# Offene Infrastruktur für bibliothekarische Daten

Linked Open Data, JSON & OpenRefine in der Praxis

Adrian Pohl & Fabian Steeg

Offene Infrastruktur, Hochschulbibliothekszentrum NRW (hbz)



WWW, 2022-02-10

Zentrum für Bibliotheks- und Informationswissenschaftliche Weiterbildung, TH Köln (ZBIW)

Diese Präsentation:

https://slides.lobid.org/2022-zbiw/ (PDF)



Fragen sind stets willkommen

Fragen sind stets willkommen

Hilf anderen, wo du kannst

Fragen sind stets willkommen

Hilf anderen, wo du kannst

Sei offen für unterschiedliche Fachkenntnisse & Erfahrungen

## Adrian

## Adrian

Seit 2008 im hbz, seit 2019 Leitung Offene Infrastruktur

## Adrian

Seit 2008 im hbz, seit 2019 Leitung Offene Infrastruktur

Schwerpunkte: Datenmodellierung, Offene Standards, Kommunikation/Koordination

# Fabian

## Fabian

Seit 2012 im hbz, bis 2019 LOD, dann Offene Infrastruktur

## Fabian

Seit 2012 im hbz, bis 2019 LOD, dann Offene Infrastruktur

Schwerpunkte: Webentwicklung und Datentransformation für lobid.org (etc.)

# Und ihr?

## Und ihr?

Wo und was arbeitet ihr?

## Und ihr?

Wo und was arbeitet ihr?

Was erwartet ihr von diesem Seminar?

### Agenda

- 1. Offene Infrastruktur
- 2. Linked Data & JSON-LD
- 3. OpenRefine Reconciliation

## Grober Zeitplan

1. Offene Infrastruktur	lobid-Dienste und Nutzungsbeispiele	10:00–10:30
	GND, lobid-gnd, Rechercheoberfläche	10:30–11:00
2. Linked Data & JSON-LD	RDF, Linked Data	11:00–11:30
	JSON, JSON-LD, LOUD	11:30–12:00
	Daten und Abfragemöglichkeiten GND	12:00–12:30
	Pause	
	Daten und Abfragemöglichkeiten Verbundkatalog	13:30–14:00
3. OpenRefine Reconciliation	Grundlagen von OpenRefine und Reconciliation	14:00–14:30
	Verbesserung des Matching auf GND-Einträge	14:30–15:00
	Datenanreicherung auf Basis der Matches	15:00–15:30

# Teil 1: Offene Infrastruktur

lobid	lobid-Dienste und Nutzungsbeispiele	10:00–10:30
GND	GND, lobid-gnd, Rechercheoberfläche	10:30–11:00





Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, seit 1973



#### Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, seit 1973

Dienstleistungs- und Entwicklungseinrichtung für digitale Services in Hochschulbibliotheken



Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, seit 1973

Dienstleistungs- und Entwicklungseinrichtung für digitale Services in Hochschulbibliotheken

Einige Arbeitsbereiche: Verbundkatalog, Suchportal DigiBib, Fernleihe, Konsortiale Erwerbung

Nutzung und Pflege von Infrastruktur für die Publikation strukturierter Daten im Web:

Nutzung und Pflege von Infrastruktur für die Publikation strukturierter Daten im Web:

Offene Webstandards und darauf aufsetzend kooperative Entwicklung von Spezifikationen (z.b AMB)

Nutzung und Pflege von Infrastruktur für die Publikation strukturierter Daten im Web:

Offene Webstandards und darauf aufsetzend kooperative Entwicklung von Spezifikationen (z.b AMB)

Transformation von Alt- und Fremddaten in web-kompatible Strukturen (mit Metafacture)

Nutzung und Pflege von Infrastruktur für die Publikation strukturierter Daten im Web:

Offene Webstandards und darauf aufsetzend kooperative Entwicklung von Spezifikationen (z.b AMB)

Transformation von Alt- und Fremddaten in web-kompatible Strukturen (mit Metafacture)

Publikation strukturierter Daten im Web inkl. APIs (lobid, OERSI, SkoHub)





lobid steht für Linking Open Bibliographic Data



#### lobid steht für Linking Open Bibliographic Data

Seit über zehn Jahren die zentrale Komponente der offenen Infrastruktur im hbz



#### lobid steht für Linking Open Bibliographic Data

Seit über zehn Jahren die zentrale Komponente der offenen Infrastruktur im hbz

Stellt intuitive, web-basierte Rechercheoberflächen und Schnittstellen bereit

lobid-resources: Daten des hbz-Verbundkatalogs

lobid-resources: Daten des hbz-Verbundkatalogs

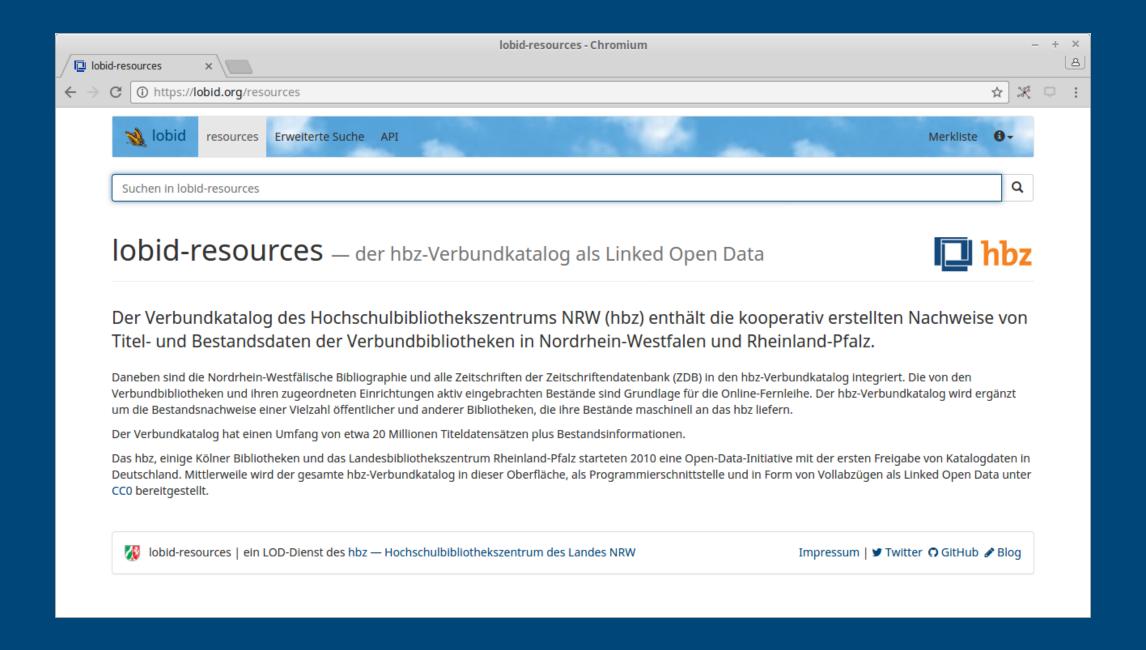
lobid-organisations: Daten des deutschsprachigen Sigelverzeichnisses und DBS-Stammdaten

lobid-resources: Daten des hbz-Verbundkatalogs

lobid-organisations: Daten des deutschsprachigen Sigelverzeichnisses und DBS-Stammdaten

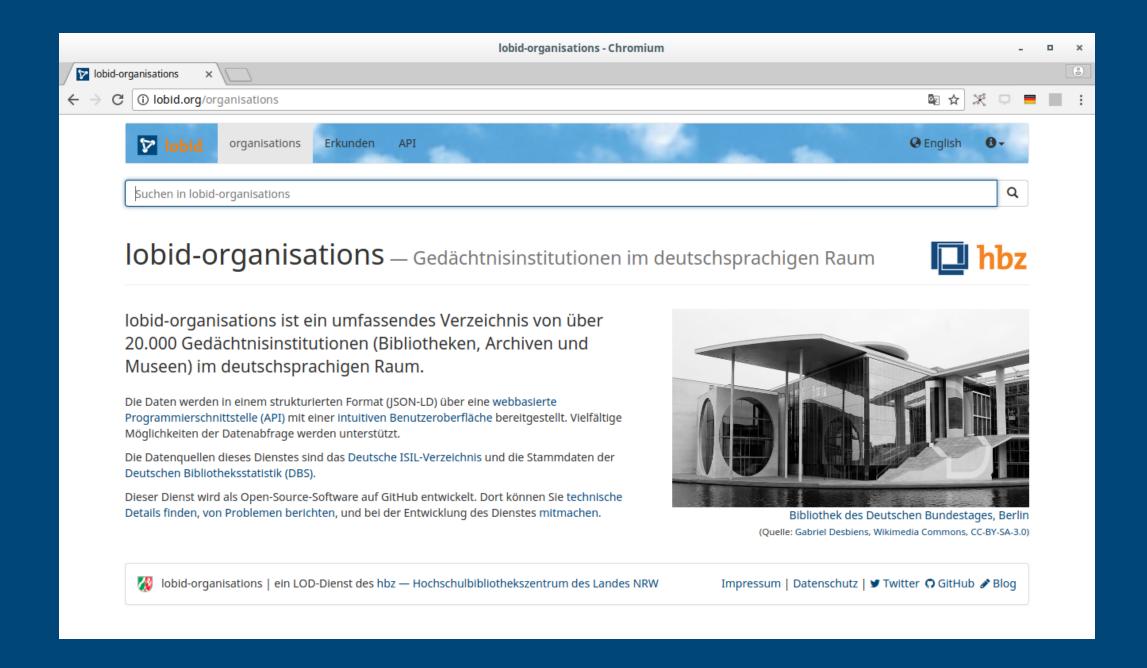
lobid-gnd: Gemeinsame Normdatei

#### lobid-resources



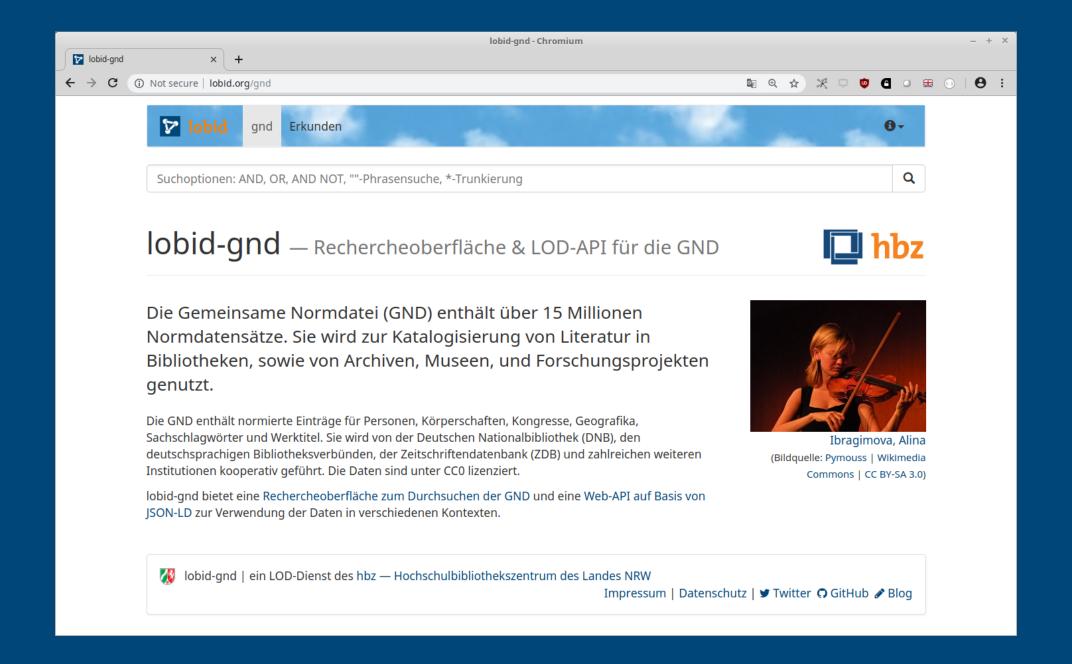
Dokumentation der lobid-resources-API: https://lobid.org/resources/api

#### lobid-organisations



Dokumentation der lobid-organisations-API: https://lobid.org/organisations/api

## lobid-gnd



Dokumentation der lobid-gnd-API: https://lobid.org/gnd/api

## lobid – ein zentraler Knoten bibliothekarischer Dateninfrastruktur



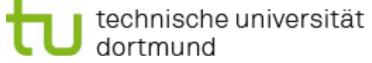
# lobid: Nutzungsbeispiele



Die Nordrhein-Westfälische Bibliographie



Biographisches Portal NRW ulb Münster



### bauhaus-archiv

deutsches literatur archiv marbach



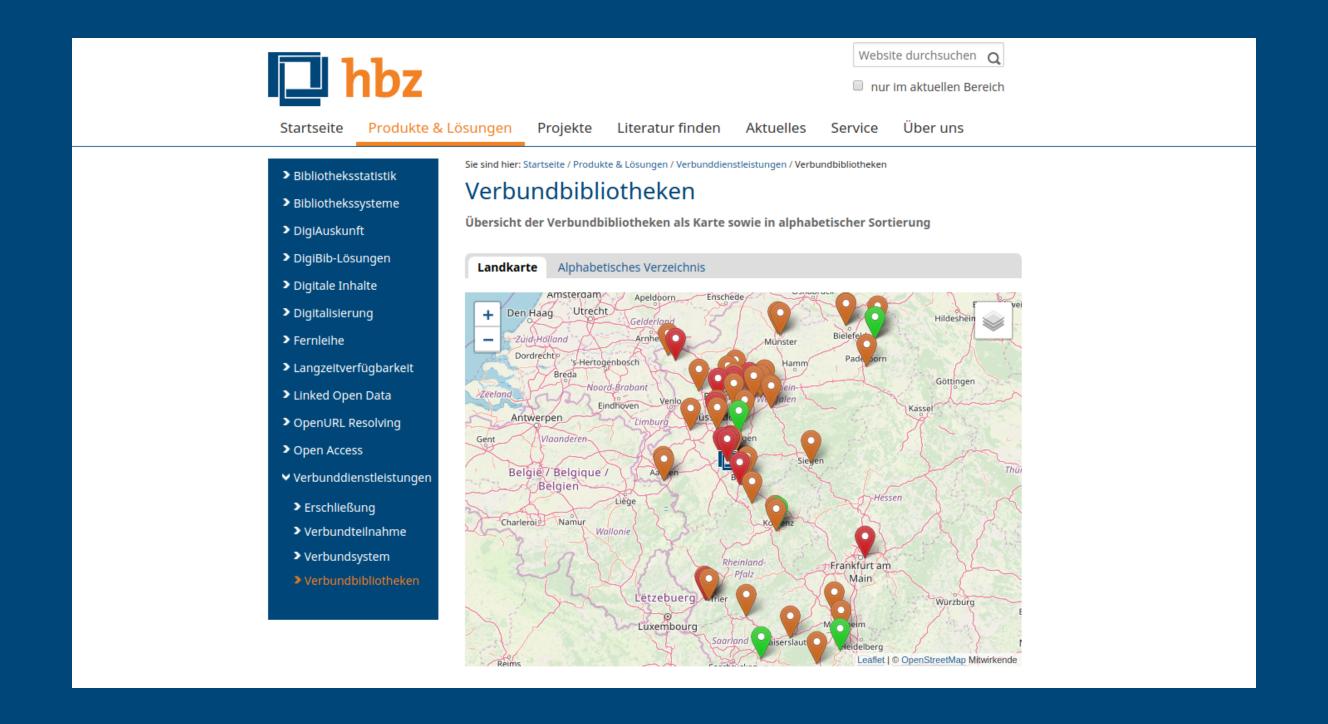


correspSearch

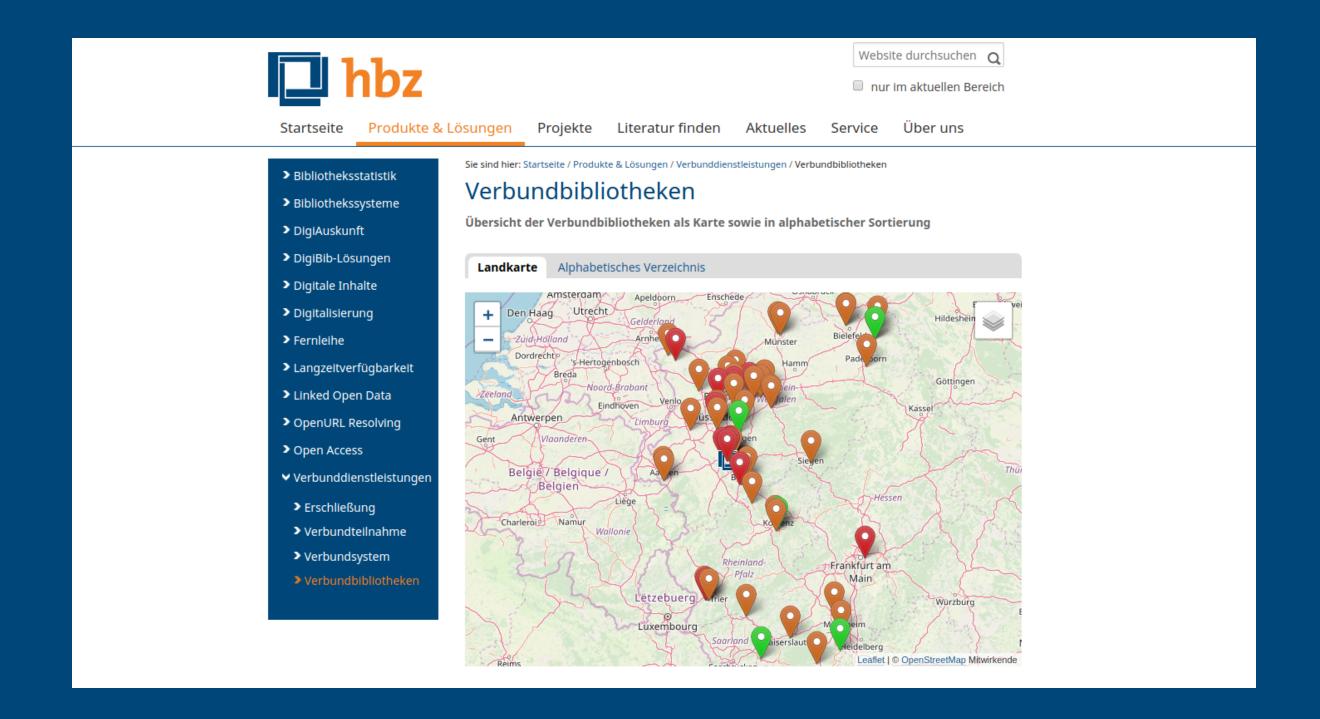
Briefeditionen durchsuchen und vernetzen

