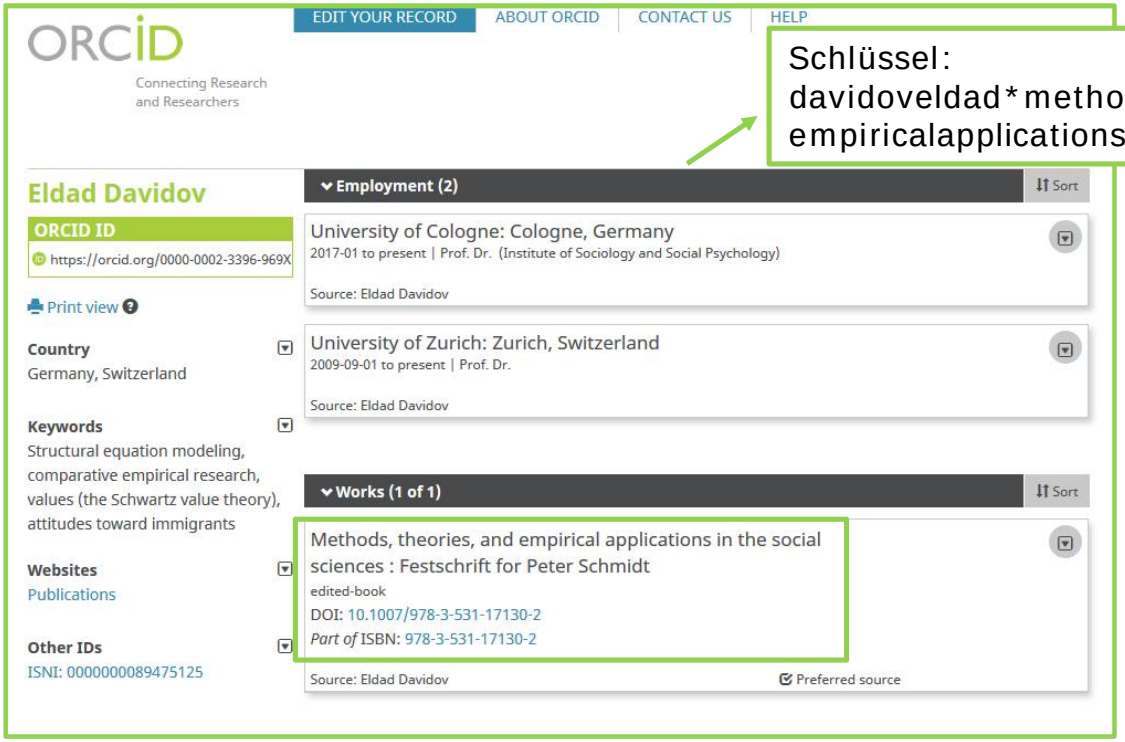
ORCiD

- Open Researcher and Contributor ID
- Eindeutige Kennung zur Identifikation einer Person
- Wissenschaftler*innen erstellen Record mit der Möglichkeit, Informationen zu Werdegang, Publikationen, u.a. zu hinterlegen

Verfügbare Ressourcen



Abgleich von ORCiD und GND-Sätzen



Schlüssel:
davidoveldad*methodstheoriesand
empiricalapplicationsinsoc

DEUTSCHE NATIONAL BIBLIOTHEK

Schlüssel:
davidoveldad*methodstheoriesand
empiricalapplicationsinsoc

GND

```
005 Tp3
006 http://d-nb.info/gnd/1022834061$zhttp://d-nb.info/gnd/129032956
008 piz
011 f
012 v
024 orcid: 0000-0002-3396-969X$vHerkunft: orcid
035 gnd/1022834061
039 gnd/129032956
039 pnd/129032956$vzg
043 XA-DE;XA-CH
100 Davidov, Eldad
510 !004815750!Universität zu Köln [Tb1]$4affi
510 !000361909!Universität Zürich [Tb1]$4affi
548 1971$4datl
550 !040559300!Soziologe [Ts1]$4berc
```

