



Angela Vorndran, Stefan Grund



Hervorholen, was in unseren Daten steckt

Mehrwerte durch Analysen großer Bibliotheksdatenbestände
(Culturegraph)



- Enthält Metadaten deutschsprachiger Bibliotheksverbände (BVB, SWB, GBV, hbz, HeBIS, KOBV, OBV) und der DNB
- >170 Mio. Datensätze, tägliches Update durch GVI-Datenlieferungen in Arbeit
- Ziel
 - Vernetzung der Datenbestände
 - Plattform für Datenanalysen (Clustering, Datenabgleich)
 - Übergreifende statistische Auswertungen

Werkclustering

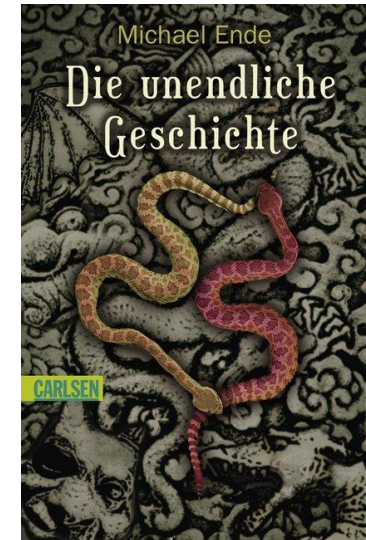
Werkdefinition

The essence of the *work* is the constellation of concepts and ideas that form the shared content of what we define to be *expressions* of the same *work*. A *work* is perceived through the identification of the commonality of content between and among various *expressions*. However, similarity of factual or thematic

IFLA Library Reference Model. A Conceptual Model for Bibliographic Information, August 2017, S. 21f., <https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf>.

- Idee: Zusammenfassen aller Ausprägungen eines Werks, d.h. z.B. verschiedene Ausgaben, Auflagen, Übersetzungen

DEUTSCHE
NATIONAL
BIBLIOTHEK



Erzeugen eines Vorverarbeitungsformats

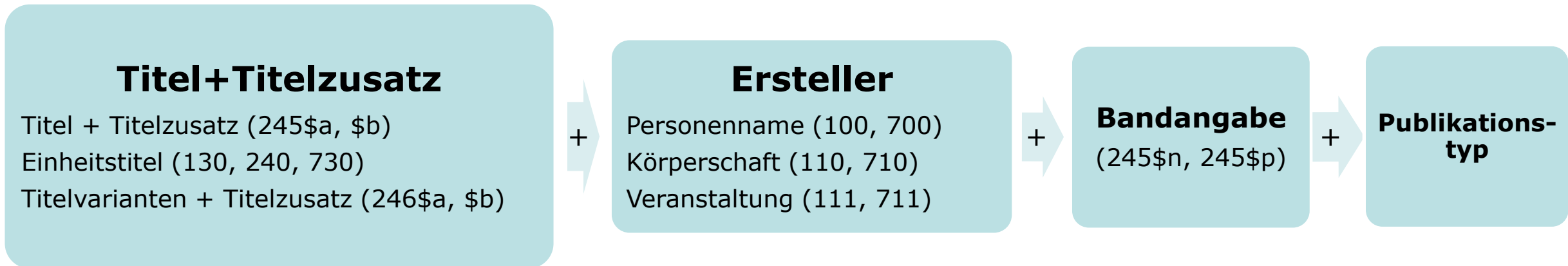
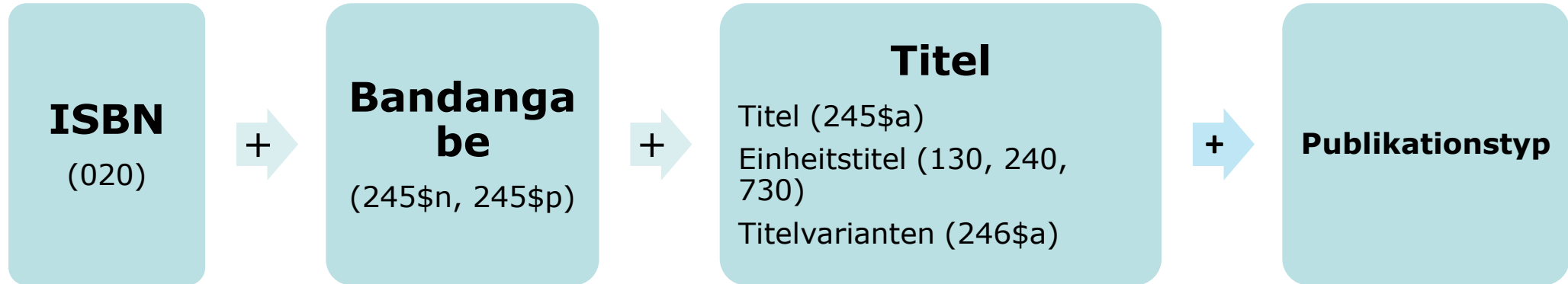
- Nach definierten Regeln werden relevante Daten aus dem MARC-Format extrahiert und zum Teil umgewandelt, normalisiert oder kombiniert (Metamorph)
- Nutzung des Vorverarbeitungsformats zur
 - Erzeugung von Schlüsseln zur Bündelung gleicher Werke
 - Darstellung auf der Culturegraph-Webseite

```
<record>
  <type>bib</type>
  <material>book</material>
  <multipartLevel>single</multipartLevel> <_id>1132025389</_id>
  <publicationType>c</publicationType>
  <form>o</form> <issued>2017</issued>
  <dnb>1132025389</dnb>
  <isbn13>9783647253244</isbn13>
  <isbn13>9783647253244</isbn13>
  <gtin14>9783647253244</gtin14>
  <urn>urn:nbn:de:101:1-2017051417349</urn>
  <eki>DNB1132025389</eki>
  <oclc>989820124</oclc>
  <language>ger</language>
  <country>XA-DE-NI</country>
  <ddc>938.05</ddc>
  <ddc>930</ddc>
  <ddc>880</ddc>
  <ddc>355</ddc>
```