4

### Verbundbibliotheken auf hbz-Website

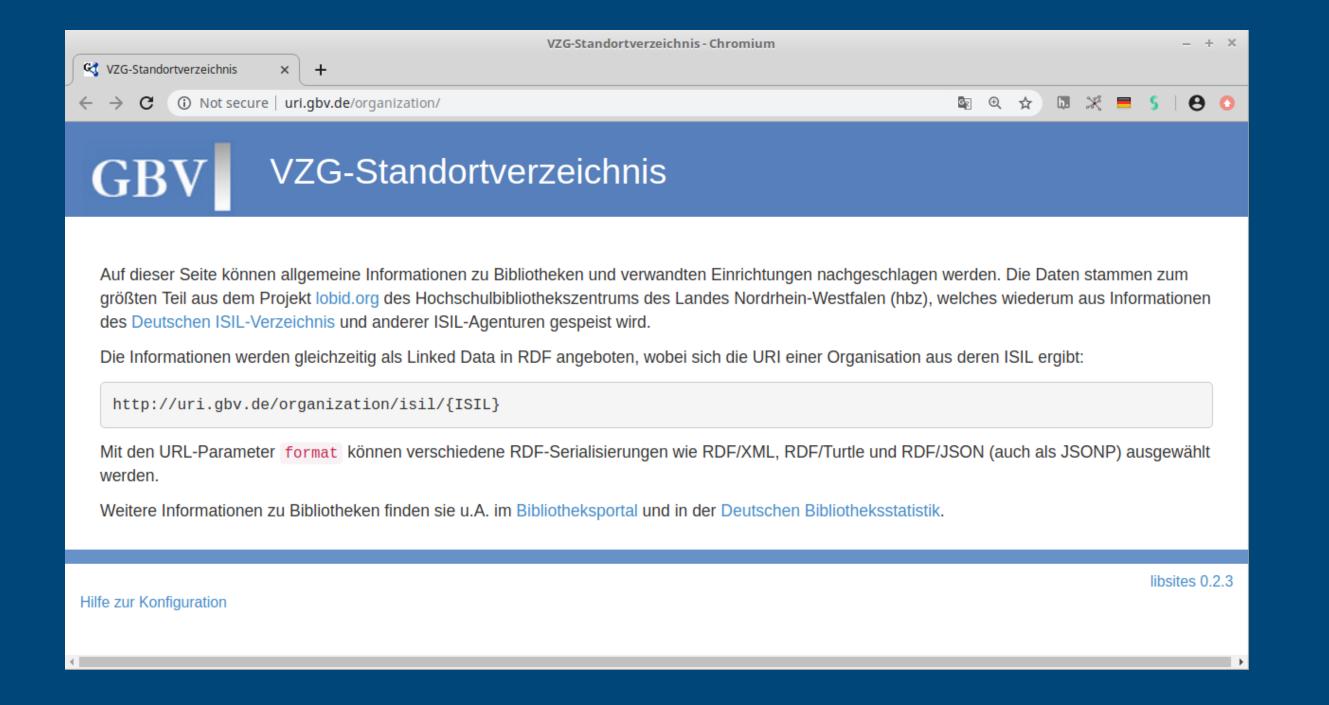


### Verbundbibliotheken auf hbz-Website



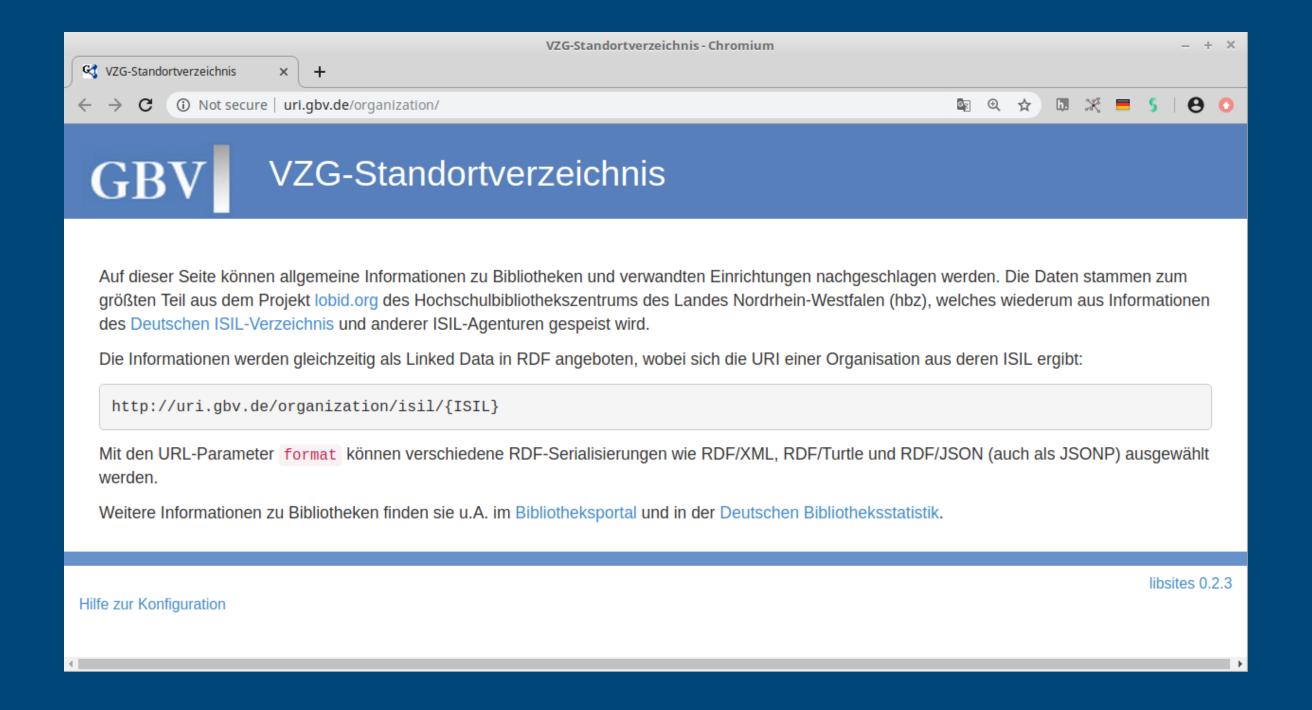
Basiert auf lobid-Organisationsdaten

## VZG-Standortverzeichnis



4

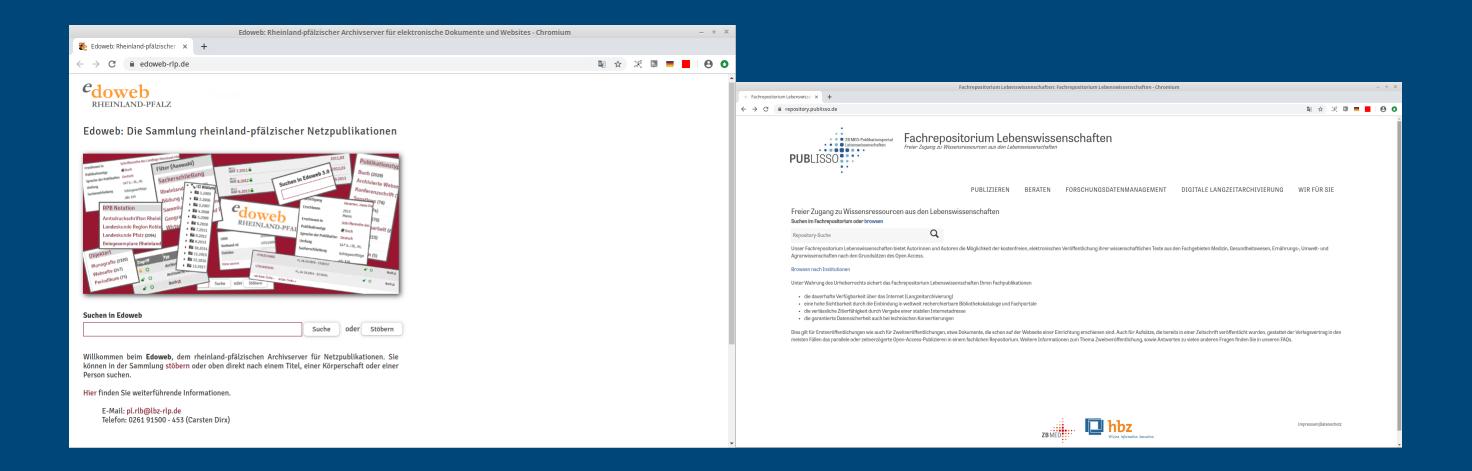
# VZG-Standortverzeichnis



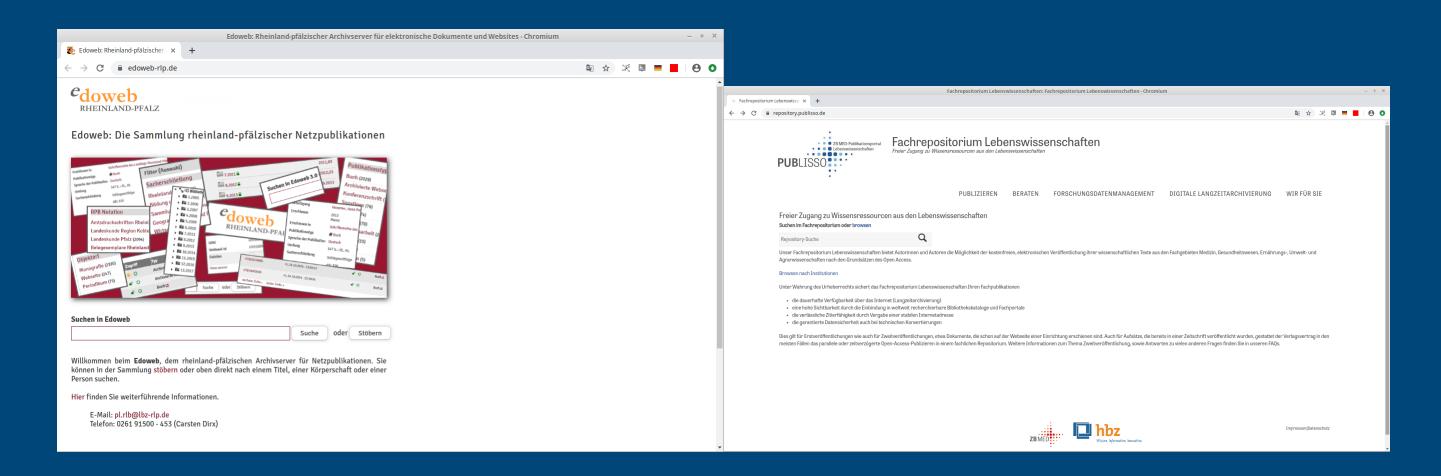
Basiert auf lobid-Organisationsdaten



# Edoweb und Fachrepositorium Lebenswisschenschaften (FRL)

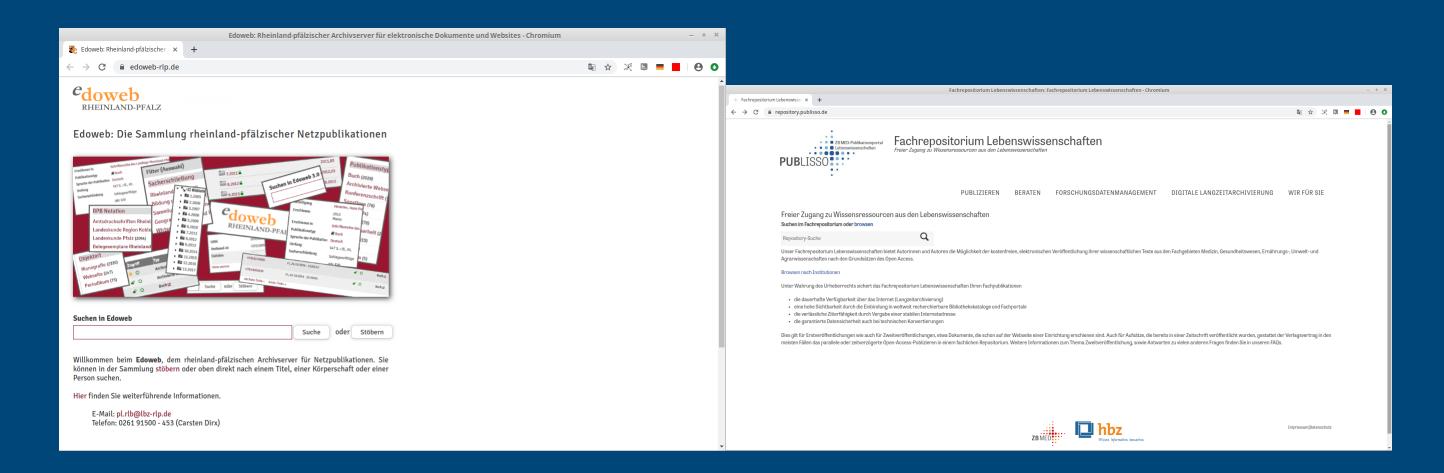


# Edoweb und Fachrepositorium Lebenswisschenschaften (FRL)



Enge Integration der lobid-Verbunddaten und des lobid-Datenmodells

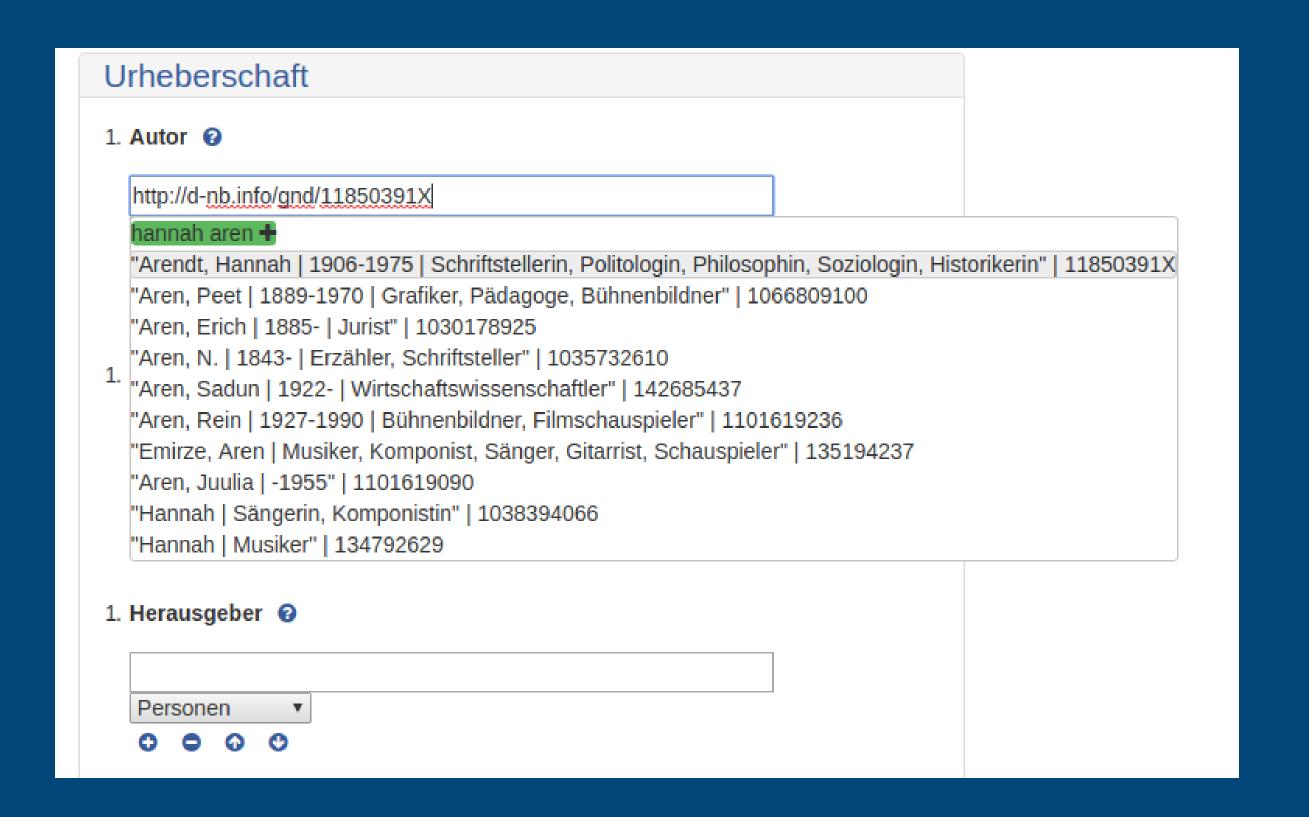
# Edoweb und Fachrepositorium Lebenswisschenschaften (FRL)



Enge Integration der lobid-Verbunddaten und des lobid-Datenmodells

In Publisso auch GND-ID-Lookup via lobid

# GND-Lookup im Fachrepositorium Lebenswissenschaften



Α







Die Nordrhein-Westfälische Bibliographie







Meta-Katalog zum Nachweis landeskundlicher Literatur in Deutschland







Biographisches Portal NRW ulb Münster



technische universität dortmund





correspSearch

Briefeditionen durchsuchen und vernetzen

### bauhaus-archiv

deutsches literatur archiv marbach

Webanwendung, die nahezu vollständig auf lobid-Daten basiert

Webanwendung, die nahezu vollständig auf lobid-Daten basiert

Nutzt das NWBib-Subset in den lobid-Verbunddaten

Webanwendung, die nahezu vollständig auf lobid-Daten basiert

Nutzt das NWBib-Subset in den lobid-Verbunddaten

Informationen zu besitzenden Bibliotheken aus lobid-Organisationsdaten

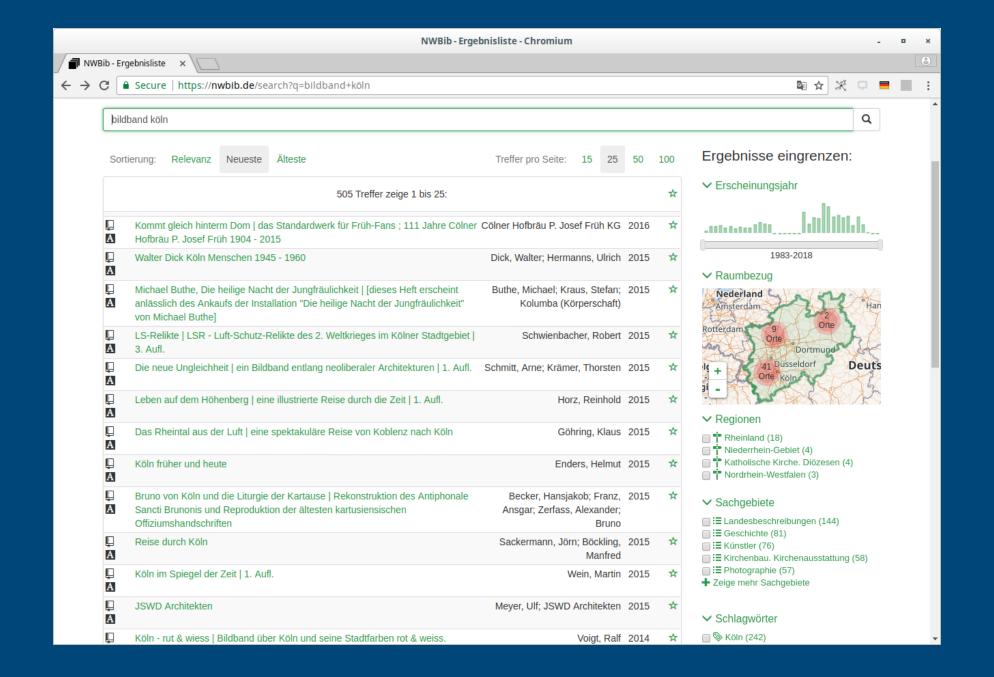
Webanwendung, die nahezu vollständig auf lobid-Daten basiert

Nutzt das NWBib-Subset in den lobid-Verbunddaten

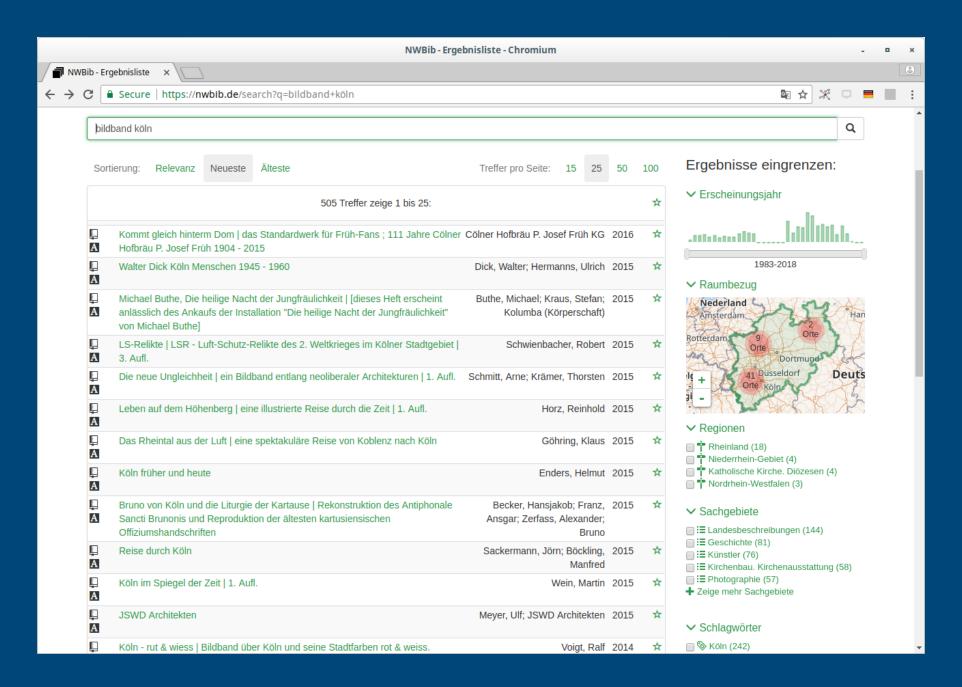
Informationen zu besitzenden Bibliotheken aus lobid-Organisationsdaten