GND-ID:
1022834061

Wenn GND-ID
vorhanden,
Eintrag in Tp-Satz

Culturegraph

Titel	Methods, Theories, And Empirical Applications In The Social Sciences
Zusatz	Festschrift For Peter Schmidt
Person	edt GND Samuel Salzborn edt GND Eldad Davidov edt GND Rainer Reinecke hnr GND Peter Schmidt
Umfang	351 S.
Erscheinungsjahr	2012
Schlagwort	GND Rosenstock-Huessy, Eugen GND Empirische Sozialforschung GND Methodologie GND Politische Wissenschaft GND Sozialpsychologie GND Empirische Sozialforschung Aufsatzsammlung Aufsatzsammlung
Klassifikation	DDC 300 DDC 301.0723

DEUTSCHE NATIONAL BIBLIOTHEK

Standardidentifier

- Herstellen von Verknüpfungen zur GND via gelieferte ORCiDs, ISNIs, etc.
- Einbringen von weiteren ORCiDs in Titel- und GND-Datensätze aus anderen Quellen (DNB-Claiming-Logs, BASE-Claiming-Logs, ORCiD-Werke-Dump)
- Jede Übernahme/Verknüpfung wird zusätzlich geprüft (z.B. stimmen Name im Titel- und GND-Satz überein, passt der geclaimte Name zum ORCiD-Record)

DEUTSCHE NATIONAL BIBLIOTHEK

Personen-Vorschlagssätze mit Informationen aus ORCiD

- Automatisierte Erstellung von Vorschlägen für neue Personennormdatensätze
- Informationen aus dem Titeldatensatz und dem ORCiD-Record
  - Andere Standardnummern für Personen (z.Zt. ISNI)
  - Ländercode
  - Abweichende Namensformen
  - Affiliation
  - Keywords
  - Wirkungsdaten
  - Akademischer Grad
- Vorschläge können perspektivisch aus verschiedenen Quellen stammen

Beispiel Vorschlagsdatensatz Person

Titel

Eingabe: 1140:05-09-17 Änderung: 9999:20-08-20 12:35:34 Status: 1140:05-09-17
0500 Oaf
0501 Text$btxt
0502 Computermedien$bc
0503 Online-Ressource$bcr
0550 Webformular
0551 X$bm
0598 NPL004
0600 ro;rb
1100 2016
1101 cr
1500 /1ger
1700 /1XA-DE-NW
2050 urn:nbn:de:101:1-2017090535307
2105 17,O10
2150 NPSozialer Zusammenhalt in Bremen: Anlage 1 Codebuch
2198 535658540
2240 DNB:1139306340
2242 |o|1005511653
3000 Arant, Regina$BVerfasser$4autEaHnpi$D2020-06-12
3010 Larsen, Mandi$BVerfasser$4aut$y(orcid)0000-0001-5057-0085EaHorcid$D2020-08-20
3010 Boehnke, Klaus$BVerfasser$4autEaHnpi$D2020-06-12
3119 Bertelsmann Stiftung$BSonstige$4oth
4000 Sozialer Zusammenhalt in Bremen : Anlage 1 Codebuch / Regina Arant, Mandi Larsen, Klaus Boe
Bertelsmann Stiftung
4030 Gütersloh : Bertelsmann Stiftung
4060 Online-Ressource
4083 =A $
4085 *HTTP*=q pdf=u
http://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/codebuch-sozialer-zusammenhalt-
H=z LF
4233 aaa5DE-101
5050 300EmHaepsg$K0,966$D2017-09-06
5050 300EaHnpi$D2017-09-05
5051 $KK_A8_03_20160930_de
5052 $f360$F0,952$g390$G0,839$D2017-09-06
5450 [noScheme]nbbre
5550 [ckw]!040081354!*Bremen [Tg f]*EaHstwgnd$K1$D2019-04-25