Titel 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 Erschließungsmerkmalen auf Bündelmitglieder in den liefernden Verbänden
 - Sacherschließungsdaten
 - Anreicherung durch Normdatenverknüpfungen

Vergleichsschlüssel erzeugen



Beispiel: Schlüssel einer Publikation

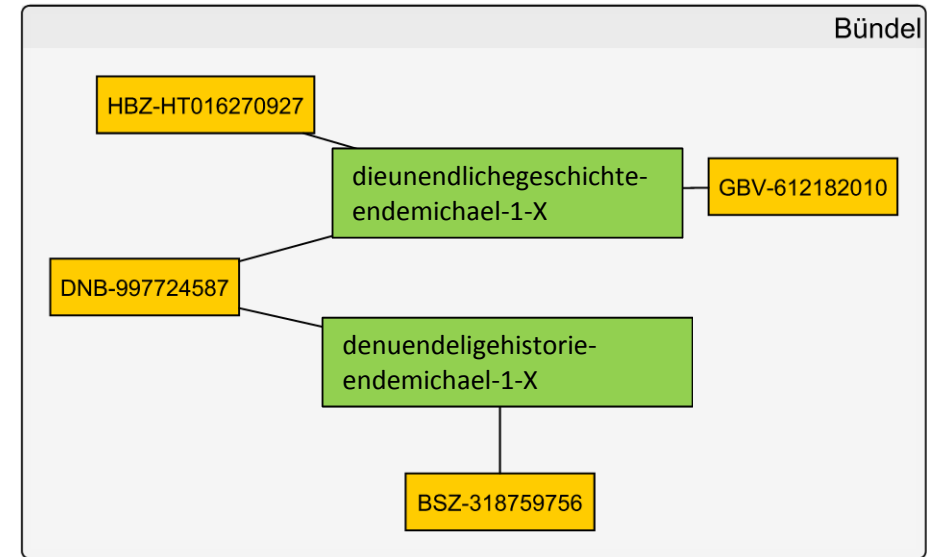
		
Link zu diesem Datensatz	http://d-nb.info/1115738097	
Art des Inhalts	Hochschulschrift	
Titel	Das Meer als Versprechen : Bedeutung und Funktion von Seeherrschaft bei Thukydides / Hans Kopp	
Person(en)	Kopp, Hans (Verfasser)	
Organisation(en)	Vandenhoeck & Ruprecht (Verlag)	
Verlag	Göttingen : Vandenhoeck & Ruprecht	
Zeitliche Einordnung	Erscheinungsdatum: 2017	
Umfang/Format	303 Seiten ; 24 cm	
Andere Ausgabe(n)	Erscheint auch als Online-Ausgabe: Kopp, Hans: Das Meer als Versprechen	
Hochschulschrift	Dissertation, Freie Universität Berlin, 2015	
ISBN/Einband/Preis	978-3-525-25324-3 Festeinband : circa EUR 90.00 (DE), circa EUR 92.60 (AT), circa CHF 112.00 (freier Preis) 3-525-25324-9	
Bestellnummer(n)	Bestellnummer: 25324	
EAN	Erzeugte Schlüssel	
Sprach		
Beziel	<isbnVolumeTitlePubtyp>9783525253243-X-dasmeeralsversprechen-norm</isbnVolumeTitlePubtyp>	
Schlag		
DDC-N		
Sachg	<titleCreator>dasmeeralsversprechenbedeutungufunktionvonseeherrschaftbeithukydidesskopphans-X-norm</titleCreator>	
Weitere		
	Inhaltsverzeichnis	