Themenvorschläge auf Basis von lobid-gnd

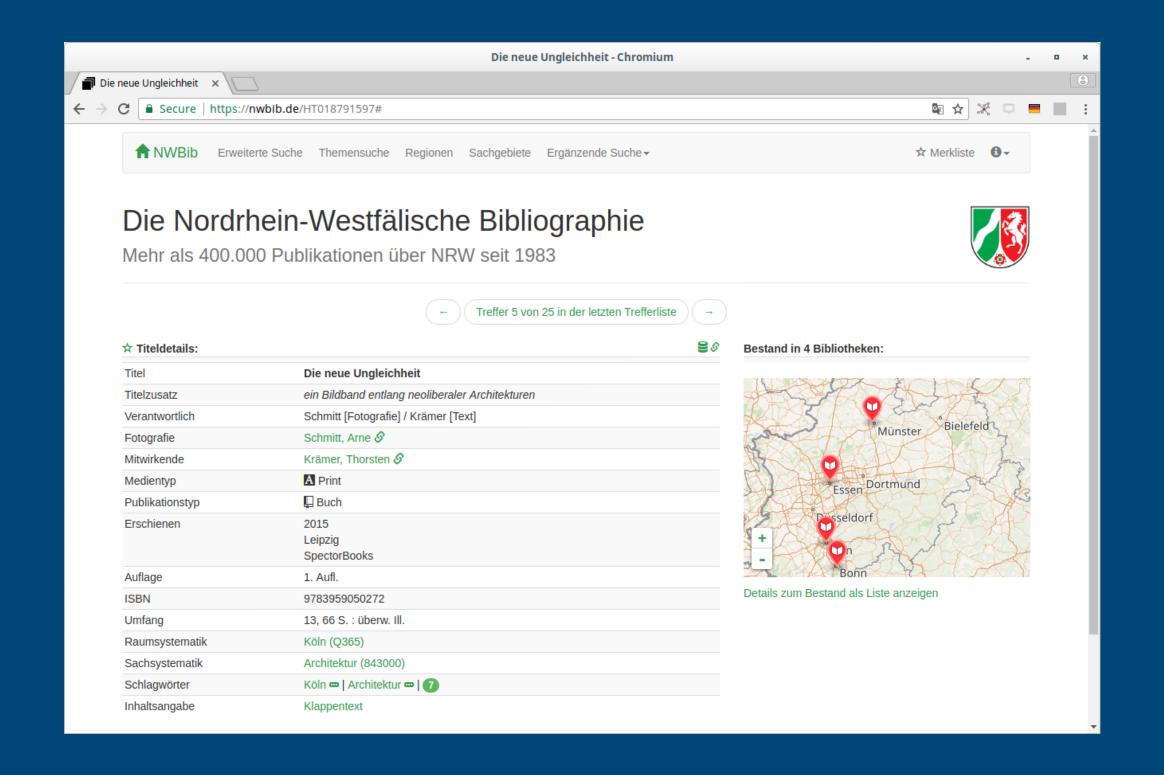
# NWBib-Suchergebnisliste

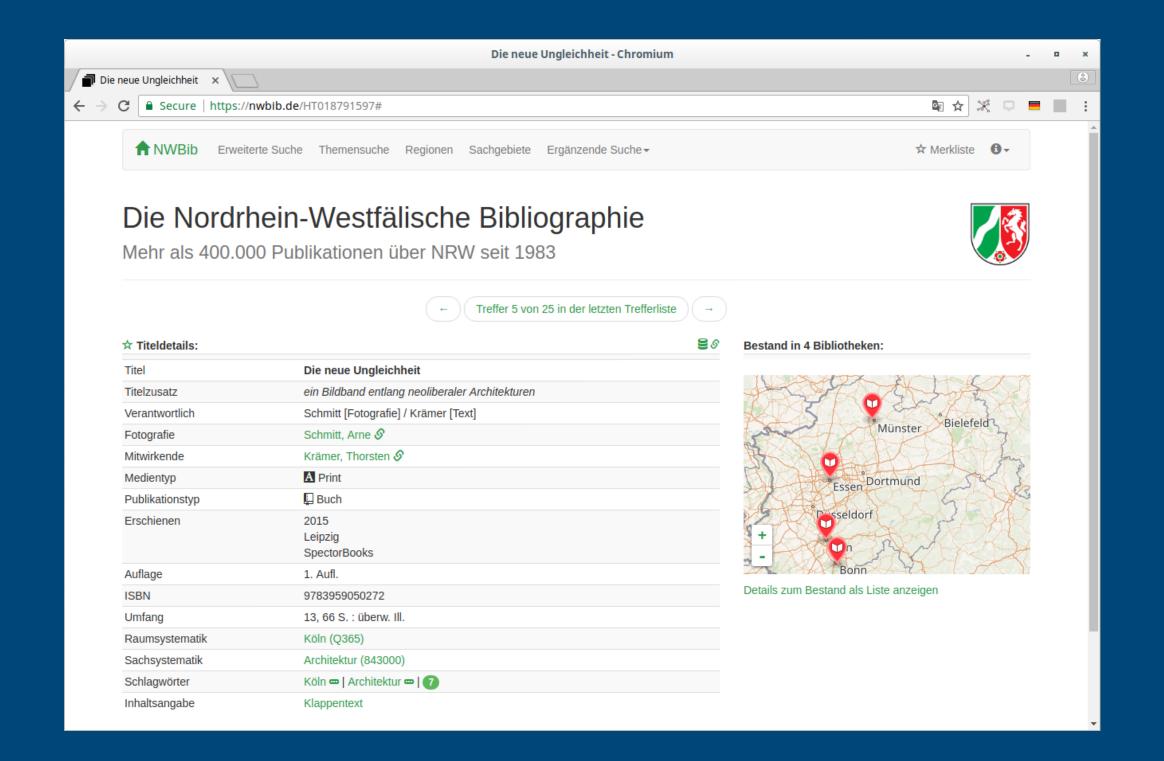


# NWBib-Suchergebnisliste



Anfrage gegen das NWBib-Subset in lobid-Verbunddaten





Informationen zu besitzenden Bibliotheken (Name, URL, Standort) werden aus lobid geladen







Die Nordrhein-Westfälische Bibliographie







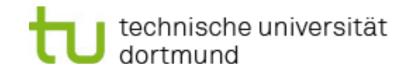


Meta-Katalog zum Nachweis landeskundlicher Literatur in Deutschland

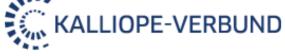




Biographisches Portal NRW ulb Münster







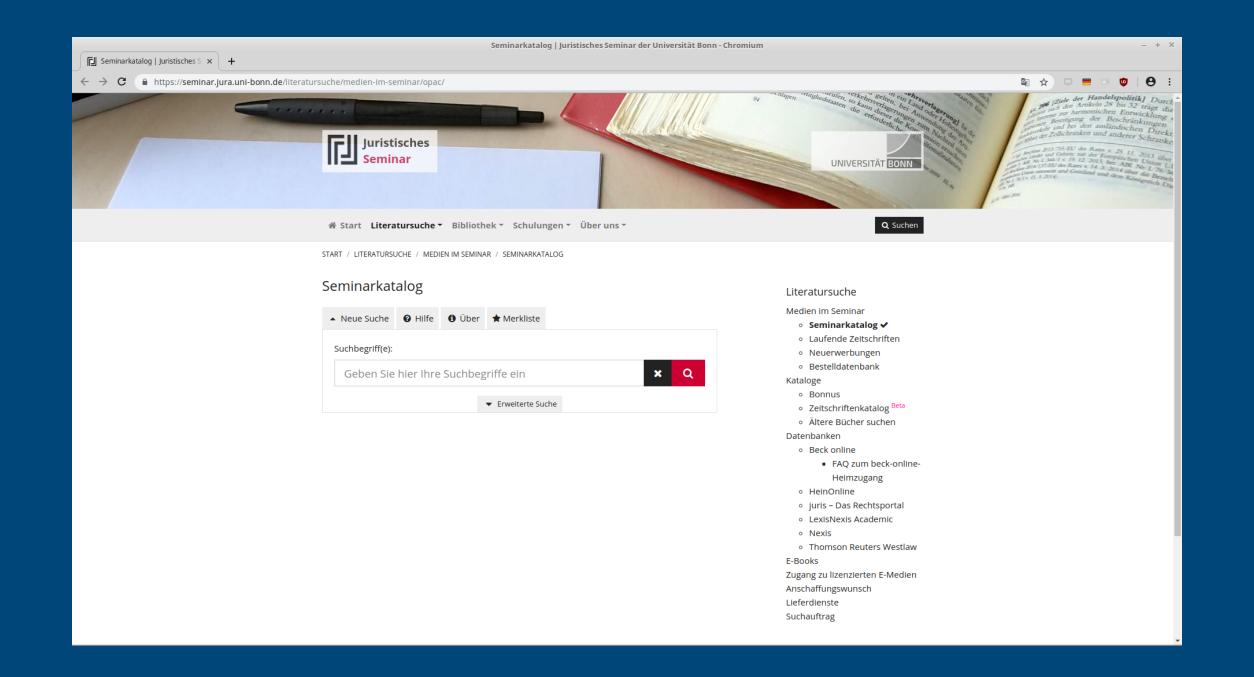


correspSearch Briefeditionen durchsuchen und vernetzen

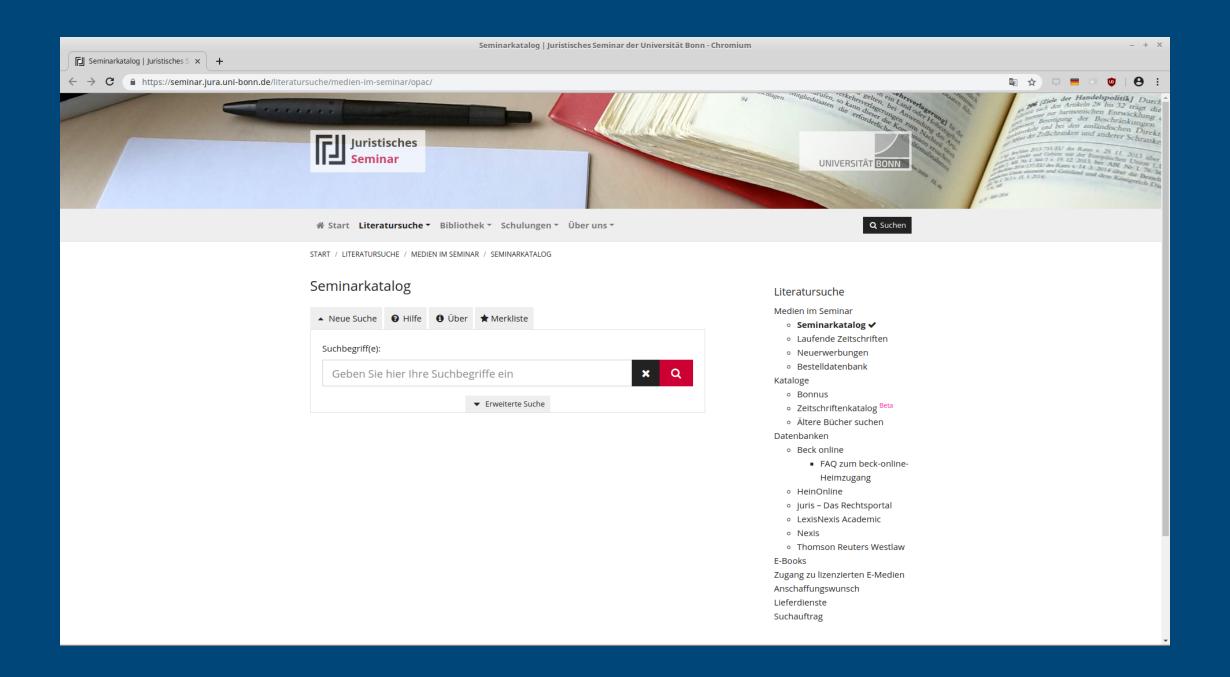
### bauhaus-archiv

deutsches literatur archiv marbach

# Bibliothekskatalog des Juristischen Seminars der ULB Bonn



## Bibliothekskatalog des Juristischen Seminars der ULB Bonn



Basiert auf täglich aktualisierten lobid-Daten

### Über den Seminarkatalog

×

#### Inhalt

Der Seminarkatalog verzeichnet unseren Buch- und Zeitschriftenbestand soweit er elektronisch erfasst ist. Ergänzend haben wir ca. 7.000 juristische Titel elektronischer Bücher und Zeitschriften hinzugefügt:

• Buch- und Zeitschriftenbestand des Juristischen Seminars

(Quelle: lobid-resources)

· Bestellte und im Geschäftsgang befindliche Bücher

(Quelle: Juristisches Seminar)

• Rund 6.600 lizenzierte juristische eBooks und eZeitschriften

(Quelle: lobid-resources)

• Rund 500 Open Access eBooks

(Quelle: DOAB)

### Herkunft der Daten

Die Daten stammen aus dem HBZ-Verbundkatalog und werden vom Hochschulbibliothekszentrum NRW (HBZ) über die *lobid-resources-API* bereitgestellt.

Weitere Daten stammen aus dem Directory of Open Access Books (DOAB).

### Aktualisierung

Die Daten werden täglich (werktags) aktualisiert.







Die Nordrhein-Westfälische Bibliographie







### Virtuelle Deutsche Landesbibliographie

Meta-Katalog zum Nachweis landeskundlicher Literatur in Deutschland











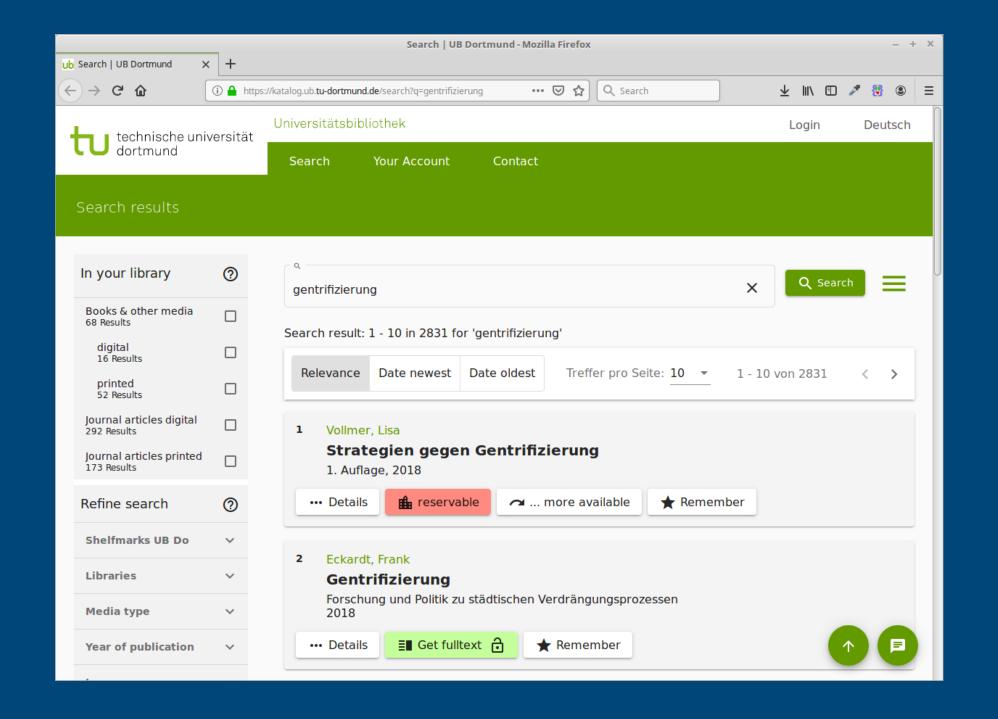


correspSearch Briefeditionen durchsuchen und vernetzen

### bauhaus-archiv

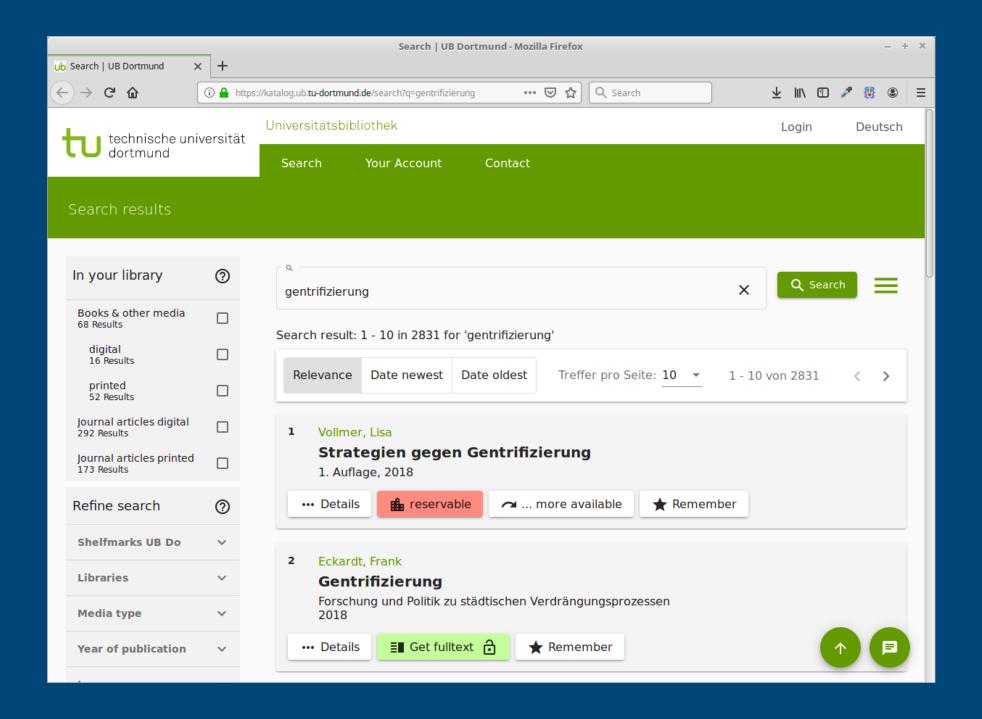
deutsches literatur archiv marbach

### Discovery-Index der UB Dortmund



Д

## Discovery-Index der UB Dortmund



Titeldaten zu den Beständen der UB & einiger umliegender Bibliotheken werden täglich von lobid geholt, transformiert und in den Index geladen

4



correspSearch

Die Nordrhein-Westfälische Bibliographie



Biographisches Portal NRW ulb Münster



### bauhaus-archiv

deutsches literatur archiv marbach

Briefeditionen durchsuchen und vernetzen

## Virtuelle Deutsche Landesbibliographie



- 1. The road to recovery: from the Central Collecting Points to a safe haven the J.R.S.O. Dossier // Steinberg, Shlomit. In: Kunst sammeln, Kunst handeln. 2013, S. 119-131
- 2. Affinität wider Willen?: Hannah Arendt, Theodor W. Adorno und die Frankfurter Schule // Weissberg, Liliane. - Grünwald: Preselect.media., 2011
- 3. Affinität wider Willen?: Hannah Arendt, Theodor W. Adorno und die Frankfurter Schule / 1. Aufl. Frankfurt am Main: Campus Verlag, 2011
- 4. Konzeptionen des Judentums zwischen Säkularisierung und Marxismus: Hannah Arendt und Max Horkheimer / / Boll, Monika. In: Affinität wider Willen?. 2011, S. 103-117
- 5. Affinität wider Willen?: Hannah Arendt, Theodor W. Adorno und die Frankfurter Schule // Weissberg, Liliane. Frankfurt [u.a.]: Campus-Verl., 2011
- 6. Hannah Arendt und Max Horkheimer: Konzeptionen des Judentums zwischen Säkularisierung und Marxismus // Boll, Monika. In: Einsicht. ISSN 1868-4211. Bd. 2.2010, H. 3, S. 24-29
- 7. Zur Konferenz "Hannah Arendt und die Frankfurter Schule" / / Weissberg, Liliane. In: Einsicht. ISSN 1868-4211. Bd. 2.2010, H. 3, S. 18-22
- 8. Hannah Arendt und die Frankfurter Schule : eine Einleitung // Konitzer, Werner. In: Einsicht. ISSN 1868-4211. Bd. 2.2010, H. 3, S. 16-17
- 9. Die andere Erinnerung : jüdische Autoren in der westdeutschen Nachkriegsliteratur / / Braese, Stephan. Unveränd. Nachdr. der Erstausg. Berlin, Wien, Philo-Verl-Ges., 2001 München : Ed. Text + Kritik, 2010
- 10. Die 100 größten Deutschen : auf den Spuren bedeutender Persönlichkeiten / / Rasper, Martin. Genehmigte Sonderausg. Augsburg : Weltbild, 2008

Weitere Treffer..