Vorschlagsdatensatz

Eingabe: 9999:04-02-22 Änderung: 9999:04-02-22 17:00:52 Status: 9999:04-02-22
005 TX
008 piz
011 t;f
024 orcid: 0000-0001-5057-0085$vHerkunft: orcid
043 DE;US
100 Larsen, Mandi
400 Larsen, Mandi M.$vName aus Titel
510 Jacobs University Bremen gGmbH$gBremen, Bremen, DE$4affi
510 Universitätsklinikum Hamburg-Eppendorf$gHamburg, Hamburg, DE$4affi
510 Safe Horizon$gNew York, New York, US$4affi
510 Columbia University $gNew York, NY, US$4affi
548 2004$4datw
550 Dr.$4akad
550 social cohesion$4them
550 intimate partner violence$4them
550 health inequalities$4them
667 Vorschlags-Ranking:7$5DE-101
667 Quelle: Auswertung DNB-ORCID-Claiming-Logs; zugehöriger Titel: !1080799184!$5DE-101
667 Quelle: Auswertung DNB-ORCID-Claiming-Logs; zugehöriger Titel: !1083610511!$5DE-101
667 Quelle: Auswertung DNB-ORCID-Claiming-Logs; zugehöriger Titel: !1139306340!$5DE-101
667 Quelle: Auswertung DNB-ORCID-Claiming-Logs; zugehöriger Titel: !1140167820!$5DE-101
667 Quelle: Auswertung DNB-ORCID-Claiming-Logs; zugehöriger Titel: !1148691146!$5DE-101
667 Quelle: Auswertung DNB-ORCID-Claiming-Logs; zugehöriger Titel: !1175694444!$5DE-101
667 Quelle: Auswertung DNB-ORCID-Claiming-Logs; zugehöriger Titel: !117569505X!$5DE-101
672 Health Inequities Related to Intimate Partner Violence Against Women$bThe Role of Social Policy in the
States, Germany, and Norway$f2016$w(DE-101)1080799184
672 Health Inequities Related to Intimate Partner Violence Against Women$bThe Role of Social Policy in the
States, Germany, and
Norway$f2016$w(DE-101)1083610511$0(doi)10.1007/978-3-319-29565-7$0(urn)urn:nbn:de:1111-2016022
672 Sozialer Zusammenhalt in Bremen$bAnlage 1
Codebuch$f2016$w(DE-101)1139306340$0(urn)urn:nbn:de:101:1-2017090535307
672 Sozialer Zusammenhalt in Bremen$f2016$w(DE-101)1140167820$0(urn)urn:nbn:de:101:1-201709211
672 Sozialer Zusammenhalt in Bremen$bAnlage 2
Methodenbericht$f2016$w(DE-101)1148691146$0(urn)urn:nbn:de:101:1-2017121418157
672 What Holds Asian Societies Together?$bInsights from the Social Cohesion Radar;
Codebook$f2017$w(DE-101)1175694444$0(urn)urn:nbn:de:101:1-201901181540442919454 5
672 What Holds Asian Societies Together?$bInsights from the Social Cohesion Radar; Methods

ORCiD record

https://orcid.org/
0000-0001-5057-0085

Websites & social links
BIGSSS

Other IDs
Scopus Author ID: 7201731343
Loop profile: 1199434
Scopus Author ID: 57202743638

Keywords
social cohesion, intimate partner violence, health inequalities

Countries
Germany

Is this you? Sign in to start editing
Printable version
Name
Mandi Larsen
Also known as
Mandi M. Larsen

Activities
Collapse all
Employment (5)
Sort

Jacobs University Bremen gGmbH: Bremen, Bremen, DE
2016-10-01 to present | University Lecturer / Methods Center Coordinator (Bremen International Graduate School for Social Sciences)
Employment
Show more detail
Source: Mandi Larsen

Jacobs University Bremen gGmbH: Bremen, Bremen, DE
2015-06-01 to 2016-09-30 | Postdoctoral Researcher (Diversity Focus Area)
Employment
Show more detail
Source: Mandi Larsen

Universitätsklinikum Hamburg-Eppendorf: Hamburg, Hamburg, DE
2008-10-01 to 2010-05-31 | German Chancellor Fellow (Institute of Forensic Medicine)
Employment
Show more detail
Source: Mandi Larsen

Safe Horizon: New York, New York, US
2004-11-01 to 2008-05-31 | Research and Evaluation Associate (Research and Evaluation Department)
Show more detail

Vernetzung von Normdaten stärken

- Grundlage dieser Verfahren: Bereits vorhandene Normdaten und ihre Verknüpfungen sollen nachgenutzt werden
- Notwendig: Mehr (verknüpfte) Personennormdatensätze, z.B. auch für Autor*innen von neuen Netzpublikationen, die nicht mehr intellektuell bearbeitet werden

Danke für Ihre Aufmerksamkeit!

Dr. Angela Vorndran
Team Datenmanagement
Deutsche Nationalbibliothek
a.vorndran@dnb.de