ISBN-Bandangabe-Titel-Publikationstyp

Titel-Titelzusatz-Ersteller (MARC-Feld 100 bzw. 700)-
Bandangabe-Publikationstyp

Clustering der Schlüssel nach dem Verfahren der Breitensuche

- Für jeden Datensatz liegen ein oder mehrere Schlüssel vor
- Bündel werden anhand von Schlüsseln zusammengestellt
- Dabei werden in einem Bündel alle Datensätze gesammelt, die einen gemeinsamen Schlüssel mit einem anderen Mitglied besitzen



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	Inheritance and transfer tax GND Nachlass GND Steuerrecht GND Erbe
Klassifikation	DDC 343/.43/053 RVK PP 5345 RVK QL 500

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

Weiterverarbeitung / Präsentation der Ergebnisse

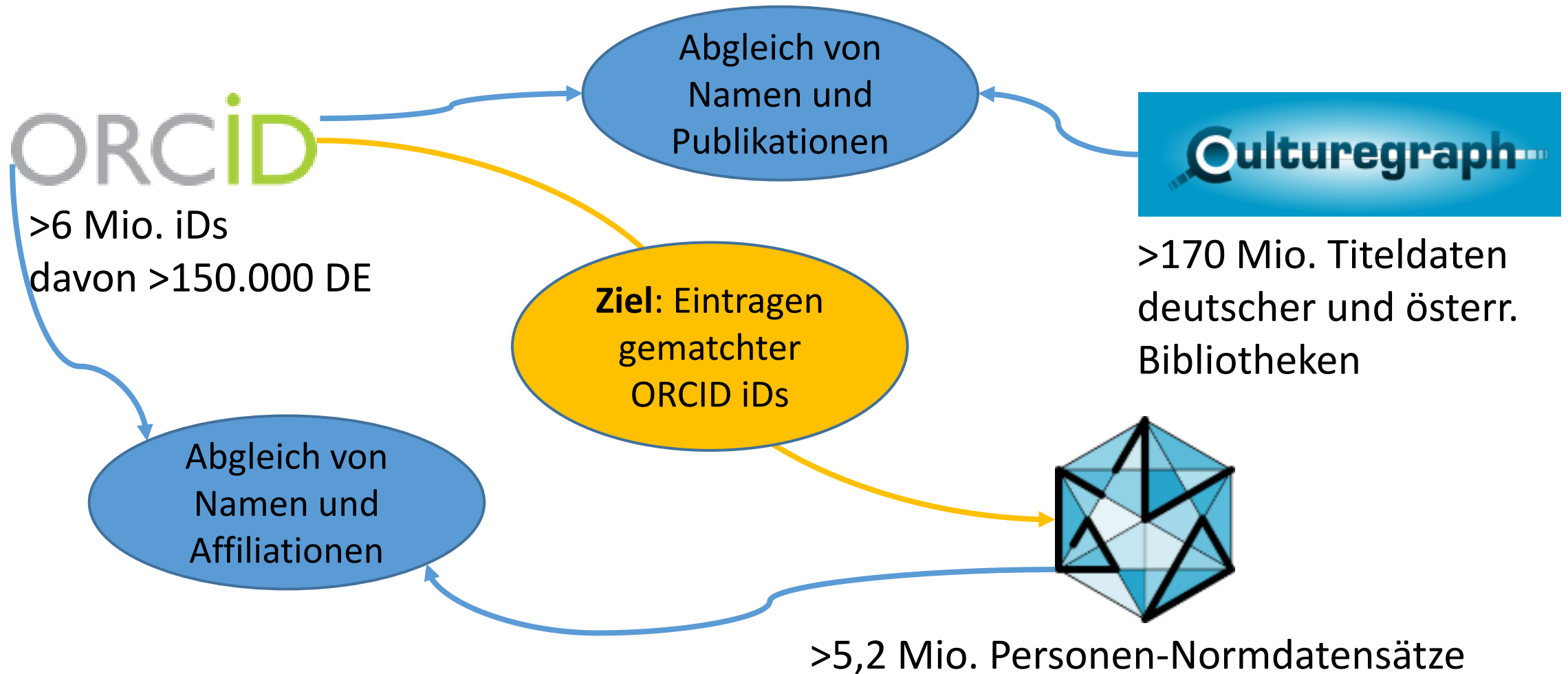
- Suchmöglichkeit nach einzelnen Titeln über die Culturegraph-Webseite, dann Anzeige aller dazugehörigen Titel eines Clusters
- Dump aller Bündel in einer unter den Verbänden abgesprochenen Form zur Datenübernahme (Aggregat)
- In Planung: API