Treffer insgesamt: 15

Benötigte Zeit: 0.687 Sekunder

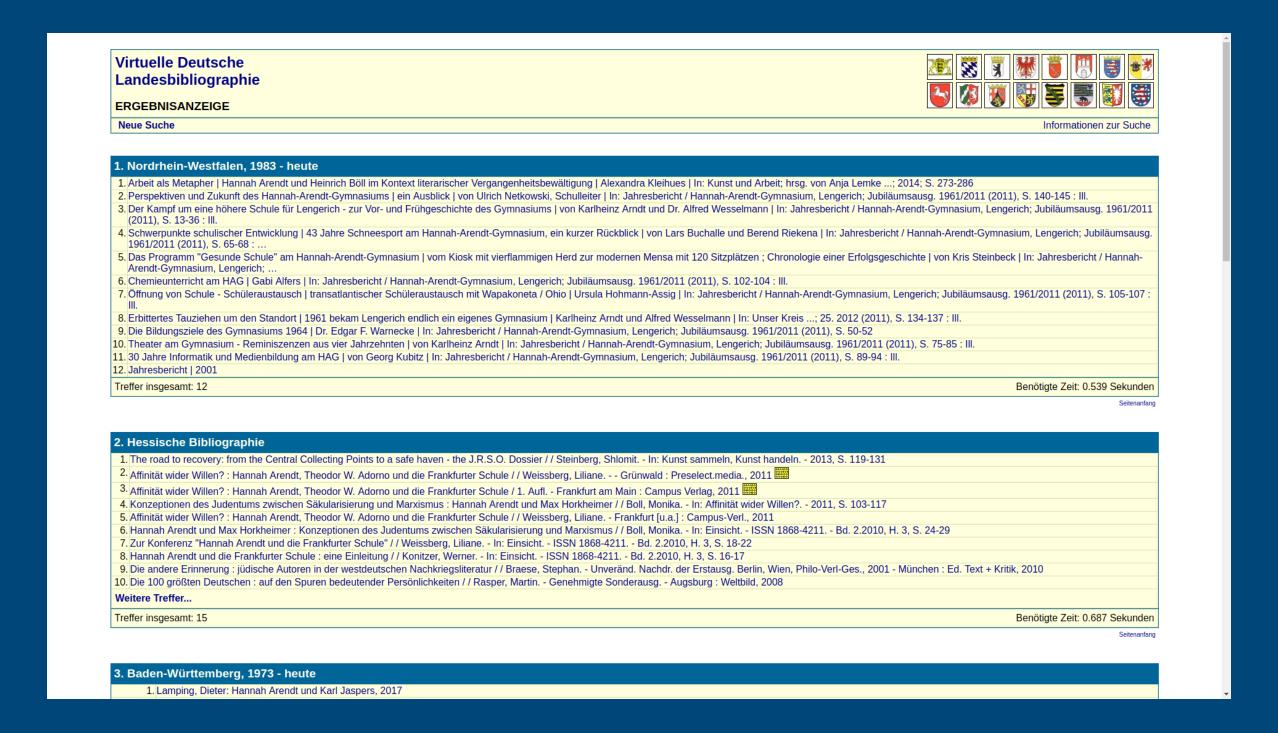
Seitenanfang

#### 3. Baden-Württemberg, 1973 - heute

1. Lamping, Dieter: Hannah Arendt und Karl Jaspers, 2017

Α

## Virtuelle Deutsche Landesbibliographie



Seit 2018 ist nwbib.de in die VDL integriert







Die Nordrhein-Westfälische Bibliographie





Biographisches Portal NRW ulb Münster



technische universität dortmund

### bauhaus-archiv

deutsches literatur archiv marbach

### Virtuelle Deutsche Landesbibliographie

Meta-Katalog zum Nachweis landeskundlicher Literatur in Deutschland





RHEINLAND-PFALZ

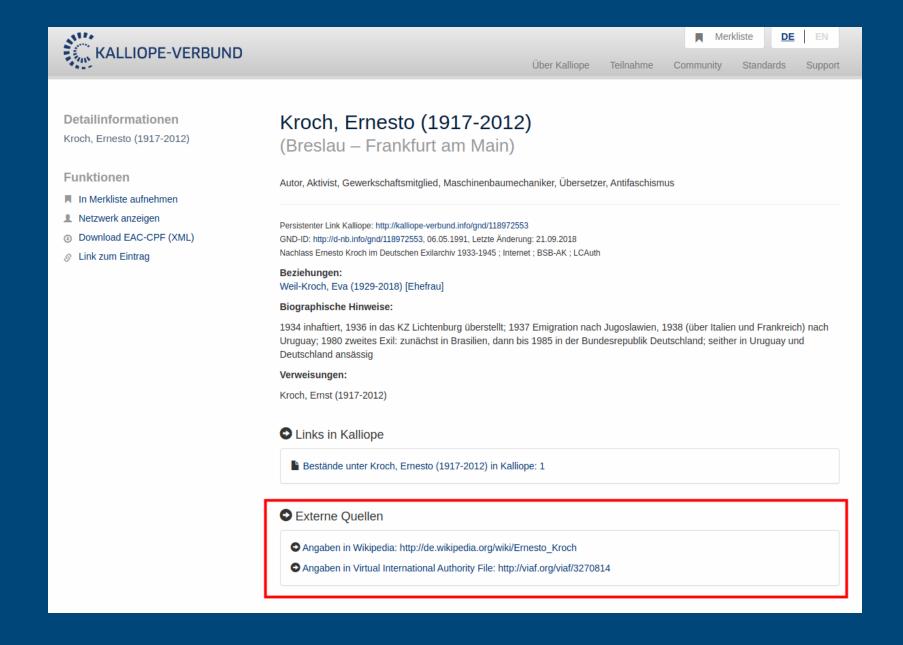




correspSearch

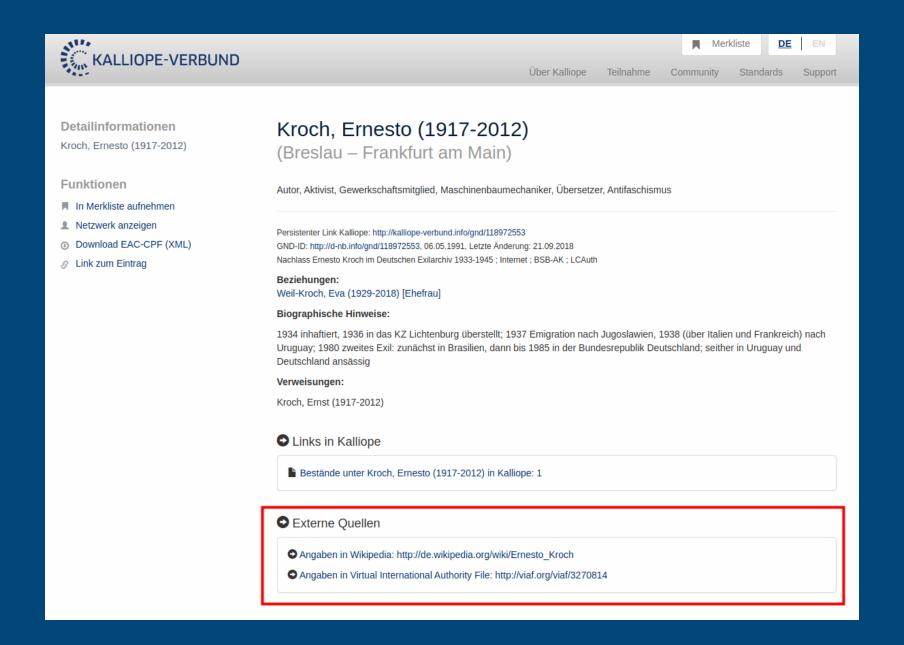
Briefeditionen durchsuchen und vernetzen

# Kalliope



 $\lambda$ 

# Kalliope



Anzeige von externen Links (z.B. Wikipedia) aus lobid-gnd auf Personenseiten

Samuel Control of the Control of the







Die Nordrhein-Westfälische Bibliographie







#### Virtuelle Deutsche Landesbibliographie

Meta-Katalog zum Nachweis landeskundlicher Literatur in Deutschland





Biographisches Portal NRW ulb Münster







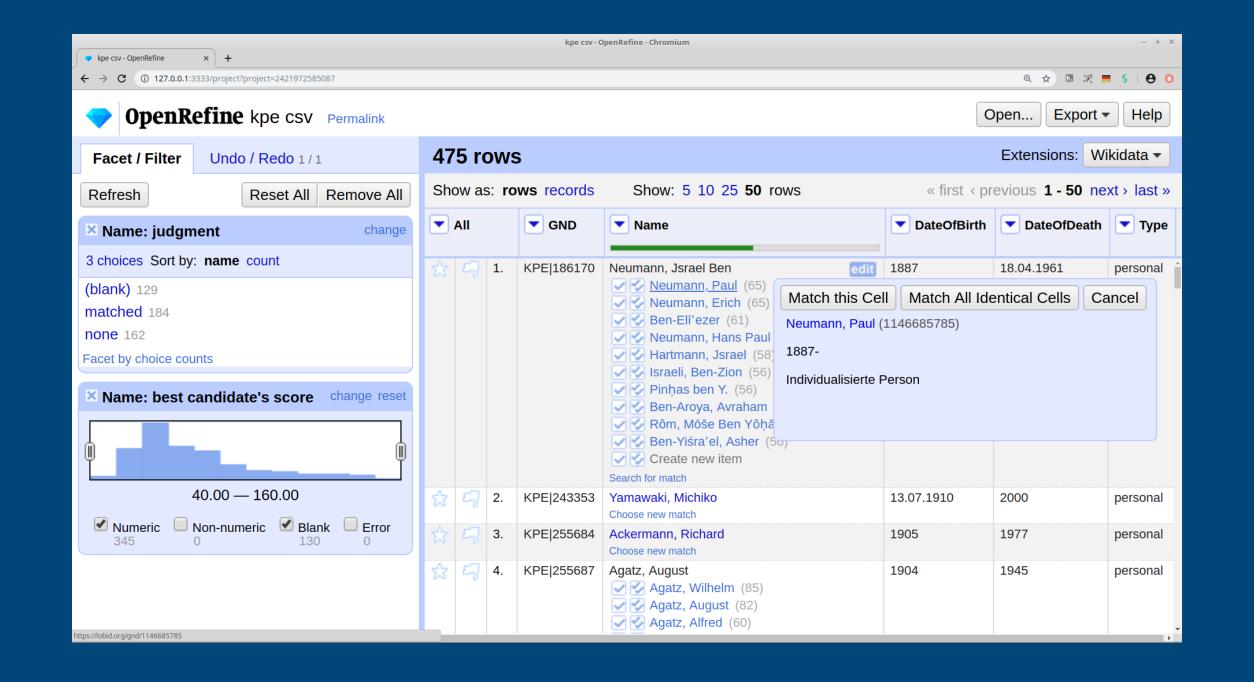
correspSearch

Briefeditionen durchsuchen und vernetzen

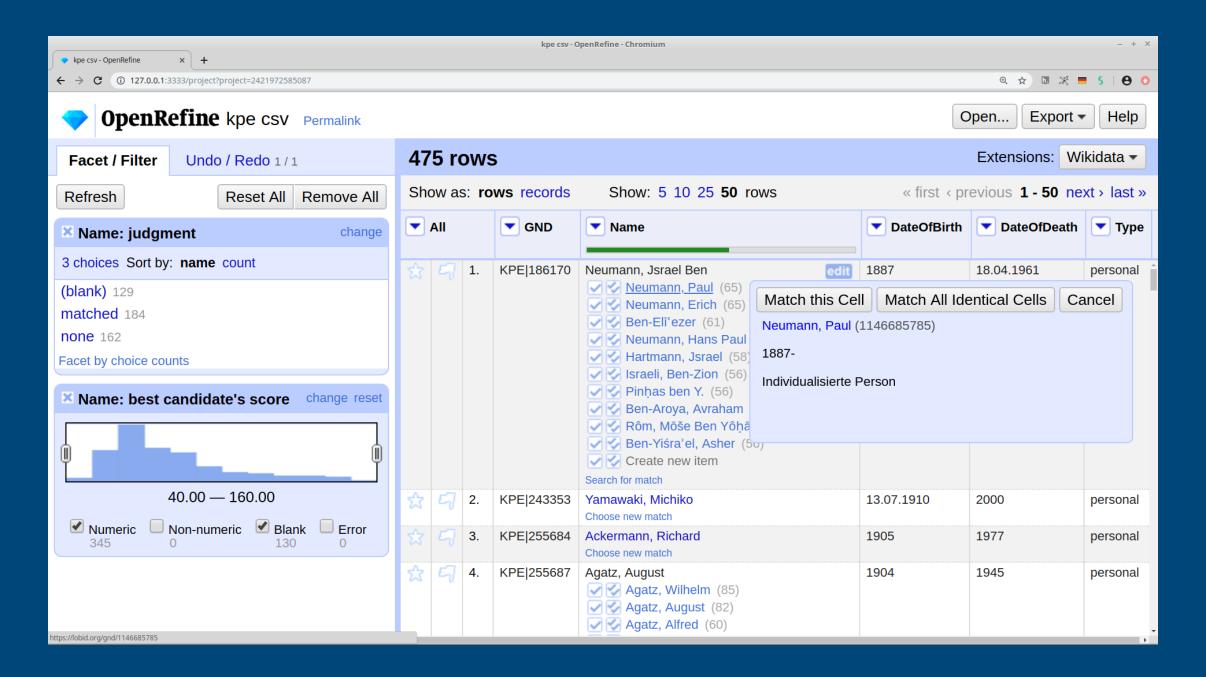
#### bauhaus-archiv

deutsches literatur archiv marbach

# Abgleich und Anreicherung lokaler Archivdaten mit der GND



# Abgleich und Anreicherung lokaler Archivdaten mit der GND



Unter Nutzung von OpenRefine und der entsprechenden lobid Reconciliation API

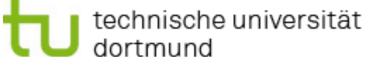




Die Nordrhein-Westfälische Bibliographie



Biographisches Portal NRW ulb Münster





deutsches literatur archiv marbach



RHEINLAND-PFALZ

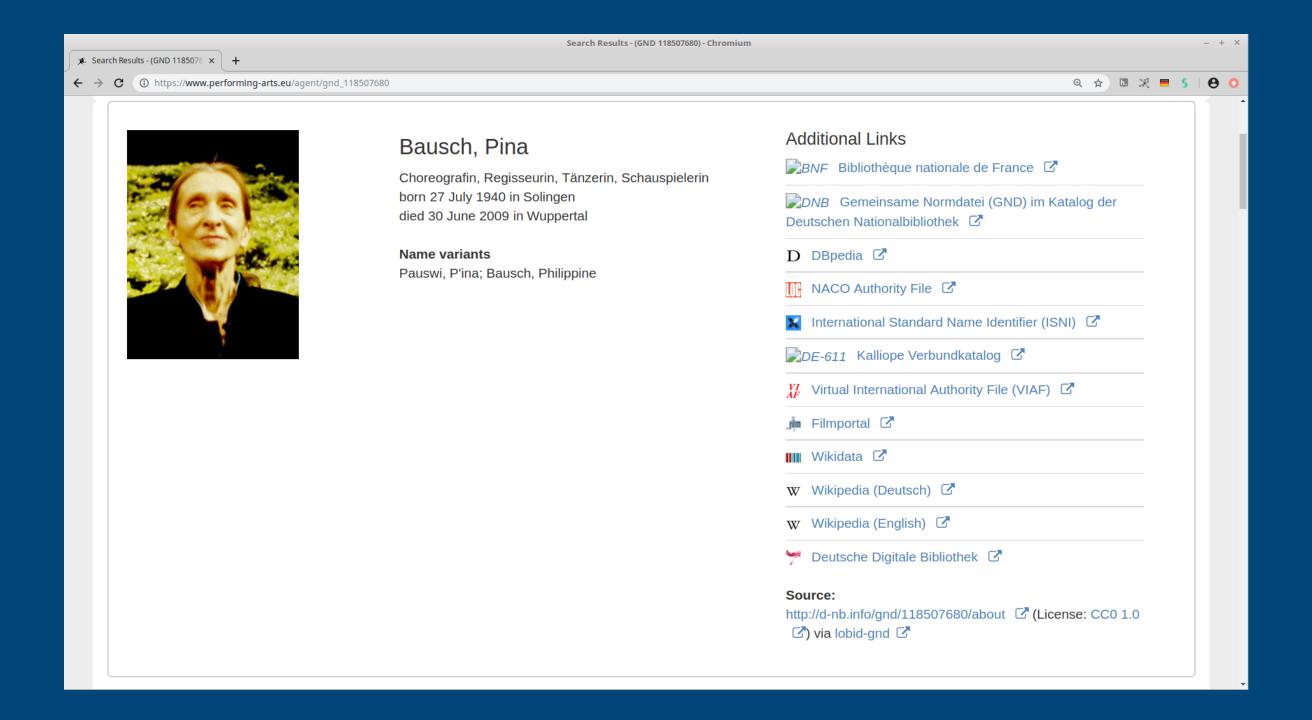


correspSearch

Briefeditionen durchsuchen und vernetzen

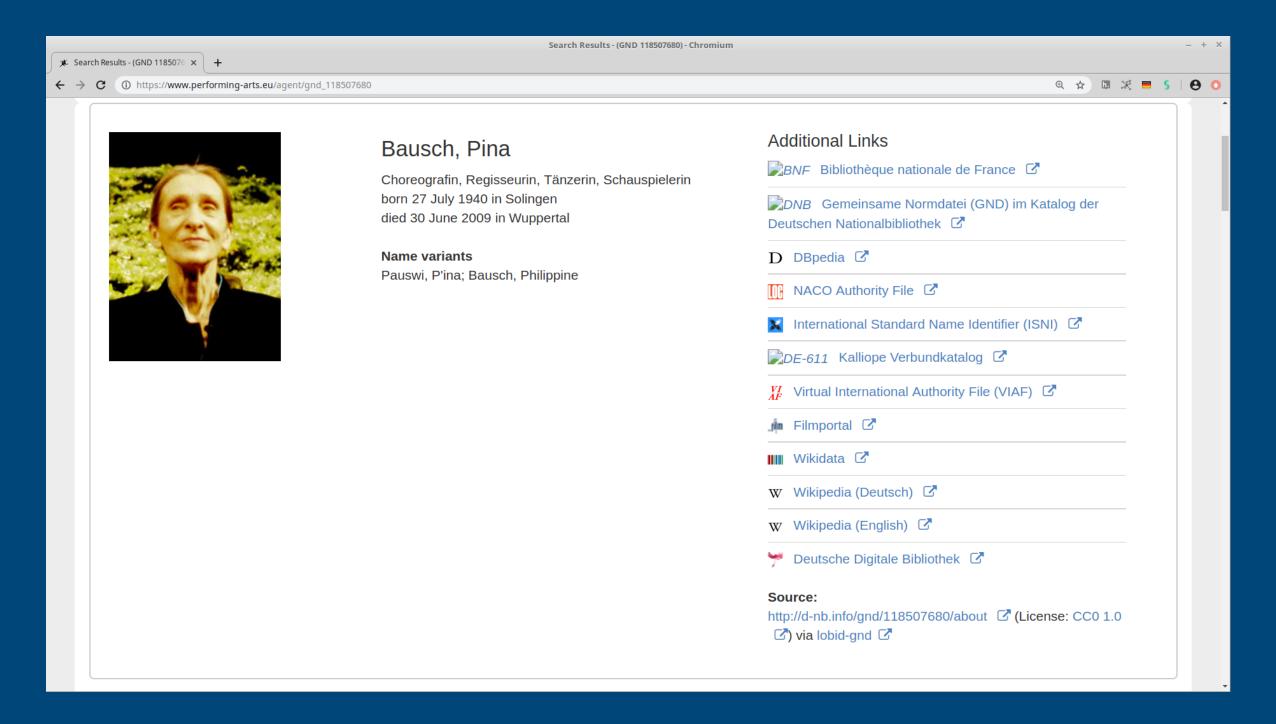
Д

#### Infoboxen zu Personen aus der GND



Д

#### Infoboxen zu Personen aus der GND



Außerdem Nutzung der lobid Reconciliation API zum Mappen großer Mengen von Personendaten auf die GND

4







Die Nordrhein-Westfälische Bibliographie







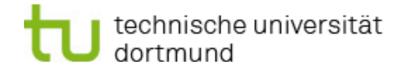
Virtuelle Deutsche Landesbibliographie

Meta-Katalog zum Nachweis landeskundlicher Literatur in Deutschland





Biographisches Portal NRW ulb Münster









correspSearch Briefeditionen durchsuchen und vernetzen

#### bauhaus-archiv

deutsches literatur archiv marbach

# Beispiel: lobid-gnd

Rechercheoberfläche & LOD-API für die GND

Normdatei für Bibliotheken im deutschsprachigen Raum

Normdatei für Bibliotheken im deutschsprachigen Raum

Datensätze für Personen, Körperschaften, Konferenzen & Veranstaltungen, Geografika, Schlagwörter, Werke

Normdatei für Bibliotheken im deutschsprachigen Raum

Datensätze für Personen, Körperschaften, Konferenzen & Veranstaltungen, Geografika, Schlagwörter, Werke

Für die formale Erfassung einer Ressource und für die inhaltliche Erschließung

Normdatei für Bibliotheken im deutschsprachigen Raum

Datensätze für Personen, Körperschaften, Konferenzen & Veranstaltungen, Geografika, Schlagwörter, Werke

Für die formale Erfassung einer Ressource und für die inhaltliche Erschließung

In den letzten Jahren wird die GND vermehrt auch von Archiven, Museen und Wissenschaftler\*innen genutzt

### Die GND-Kooperative

### Die GND-Kooperative

Verantwortlich für die GND ist die GND-Kooperative

### Die GND-Kooperative

Verantwortlich für die GND ist die GND-Kooperative

Die GND-Kooperative besteht hauptsächlich aus den deutschsprachigen Bibliotheksverbünden, den angeschlossenen Bibliotheken sowie der Deutschen Nationalbibliothek (DNB) als technischer Hoster

Die DNB publiziert die GND unter anderem als Linked Open Data (LOD) unter https://data.dnb.de/opendata/

Die DNB publiziert die GND unter anderem als Linked Open Data (LOD) unter https://data.dnb.de/opendata/

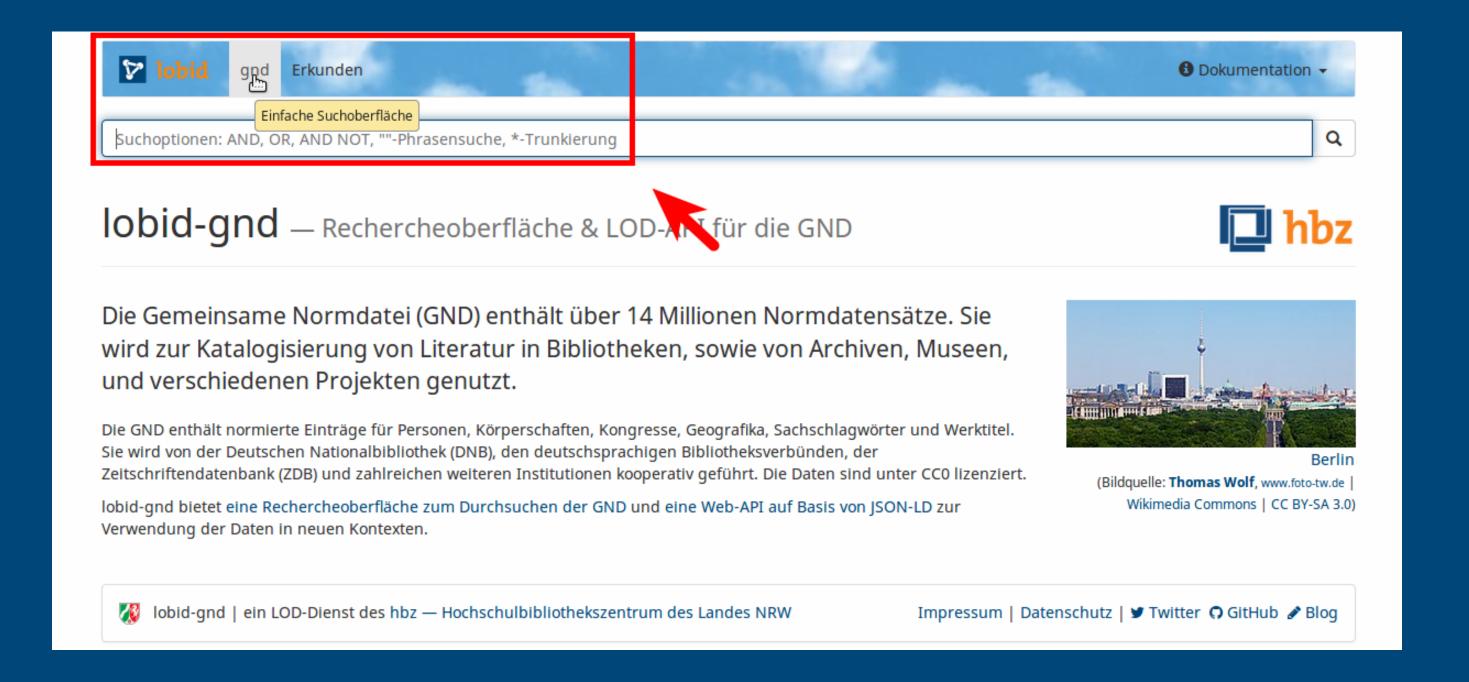
Zusätzliche Anreicherungen, z.B. Links in EntityFacts

Die DNB publiziert die GND unter anderem als Linked Open Data (LOD) unter https://data.dnb.de/opendata/

Zusätzliche Anreicherungen, z.B. Links in EntityFacts

Diese Daten bilden die Grundlage für lobid-gnd

#### Die Oberfläche

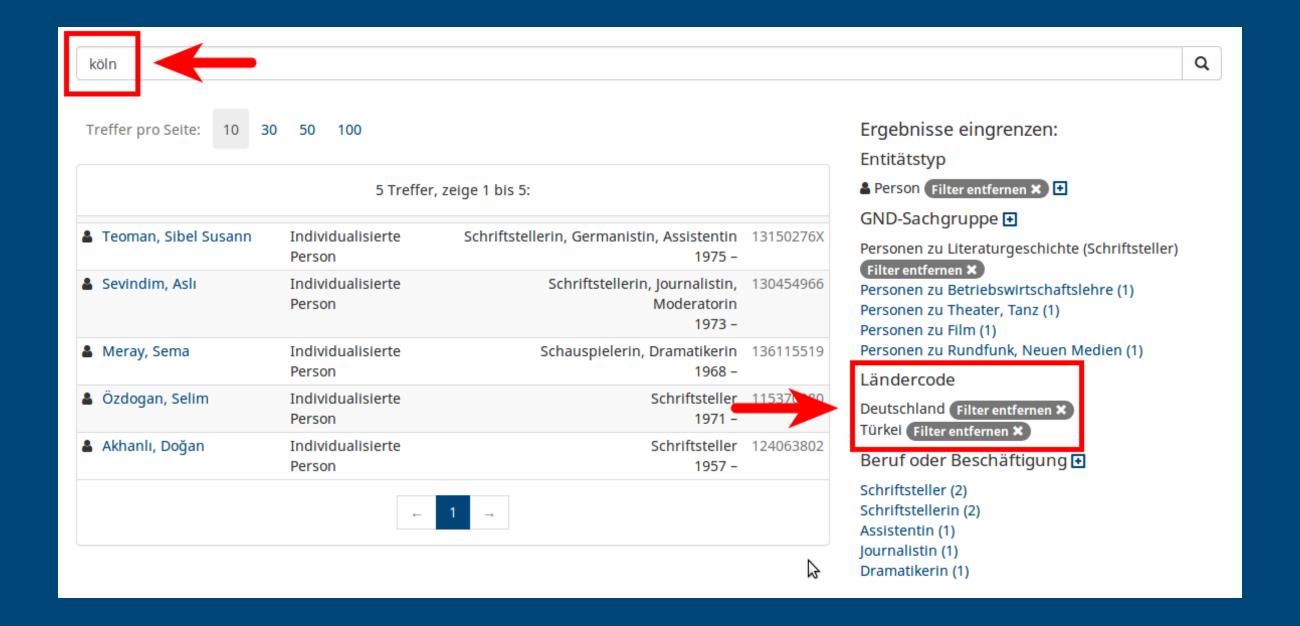


A

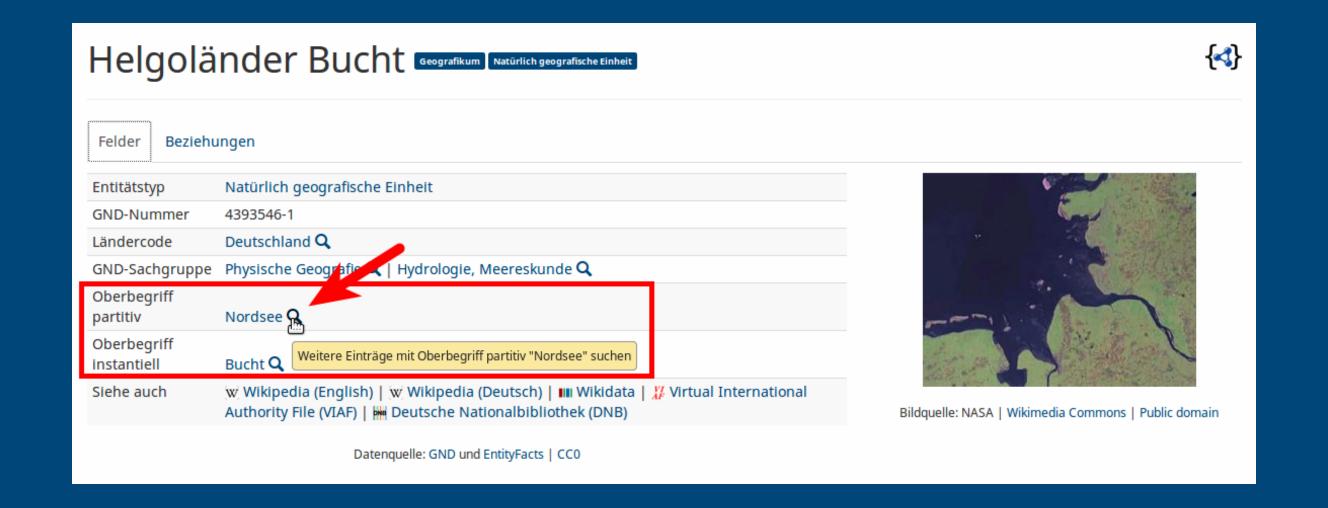
#### Auto Suggest



#### Ergebnisliste

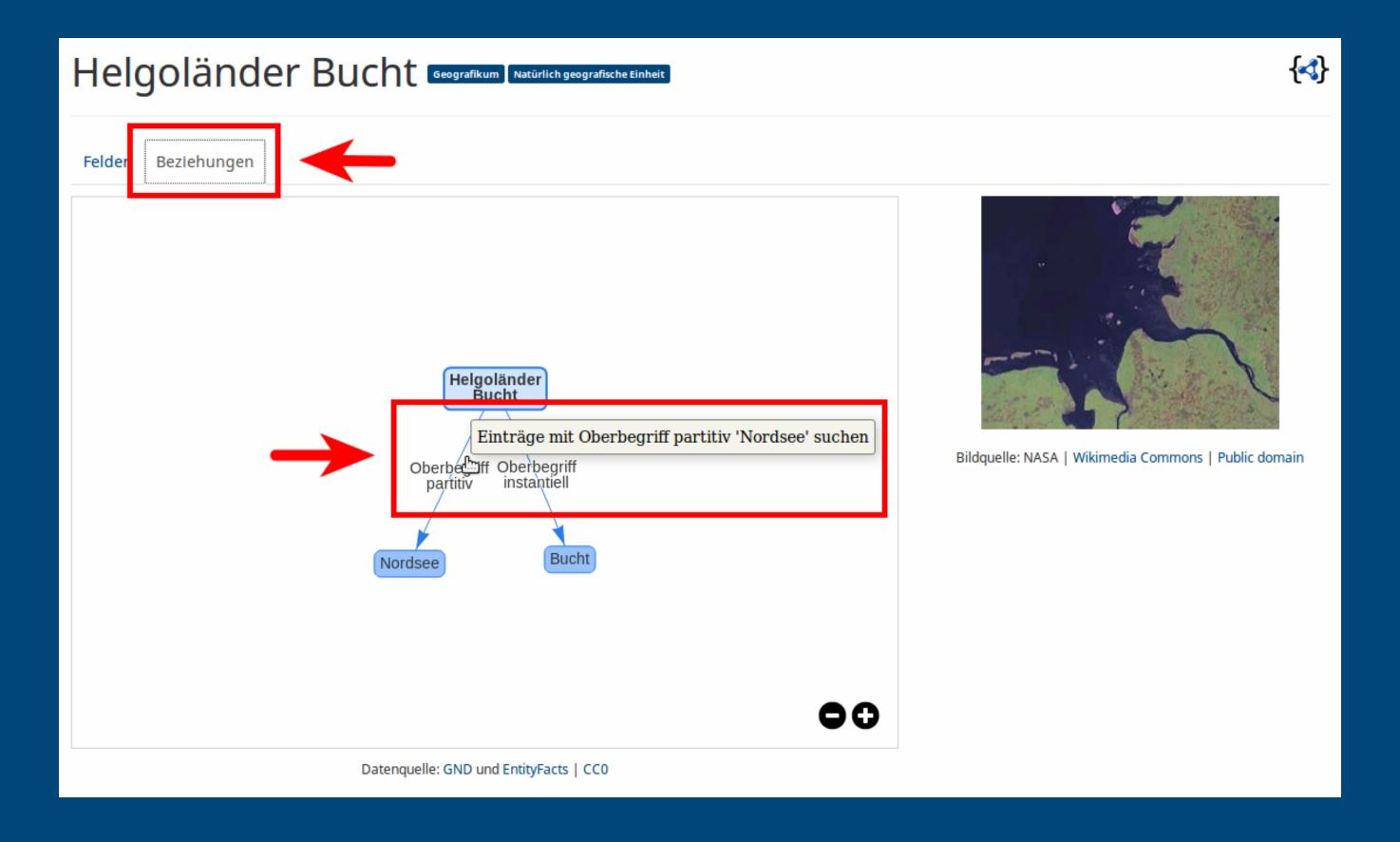


#### Einzeltreffer



 $\mathbf{A}$ 

#### Beziehungsgraph



# Übung

Probiert die auf den vorigen Folien beschriebene Oberfläche von lobid-gnd aus (die Screenshots verlinken auf die entsprechenden Seiten)

## Teil 2: Linked Data & JSON-LD

Standards	RDF, Linked Data	11:00–11:30
APIs	JSON, JSON-LD, LOUD	11:30–12:00
lobid-gnd	Daten und Abfragemöglichkeiten GND	12:00–12:30
lobid- resources	Daten und Abfragemöglichkeiten Verbundkatalog	13:30–14:00

# Übung

Verschicke einen Link zum Buch "With Reference to Reference" von Catherine Elgin

1: OPAC -> http://okeanos-www.hbz-nrw.de/F

2: lobid-resources -> https://lobid.org/resources

Überführung traditioneller bibl. Praktiken in das Web

Überführung traditioneller bibl. Praktiken in das Web Sichtbarkeit und Auffindbarkeit im Web erreichen

Überführung traditioneller bibl. Praktiken in das Web Sichtbarkeit und Auffindbarkeit im Web erreichen Nachnutzbarkeit ermöglichen

Überführung traditioneller bibl. Praktiken in das Web
Sichtbarkeit und Auffindbarkeit im Web erreichen
Nachnutzbarkeit ermöglichen
Synergieeffekte durch Verlinkung mit anderen Daten

Überführung traditioneller bibl. Praktiken in das Web
Sichtbarkeit und Auffindbarkeit im Web erreichen
Nachnutzbarkeit ermöglichen
Synergieeffekte durch Verlinkung mit anderen Daten
Verbesserung der Recherchemöglichkeiten

Überführung traditioneller bibl. Praktiken in das Web
Sichtbarkeit und Auffindbarkeit im Web erreichen
Nachnutzbarkeit ermöglichen
Synergieeffekte durch Verlinkung mit anderen Daten
Verbesserung der Recherchemöglichkeiten

Quelle: Pohl, Adrian / Ostrowski, Felix (2010): 'Linked Data' - und warum wir uns im hbz-Verbund damit beschäftigen." B.I.T. Online 13(3): S. 259-268. Preprint: http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl\_ostrowski\_2010\_linked-data.pdf

### Linked Data: Best Practices

1. Nutze URIs als Namen für Dinge

- 1. Nutze URIs als Namen für Dinge
- 2. Nutze HTTP-URIs, so dass Menschen sie aufrufen können

- 1. Nutze URIs als Namen für Dinge
- 2. Nutze HTTP-URIs, so dass Menschen sie aufrufen können
- 3. Wenn jemand einen URI aufruft, biete nützliche Informationen an unter Nutzung der Standards (RDF, SPARQL)

- 1. Nutze URIs als Namen für Dinge
- 2. Nutze HTTP-URIs, so dass Menschen sie aufrufen können
- 3. Wenn jemand einen URI aufruft, biete nützliche Informationen an unter Nutzung der Standards (RDF, SPARQL)
- 4. Nimm Links zu anderen URIs auf, so dass weitere Dinge entdeckt werden können.

- 1. Nutze URIs als Namen für Dinge
- 2. Nutze HTTP-URIs, so dass Menschen sie aufrufen können
- 3. Wenn jemand einen URI aufruft, biete nützliche Informationen an unter Nutzung der Standards (RDF, SPARQL)
- 4. Nimm Links zu anderen URIs auf, so dass weitere Dinge entdeckt werden können.

Tim Berners-Lee (2006ff): Linked Data – Design Issues

#### RDF & The Semantic Web

### RDF & The Semantic Web – cutting edge seit 1999

### Linked Data

Ultimately, RDF and the Semantic Web are of no interest to Web developers. They also have a really negative public perception problem. We should stop talking about them. Let's shift the focus to be on Linked Data, explaining the problems that Web developers face today, and concrete, demonstrable solutions to those problems.

### Linked Data

Ultimately, RDF and the Semantic Web are of no interest to Web developers. They also have a really negative public perception problem. We should stop talking about them. Let's shift the focus to be on Linked Data, explaining the problems that Web developers face today, and concrete, demonstrable solutions to those problems.

 Manu Sporny, damals Vorsitzender der RDFa Working Group beim W3C, der JSON-LD Community Group & Mitglied weiterer Semantic-Web-Gruppen, beim Schreiben an der JSON-LD-Spezifikation Sporny (2012)

## Linked Open Data

Wissen ist offen, wenn jedeR darauf frei zugreifen, es nutzen, verändern und teilen kann – eingeschränkt höchstens durch Maßnahmen, die Ursprung und Offenheit des Wissens bewahren.

## Linked Open Data

Wissen ist offen, wenn jedeR darauf frei zugreifen, es nutzen, verändern und teilen kann – eingeschränkt höchstens durch Maßnahmen, die Ursprung und Offenheit des Wissens bewahren.

http://opendefinition.org/od/2.1/de/

### Linked Open Usable Data



#### **Rob Sanderson**

@azaroth42



"Open" is not necessarily simple nor useful, just available. We need fans to get LOUD: Linked Open Useful Data:)

Tweet übersetzen

19:39 - 23. Aug. 2016

Quelle: Rob Sanderson auf Twitter

Siehe auch Rob Sanderson's Europeanatech 2018 Keynote (Slides, Aufzeichnung)

Daten werden mit existierender Software bearbeitet (ausgewertet, ergänzt, integriert etc.)

Daten werden mit existierender Software bearbeitet (ausgewertet, ergänzt, integriert etc.)