Statistik Werkbündel März 2019

- 103.898.180 Datensätze wurden verarbeitet (nicht behandelt: Teile von mehrbändigen Werken, fortlaufende Ressourcen etc.)
- 15.166.633 Bündel mit mehr als einem Mitglied
- Darin enthalten 73.282.355 Datensätze
- Durchschnittliche Bündelgröße (wenn mehr als ein Mitglied): 5
- Abschließende Evaluation intellektuell und maschinell läuft noch

Abgleich von GND-Personendatensätzen und ORCID Records

Verfügbare Ressourcen



Voraussetzungen in den Metadaten



Andreas Grus

ORCID iD

<https://orcid.org/0000-00>

Print view

Websites

Professional website at
Professional website at

Country

Germany

Other IDs

ResearcherID: J-4296-20

Metadaten für

✓	Nachname + Vorname	✓
✓	Alternative Namensformen	✓
✓	Persistenter Identifikator	✓
✓	Externe Identifikatoren	✓
✓	Affiliationen	✓
✓	Publikationen	✓



005 Tp3
006 http:/
008 piz
011 f
012 v
024 orcid
035 gnd/
039 pnd/
039 pnd/
040 \$erd
043 XA-D
100 Grus
375 m
510 !101!
548 1960



Abgleich von ORCID und GND-Sätzen

Abgleich von
Namen und
Affiliationen

ORCID

Verwendung von



PICA

<p>person:...</p> <p>.../personal-details:family-name</p> <p>.../personal-details:given-names</p>	<p>Nachname + Vorname</p>	<p>100 028A</p>
<p>person:...</p> <p>.../other-name:content</p>	<p>Alternative Namensformen</p>	<p>400 028@</p>
<p>activities:...employment:...</p> <p>.../employment:organization/common:name</p> <p>.../employment:department-name</p>	<p>Affiliationen</p>	<p>510 029R</p> <p>wenn \$4affi</p>

Abgleich mit Titeldaten aus Culturegraph

Abgleich von
Namen und
Publikationen

Schlüssel gebildet aus Name+Vorname und
Titel+Titelzusatz bis zu einer Länge von 50 Zeichen

ORCID

Culturegraph

MARC

<p>person:...</p> <ul style="list-style-type: none">.../personal-details:family-name.../personal-details:given-names.../other-name:content <p>work:contributor...</p> <ul style="list-style-type: none">...work:credit-name	<p>Name der Person</p>	<p>100, 700 +Alternative Namens- formen aus GND</p>
<p>work:work-summary...</p> <ul style="list-style-type: none">.../work:title/common:title.../common:subtitle	<p>Titel</p>	<p>245, 240, 130, 730, 246</p>

ORCID Connecting Research and Researchers

EDIT YOUR RECORD ABOUT ORCID CONTACT US HELP

4,990,483 ORCID iDs and counting. See more...

Max Mustermann
<https://orcid.org/AAA-A-BBBB-CCCC-DDDD>

Websites
[Mendeley profile](#)

Other Ids
 Scopus Author-ID: 123456789

Education (3)

- Example1 University
2005 PhD (Department X)
- University of Example2
1998 MSc (Department Y)
- University Z
1997 | BSc (Department Q)

Works (41 of 41)