Entwicklung neuer Software zur Interaktion mit Daten

Daten werden mit existierender Software bearbeitet (ausgewertet, ergänzt, integriert etc.)

Entwicklung neuer Software zur Interaktion mit Daten

LOUD: Orientierung auf Bedürfnisse und Konventionen rund um Software (Entwicklung, Standards, etc.)

Daten werden mit existierender Software bearbeitet (ausgewertet, ergänzt, integriert etc.)

Entwicklung neuer Software zur Interaktion mit Daten

LOUD: Orientierung auf Bedürfnisse und Konventionen rund um Software (Entwicklung, Standards, etc.)

Nützliche Daten: Zielgruppe kennen & eigene Angebote auf sie ausrichten

Daten werden mit existierender Software bearbeitet (ausgewertet, ergänzt, integriert etc.)

Entwicklung neuer Software zur Interaktion mit Daten

LOUD: Orientierung auf Bedürfnisse und Konventionen rund um Software (Entwicklung, Standards, etc.)

Nützliche Daten: Zielgruppe kennen & eigene Angebote auf sie ausrichten

Hauptzielgruppe: Entwickler\*innen oder Nutzer\*innen von *Software* für Datenzugriff und -manipulation

Application Programming Interfaces (Programmierschnittstellen)

Application Programming Interfaces (Programmierschnittstellen)

Software baut auf APIs auf

Application Programming Interfaces (Programmierschnittstellen)

Software baut auf APIs auf

APIs machen Softwareentwicklung handhabbar (für 1st- und 2nd-Party-Software)

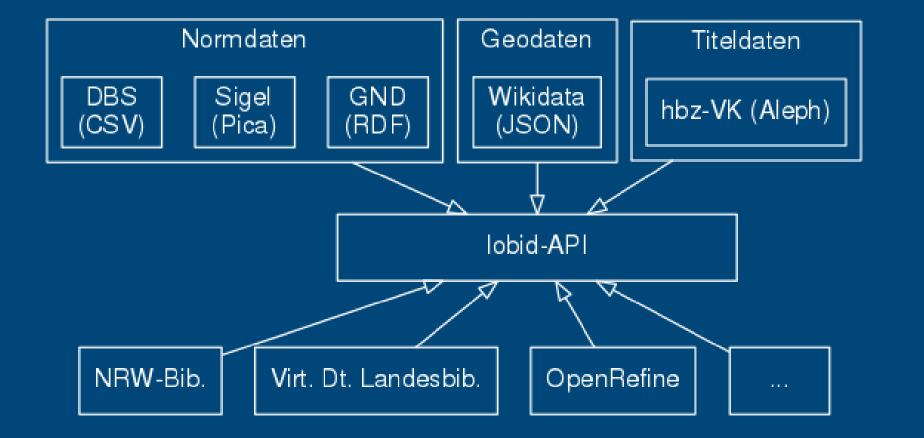
Application Programming Interfaces (Programmierschnittstellen)

Software baut auf APIs auf

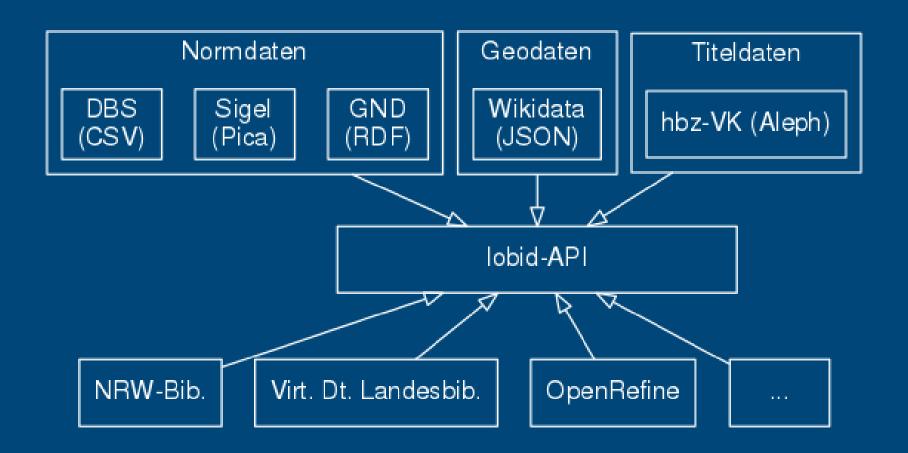
APIs machen Softwareentwicklung handhabbar (für 1st- und 2nd-Party-Software)

APIs ermöglichen Nutzung und Integration von 3rd-Party-Software

#### Zum Bsp. lobid-Formate und -Anwendungen



#### Zum Bsp. lobid-Formate und -Anwendungen



APIs entkoppeln Anwendungen von Datenquellen, Formaten und Systemen. Sie ermöglichen so modulare, zukunftsfähige Applikationen

#### Und wie APIs bereitstellen?

#### Und wie APIs bereitstellen?

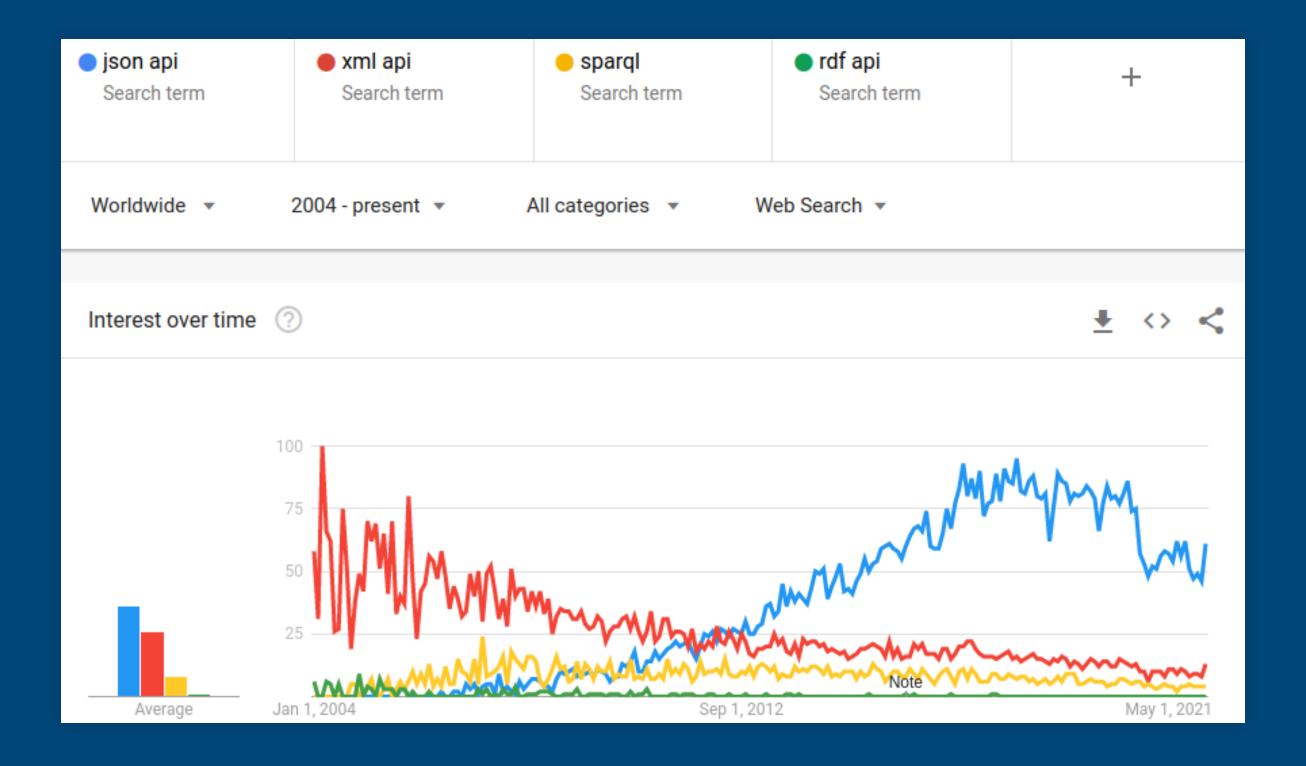
JSON über HTTP

#### Und wie APIs bereitstellen?

JSON über HTTP

Der Web-API-Standard seit Jahren, siehe z.B. Target (2017)

#### The Rise and Rise of JSON



Quellen: Google Trends, Web Data Commons, W3Techs

Ein einfaches Key-Value-Format für strukturierte Daten

Ein einfaches Key-Value-Format für strukturierte Daten Key ist immer ein String

Ein einfaches Key-Value-Format für strukturierte Daten

Key ist immer ein String

Value ist String, Number, Boolean, Array oder Object

## SON

Ein einfaches Key-Value-Format für strukturierte Daten Key ist immer ein String

Value ist String, Number, Boolean, Array oder Object

{ "foo": "bar" }

Ein einfaches Key-Value-Format für strukturierte Daten Key ist immer ein String

Value ist String, Number, Boolean, Array oder Object

```
{ "foo": "bar" }
```

Quelle: RFC 8259

# Beispiel: GET https://lobid.org/organisations/DE-1a

```
{
  "linkedTo" : {
    "id" : "http://lobid.org/organisations/DE-601#!",
    "label" : "Verbundzentrale des GBV (VZG)"
},
  "rs" : "1100000000000",
  "address" : {
    "addressLocality" : "Berlin",
    "type" : "PostalAddress",
    "addressCountry" : "DE",
    "postalCode" : "10772"
}
...
```

Kommandozeilen-Tool zum Datentransfer mit URLs

Kommandozeilen-Tool zum Datentransfer mit URLs

https://curl.haxx.se/download.html

Kommandozeilen-Tool zum Datentransfer mit URLs

https://curl.haxx.se/download.html

https://reqbin.com/curl

Kommandozeilen-Tool zum Datentransfer mit URLs

https://curl.haxx.se/download.html

https://reqbin.com/curl

\$ curl "https://lobid.org"

## Übung: cURL

```
CURL $ curl "https://lobid.org/organisations/DE-1a"

Copy  # Kopieren und im Terminal einfügen (s. Hinweise unten)
$ curl "https://lobid.org/organisations/DE-1a"

History $ † # Auf die Kommandozeilen-History zugreifen $ curl "https://lobid.org/organisations/DE-1a"
```

(Für Copy/Paste in GUI: CTRL+c bzw. CTRL+v.

*Im Terminal: SHIFT+CTRL+c bzw. SHIFT+CTRL+v. )* 

U

#### Antwort

```
$ curl https://lobid.org/organisations/DE-1a
  "linkedTo" : {
    "id" : "http://lobid.org/organisations/DE-601#!",
    "label": "Verbundzentrale des GBV (VZG)"
 },
  "rs" : "1100000000000,
  "address" : {
    "addressLocality" : "Berlin",
    "type" : "PostalAddress",
    "addressCountry" : "DE",
    "postalCode" : "10772"
 },
```

Ausgabe ist lang, oft wollen wir nur bestimmte Werte

F

### JavaScript Object Notation (JSON)

# JavaScript Object Notation (JSON)

```
var options = {
  url: 'https://lobid.org/organisations/DE-1a'
};

request(options, function (error, response, body) {
  var doc = JSON.parse(body);
  console.log('postal code:', doc.address.postalCode) // <--
});</pre>
```

# JavaScript Object Notation (JSON)

```
var options = {
  url: 'https://lobid.org/organisations/DE-1a'
};

request(options, function (error, response, body) {
  var doc = JSON.parse(body);
  console.log('postal code:', doc.address.postalCode) // <---
});</pre>
```

> postal code: 10772

# j

# jq

Ein flexibles Kommandozeilen-Tool zur JSON-Verarbeitung

# jq

Ein flexibles Kommandozeilen-Tool zur JSON-Verarbeitung

https://stedolan.github.io/jq/

# jq

Ein flexibles Kommandozeilen-Tool zur JSON-Verarbeitung

https://stedolan.github.io/jq/

https://jqplay.org/

Ein flexibles Kommandozeilen-Tool zur JSON-Verarbeitung

https://stedolan.github.io/jq/

https://jqplay.org/

\$ jq --help

# Übung: JSON mit jq (1)

```
$ curl "https://lobid.org/organisations/DE-1a" \
| jq .name # filter: .name
```

## Ubung: JSON mit jq (2)

```
Kosten-
träger
```

```
$ curl "https://lobid.org/organisations/DE-1a"
jq <???> # dt. Bezeichner für den Kostenträger
```

```
datum
```

```
Modifikations- $ curl "https://lobid.org/organisations/DE-1a"
                 jq <???> # Datum der letzten Modifikation
```

## Ubung: JSON mit jq (3)

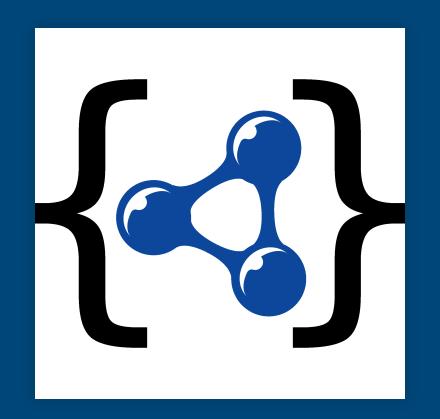
```
Kosten-
träger
```

```
$ curl "https://lobid.org/organisations/DE-1a"
jq .fundertype.label.de
```

```
datum
```

```
Modifikations- $ curl "https://lobid.org/organisations/DE-1a"
                 jq .mainEntityOfPage.dateModified
```

#### JSON + Linked Data = JSON-LD



"designed to be usable directly as JSON, with no knowledge of RDF" – Es ist richtiges JSON!

"designed to be usable directly as JSON, with no knowledge of RDF" – Es ist richtiges JSON!

"also designed to be usable as RDF"

"designed to be usable directly as JSON, with no knowledge of RDF" – Es ist richtiges JSON!

"also designed to be usable as RDF"

https://www.w3.org/TR/json-ld/

# JSON

```
$ curl -H "Accept: application/json"
https://api.github.com/users/acka47
```

```
{
  "login": "acka47",
  "avatar_url":
"https://avatars2.githubusercontent.com/u/160292?v=4",
  "url": "https://api.github.com/users/acka47",
  "type": "User",
  "name": "Adrian",
  "company": "hbz",
  "location": "Cologne, Germany",
  "bio": "Metadata, RDF, vocabularies. Working at @hbz."
}
```

F

#### JSON + @context + @id = JSON-LD

```
"@context": "http://schema.org/",
 "@id": "https://github.com/users/acka47",
 "login": "acka47",
 "avatar_url":
"https://avatars2.githubusercontent.com/u/160292?v=4",
 "url": "https://api.github.com/users/acka47",
 "type": "User",
 "name": "Adrian",
 "company": "hbz",
 "location": "Cologne, Germany",
  "bio": "Metadata, RDF, vocabularies. Working at @hbz. "
```

F

2010-2013: Alpha-Betrieb mit RDF & SPARQL

2010-2013: Alpha-Betrieb mit RDF & SPARQL

2013-2017: lobid v1.x mit JSON-LD aber noch nicht sehr intuitiv nutzbar (JSON-LD als RDF-Serialisierung)

2010-2013: Alpha-Betrieb mit RDF & SPARQL

2013-2017: lobid v1.x mit JSON-LD aber noch nicht sehr intuitiv nutzbar (JSON-LD als RDF-Serialisierung)

seit 2017: richtig LOUDe lobid v2 plus OpenRefine-Schnittstellen

2010-2013: Alpha-Betrieb mit RDF & SPARQL

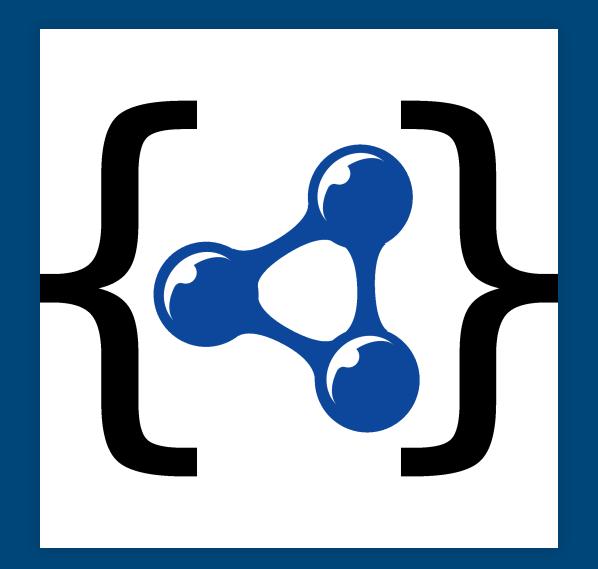
2013-2017: lobid v1.x mit JSON-LD aber noch nicht sehr intuitiv nutzbar (JSON-LD als RDF-Serialisierung)

seit 2017: richtig LOUDe lobid v2 plus OpenRefine-Schnittstellen

Von LOD zu LOUD – Erfahrungen aus zehn Jahren Linked-Open-Data-Entwicklung am hbz

# Beispiel: lobid-gnd

#### a. Die Daten



#### JSON-LD-Link

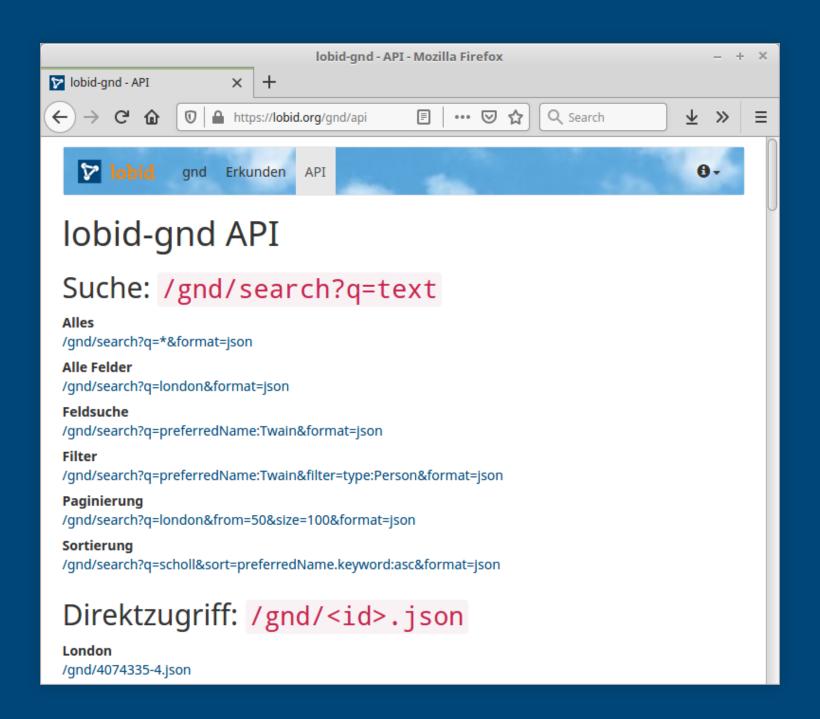
#### JSON(-LD)

```
lobid.org/gnd/11850391X.json - Chromium
lobid.org/gnd/11850391 ×
← → C (i) lobid.org/gnd/11850391X.json
                                                                                Q & X 🗆
₩ {
                                                                              Raw
                                                                                        Parsed
      "gender": [
       ₩ {
              "id": "http://d-nb.info/standards/vocab/gnd/gender#female",
              "label": "Weiblich"
    "dateOfDeath": [
          "1975-12-04"
    "placeOfDeath": [
       ₩ {
              "id": "http://d-nb.info/gnd/4042011-5",
              "label": "New York, NY"
    ▼ "familialRelationship": [
              "id": "http://d-nb.info/gnd/119378418",
              "label": "Blücher, Heinrich"
              "id": "http://d-nb.info/gnd/118502751",
              "label": "Anders, Günther"
```

```
acka47@homo: ~
Datei Bearbeiten Ansicht Suchen Terminal Hilfe
acka47@homo:~$ curl http://lobid.org/gnd/11850391X
 "gender" : [ {
    "id" : "http://d-nb.info/standards/vocab/gnd/gender#female",
    "label" : "Weiblich"
 } ],
  "dateOfDeath" : [ "1975-12-04" ],
  "placeOfDeath" : [ {
    "id" : "http://d-nb.info/gnd/4042011-5",
    "label" : "New York, NY"
 } ],
  "familialRelationship" : [ {
    "id" : "http://d-nb.info/gnd/119378418",
    "label" : "Blücher, Heinrich"
    "id" : "http://d-nb.info/gnd/118502751",
    "label" : "Anders, Günther"
  } ],
  "type" : [ "DifferentiatedPerson", "AuthorityResource", "Person" ],
  "gndSubjectCategory" : [ {
    "id" : "http://d-nb.info/standards/vocab/gnd/gnd-sc#8.1p",
    "label" : "Personen (Politologen, Staatstheoretiker)"
 }, {
    "id" : "http://d-nb.info/standards/vocab/gnd/gnd-sc#4.7p",
    "label" : "Personen zu Philosophie"
 }, {
    "id" : "http://d-nb.info/standards/vocab/gnd/gnd-sc#9.5p",
    "label" : "Personen zu Soziologie, Gesellschaft, Arbeit, Sozialgeschichte"
```

F

#### b. Web-API



 $\mathsf{F}$ 

JSON-LD-Daten in Elasticsearch-Index

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

Für Einzeltreffer: RDF per Content Negotiation

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

Für Einzeltreffer: RDF per Content Negotiation

API für Auto Suggest

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

Für Einzeltreffer: RDF per Content Negotiation

API für Auto Suggest

API für Bulk Downloads

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

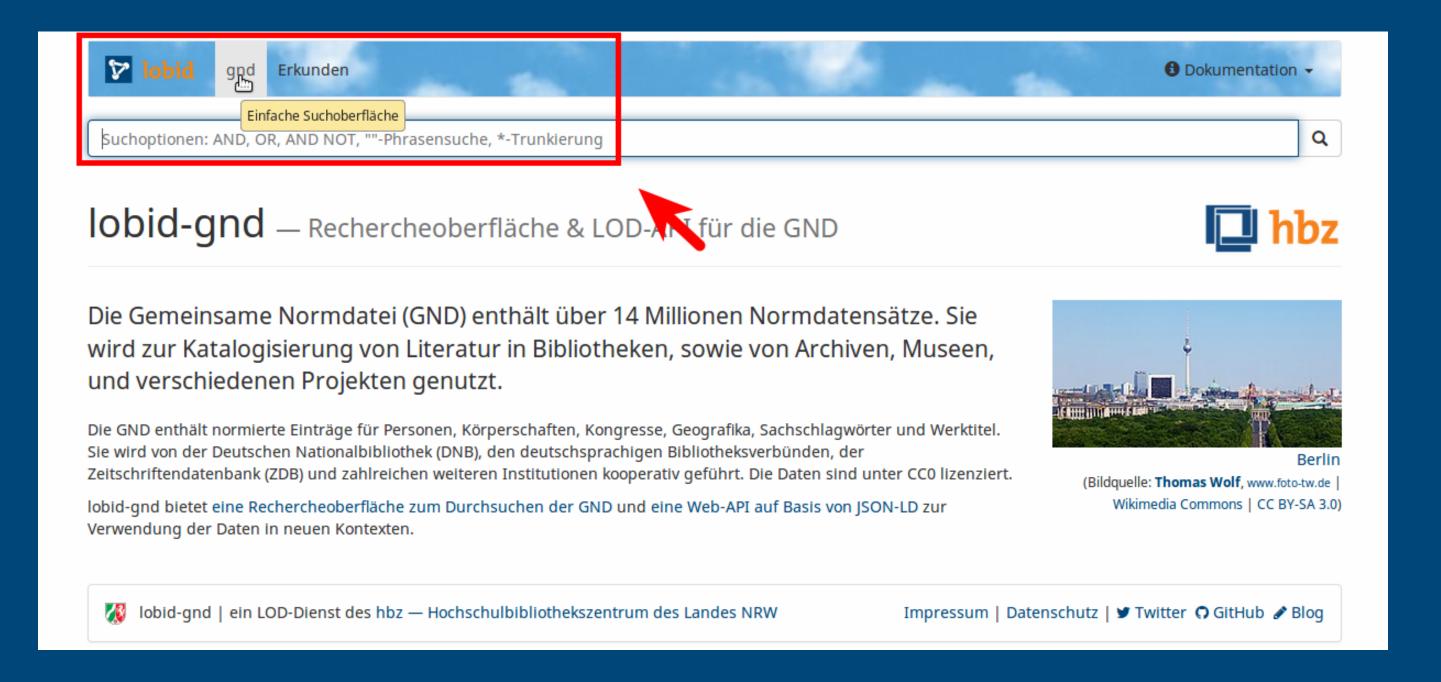
Für Einzeltreffer: RDF per Content Negotiation

API für Auto Suggest

API für Bulk Downloads

OpenRefine Reconciliation | API

### Übung: Suchsyntax



 $\sf U$ 

# Übung: Suchsyntax

```
Feldsuche
                   ??? # <field>:<term>
Boolsche
                   ??? # AND, OR
Suche
Trunkierung
                   ??? # Trunkieren mit *
Intervalle
                   ??? # <feld>:[* TO <???>]
Existenz
                   ??? # _exists_:<field>
```

Elasticsearch 'query string' Dokumentation

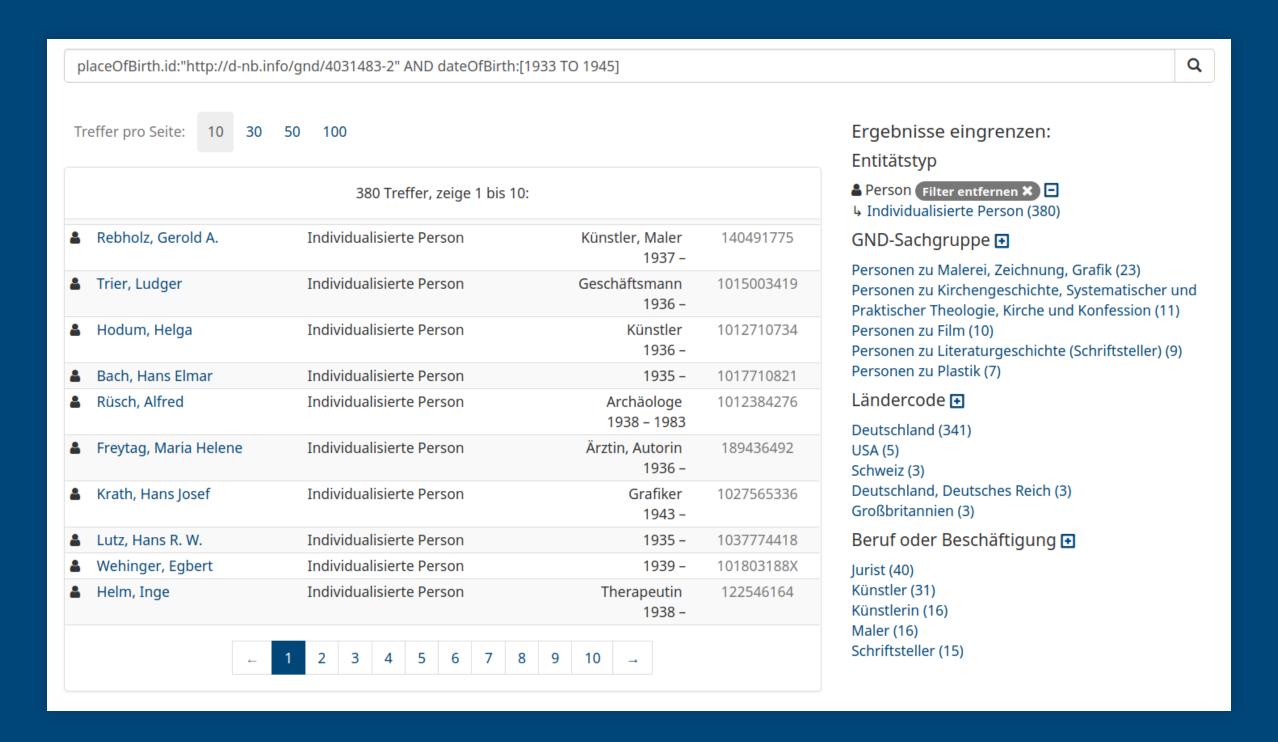
## Suchsyntax: Lösungen

Feldsuche	abbreviatedNameForTheCorporateBody:zbiw
Boolsche Suche	Kunst AND (Kultur OR Wissenschaft)
Trunkierung	professionOrOccupation.label:Politiker*
Intervalle	dateOfBirth:[* TO 1500]
Existenz	_exists_:dateOfBirth

Elasticsearch 'query string' Dokumentation

# Alles kombinierbar

### Personen, die während der NS-Zeit in Köln geboren wurden



# Übung

Suche nach allen Personen in der GND, die in Köln geboren wurden und in Düsseldorf gestorben sind

lobid-gnd: Formulierung komplexer Suchanfragen

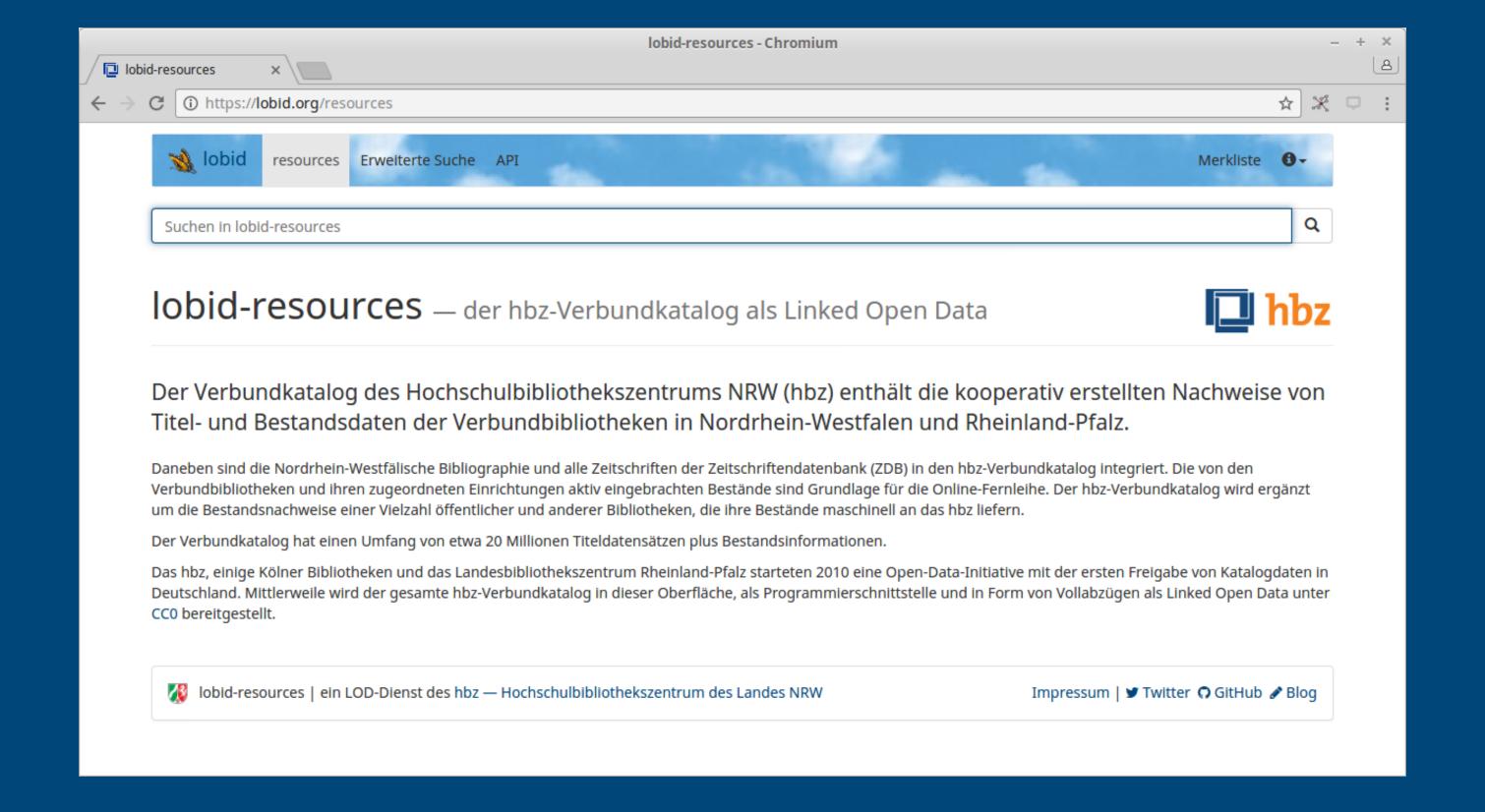
# Übung: Lösung

Suche nach allen Personen in der GND, die in Köln geboren wurden und in Düsseldorf gestorben sind