Das Beispiel als Beispiel
 2015-08 | journal-article
 DOI: 1234/56789

Schlüssel: mustermannmax*dasbeispielalsbeispiel



Schlüssel: mustermannmax*dasbeispielalsbeispiel

Culturegraph

Titel: Das Beispiel als Beispiel

Person: GND-ID 987654321 Max Mustermann

Umfang: Online-Ressource

Erscheinungsjahr: 2015

Sprache: de

Standard-Identifer: EKI XYZ DOI ABC

Bündelung: Cluster CL-WRK38925434

Eingabe: 0206:12-10-18 Änderung: 9999:01-03-19 16:21:4

005 Tp3

006 <http://d-nb.info/gnd/1168942055>

008 piz

011 f

024 orcid: AAAA-BBBB-CCCC-DDDD\$vcg_orcid

035 gnd/ 987654321

040 \$erda

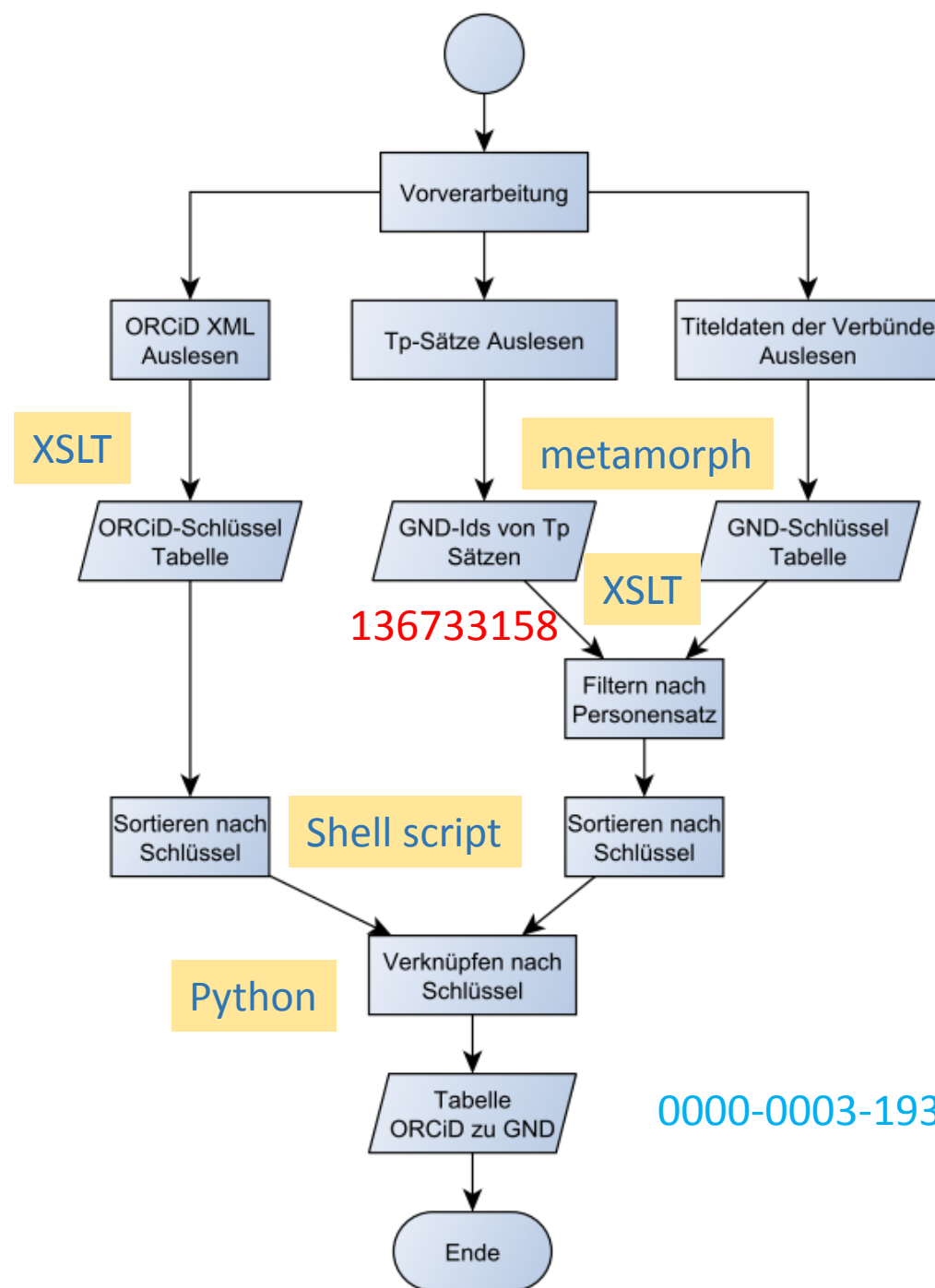
043 XB-ID

100 Max Mustermann



GND-ID
987654321

0000-0003-1931-5494
 degiannakisstavros*
 aprobitmodelforthestateofth
 egreekgdpgrowth



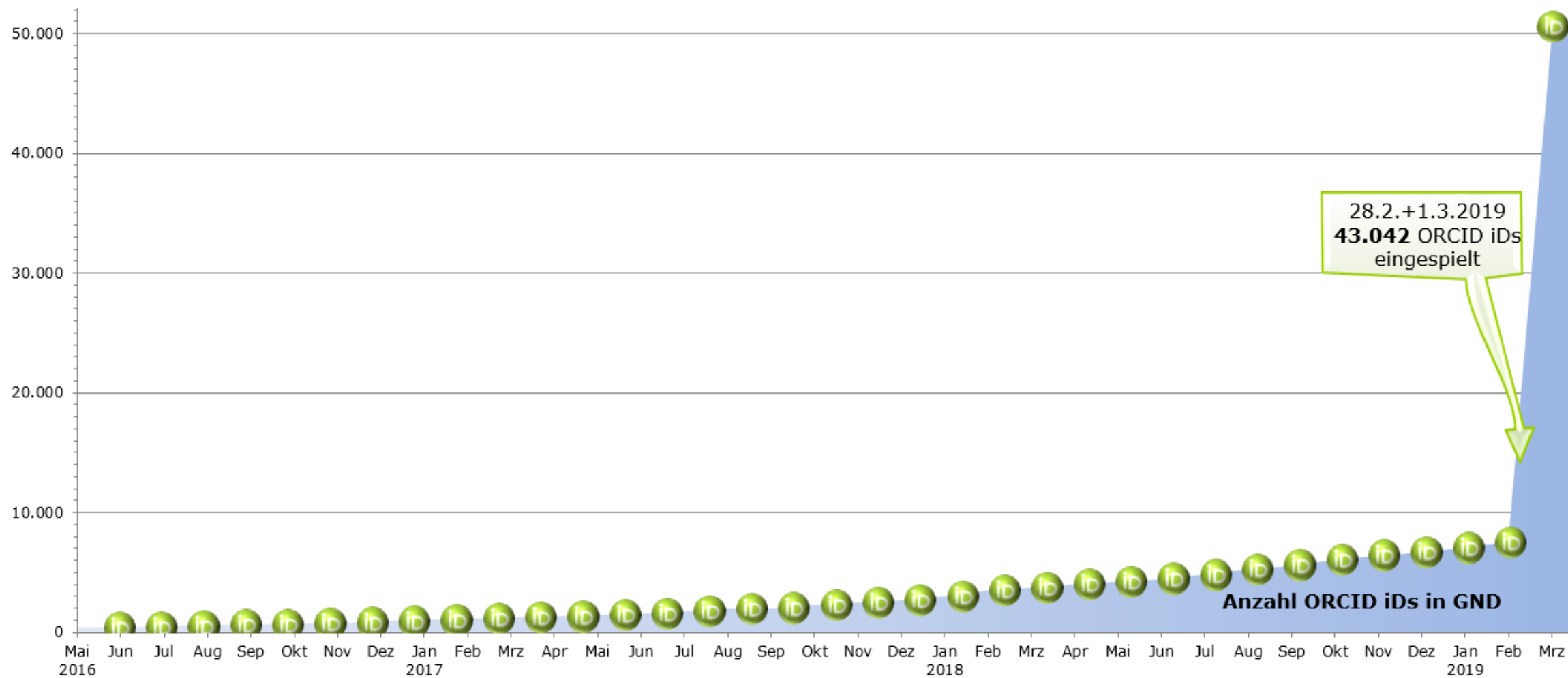
136733158
 degiannakisstavros*
 aprobitmodelforthestateofth
 egreekgdpgrowth

0000-0003-1931-5494 136733158

Datenübernahme

- 15.391 Matches Name und Affiliation (ORCiD und GND)
- 35.199 Matches Name und Titel verknüpfter Publikationen (ORCiD und Culturegraph)
- 4.811 Überschneidungen der beiden Verfahren
- 2.531 Überschneidungen mit manuell eingetragenen ORCID iDs
- **43.042** Einspielungen in die GND

ORCID iDs in der GND



Vielen Dank für Ihre
Aufmerksamkeit.

Fragen?

Kontakt: Angela Vorndran

a.vorndran@dnb.de

Stefan Grund

s.grund@dnb.de



<http://hub.culturegraph.org/>