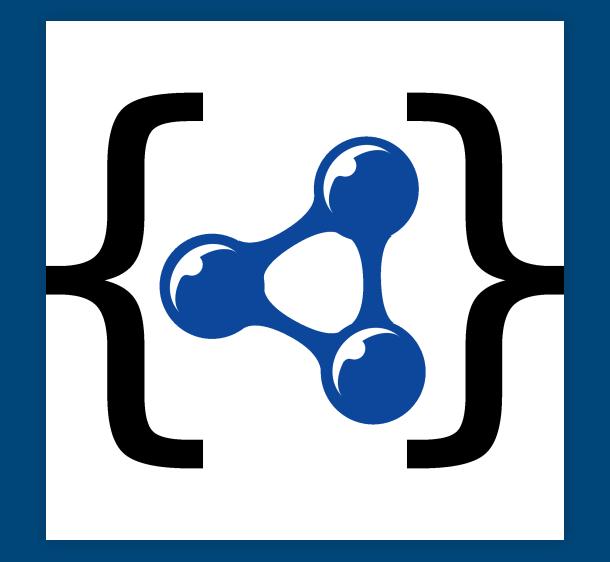
# Pause

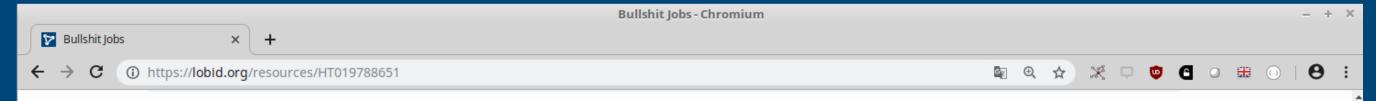
# Beispiel: lobid-resources

#### a. Die Oberfläche



### b. Die Daten



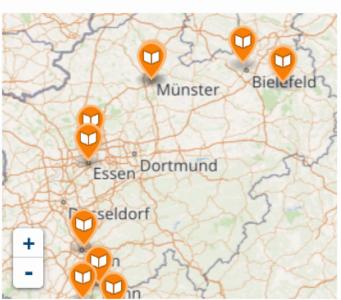


#### **Bullshit Jobs**

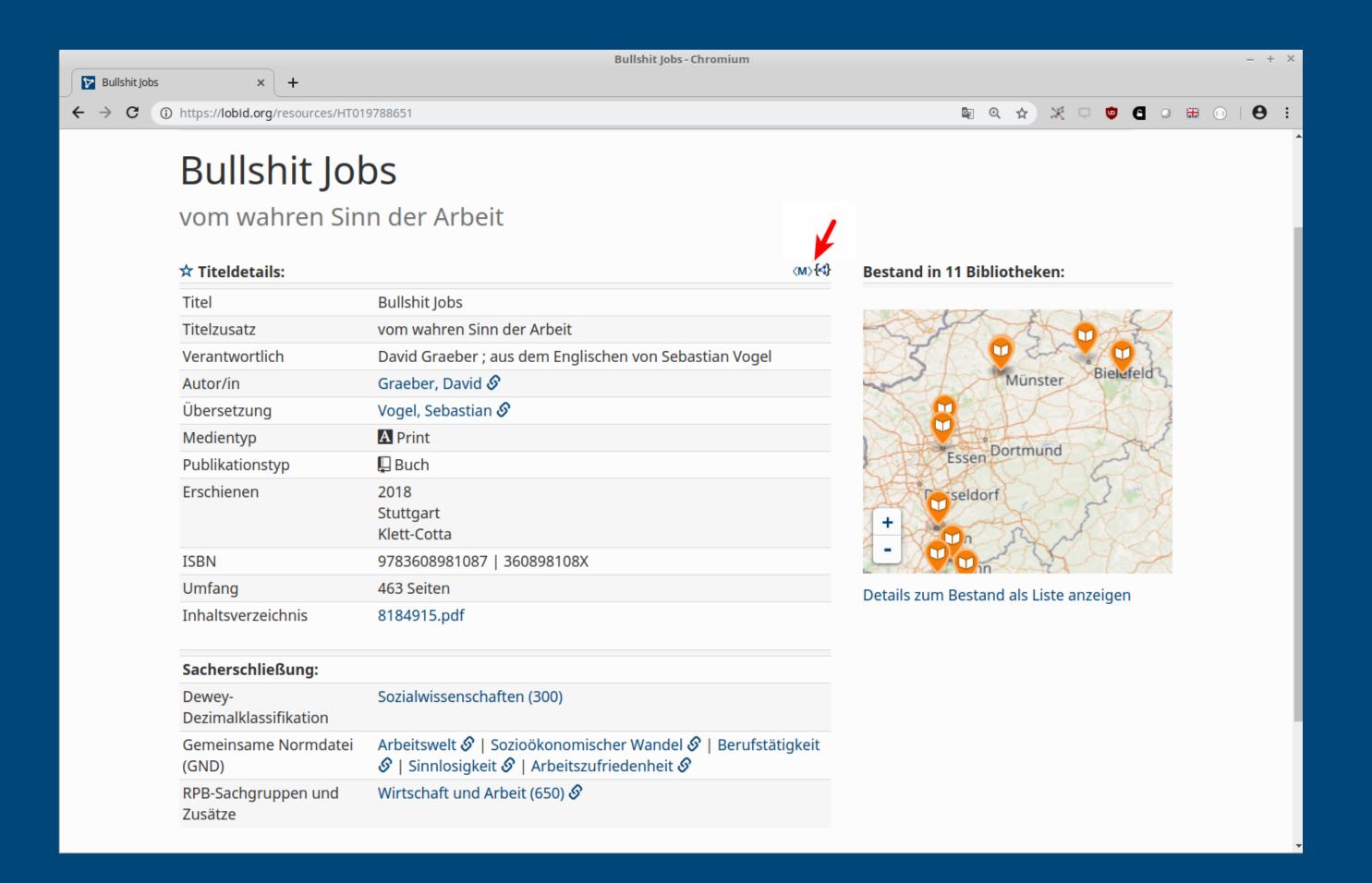
vom wahren Sinn der Arbeit

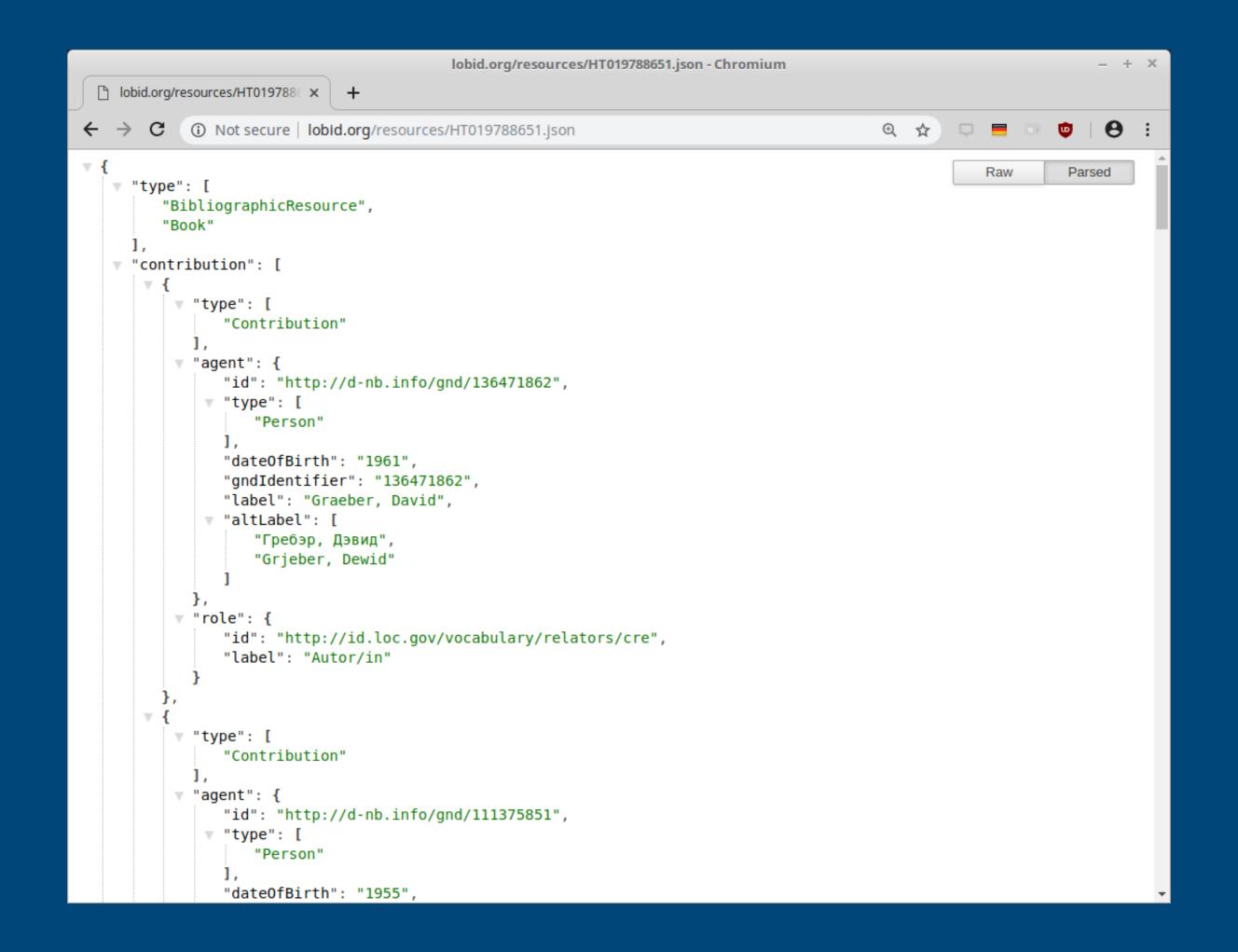


#### Bestand in 11 Bibliotheken:



Details zum Bestand als Liste anzeigen





```
acka47@metameta~
                                                                                                                                    - + \times
Datei Bearbeiten Ansicht Suchen Terminal Hilfe
acka47@metameta ~ $ curl -H "Accept: application/json" http://lobid.org/resources/HT019788651
 "type" : [ "BibliographicResource", "Book" ],
 "contribution" : [ {
   "type" : [ "Contribution" ],
   "agent" : {
     "id" : "http://d-nb.info/gnd/136471862",
     "type" : [ "Person" ],
     "dateOfBirth" : "1961",
     "gndIdentifier" : "136471862",
     "label" : "Graeber, David",
     "altLabel" : [ "Гребэр, Дэвид", "Grjeber, Dewid" ]
   "role" : {
     "id" : "http://id.loc.gov/vocabulary/relators/cre",
     "label" : "Autor/in"
 }, {
   "type" : [ "Contribution" ],
   "agent" : {
     "id" : "http://d-nb.info/gnd/111375851",
     "type" : [ "Person" ],
     "dateOfBirth" : "1955",
     "gndIdentifier" : "111375851",
     "label" : "Vogel, Sebastian"
   "role" : {
     "id" : "http://id.loc.gov/vocabulary/relators/trl",
     "label" : "Übersetzung"
 } ],
 "extent" : "463 Seiten",
 "hasItem" : [ {
   "id" : "http://lobid.org/items/HT019788651:DE-6:001%202620#!",
   "type" : [ "Item" ],
   "heldBy" : {
     "id": "http://lobid.org/organisations/DE-6#!",
     "label" : "lobid Organisation"
   "note" : "00000045",
   "callNumber" : "001 2620",
   "label" : "001 2620"
   "id" : "http://lobid.org/items/HT019788651:DE-1032:Soz%20J%20291#!",
   "type" : [ "Item" ],
```

Tägl. Bezug der XML-Daten über den Aleph-Publishing-Mechanismus

Tägl. Bezug der XML-Daten über den Aleph-Publishing-Mechanismus

Konversion nach N-Triples mit Metafacture (Morph)

Tägl. Bezug der XML-Daten über den Aleph-Publishing-Mechanismus

Konversion nach N-Triples mit Metafacture (Morph)

Dabei finden verschiedene Normalisierungsarbeiten statt

Tägl. Bezug der XML-Daten über den Aleph-Publishing-Mechanismus

Konversion nach N-Triples mit Metafacture (Morph)

Dabei finden verschiedene Normalisierungsarbeiten statt

Labels für verlinkte Ressourcen ergänzen

Tägl. Bezug der XML-Daten über den Aleph-Publishing-Mechanismus

Konversion nach N-Triples mit Metafacture (Morph)

Dabei finden verschiedene Normalisierungsarbeiten statt

Labels für verlinkte Ressourcen ergänzen

Et voilà: Linked Open Usable Data (LOUD)

MARC-Relator-Codes werden aus RAK-Nebeneintragungen generiert (Morph für Personennebeneintragungen)

MARC-Relator-Codes werden aus RAK-Nebeneintragungen generiert (Morph für Personennebeneintragungen)

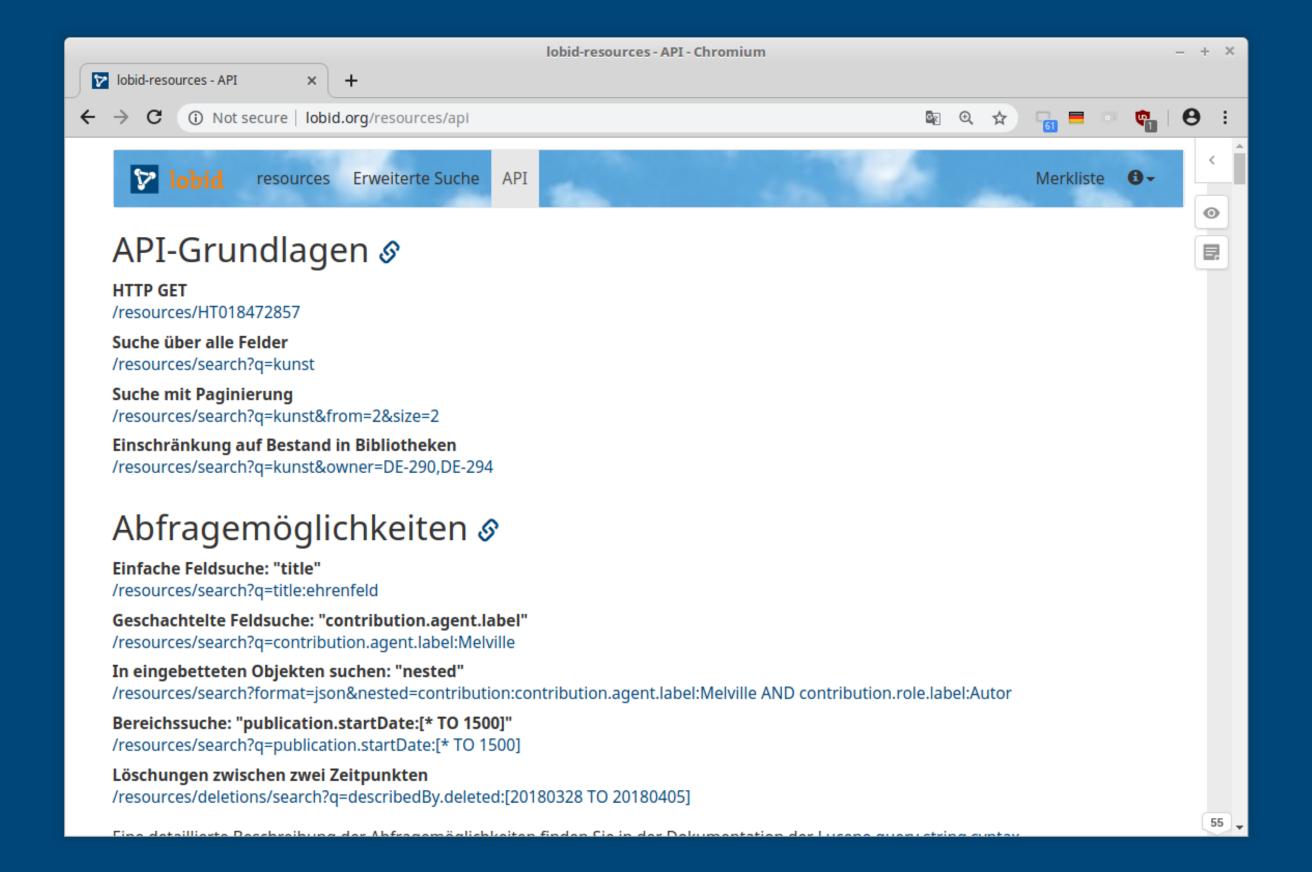
Normalisierung von Jahresangaben

MARC-Relator-Codes werden aus RAK-Nebeneintragungen generiert (Morph für Personennebeneintragungen)

Normalisierung von Jahresangaben

Mapping von Informationen aus 050, 057, 058, 334, 652a, 9--, 400/403 etc. auf eine Liste von Publikations- und Medientypen (Morph)

#### c. Web-API



JSON-LD-Daten in Elasticsearch-Index

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

#### Abfragemöglichkeiten

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

Abfrage nach gelöschten Titeln (ab 2018-04-04)

#### Abfragemöglichkeiten

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

Abfrage nach gelöschten Titeln (ab 2018-04-04)

API für Auto Suggest

#### Abfragemöglichkeiten

JSON-LD-Daten in Elasticsearch-Index

Elasticsearch bzw. Lucene Suchsyntax

Abfrage nach gelöschten Titeln (ab 2018-04-04)

API für Auto Suggest

API für Bulk Downloads

## Beispielabfragen

# Ermitteln eines hbz-Titels anhand der OCLC-Nummer

oclcNumber:601339369

#### RDA-Titel seit 2010

title:rda AND

publication.startDate:[2010 T0 \*]

#### Online-Medien mit Bestand

medium.id:

"http://rdaregistry.info/termList/ RDACarrierType/1018"

AND \_exists\_:hasItem

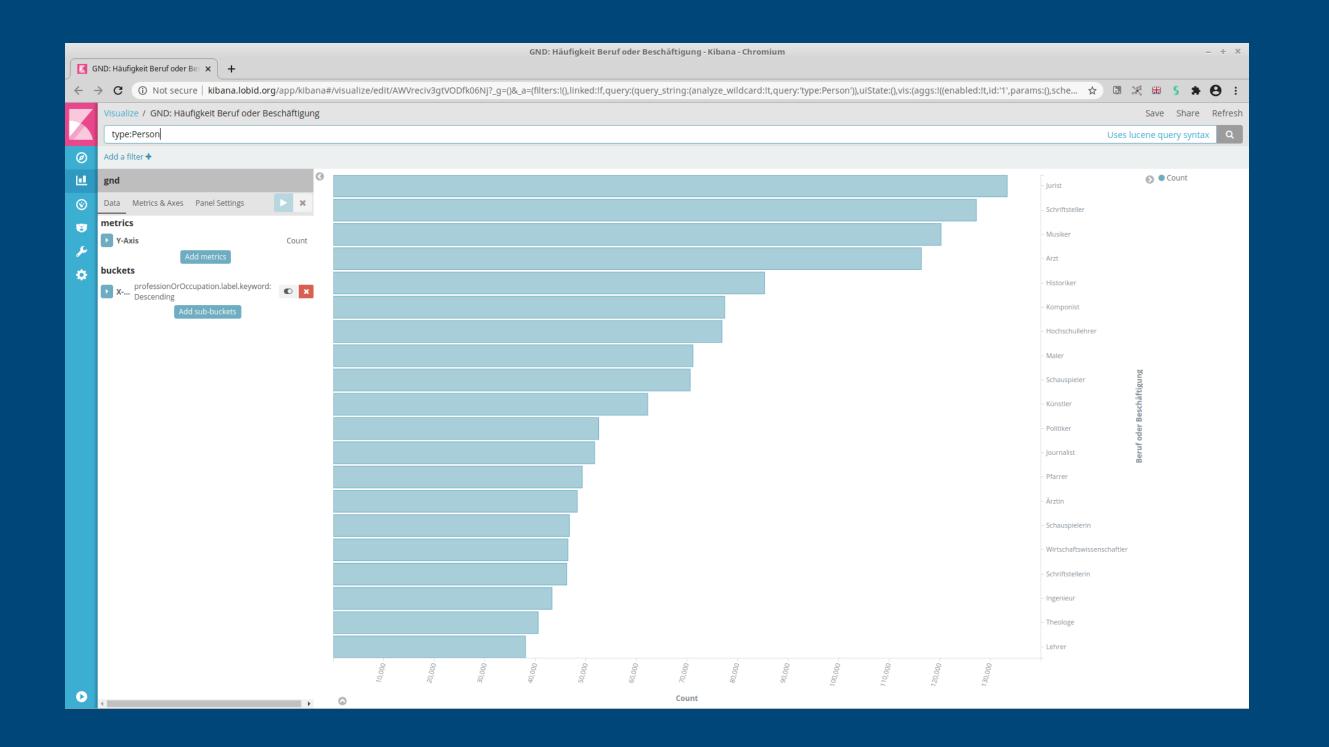
#### ZDB-Titel, die mit einer NWBib-Systematik erschlossen sind, aber nicht der NWBib zugerechnet werden und von der ULB Bonn erstellt wurden

```
subject.source.id:
    ("https://nwbib.de/subjects" OR
        "https://nwbib.de/spatial")
AND inCollection.id:
        "http://lobid.org/resources/HT014846970#!"
AND NOT inCollection.id:
        "http://lobid.org/resources/HT014176012#!"
AND describedBy.sourceOrganization.id:
        "http://lobid.org/organisations/DE-5#!"
```

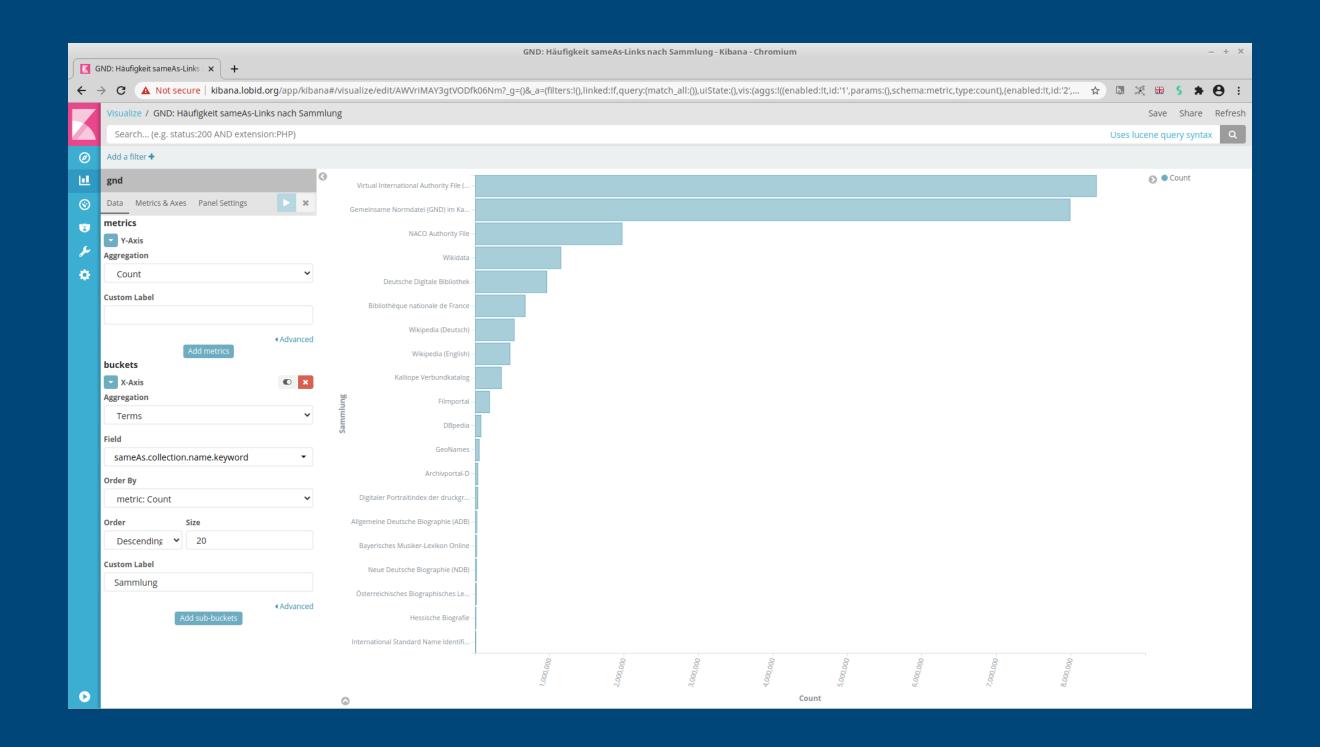
# Deletions Endpoint: Vom 14. bis 23.2.2019 gelöschte Titel

describedBy.deleted:[20190214 TO 2019023]

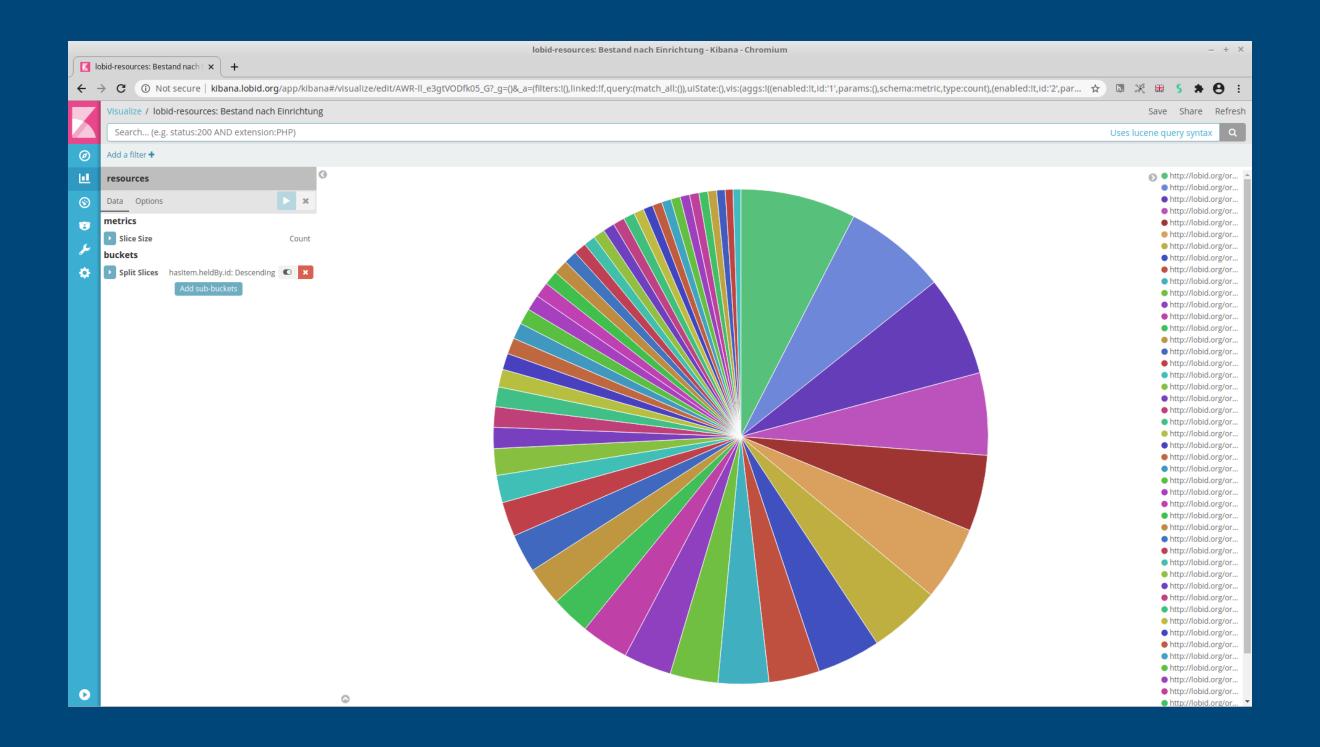
#### Kibana-Visualisierungen



Balkendiagramm: Häufigkeit von Berufsangaben in der GND



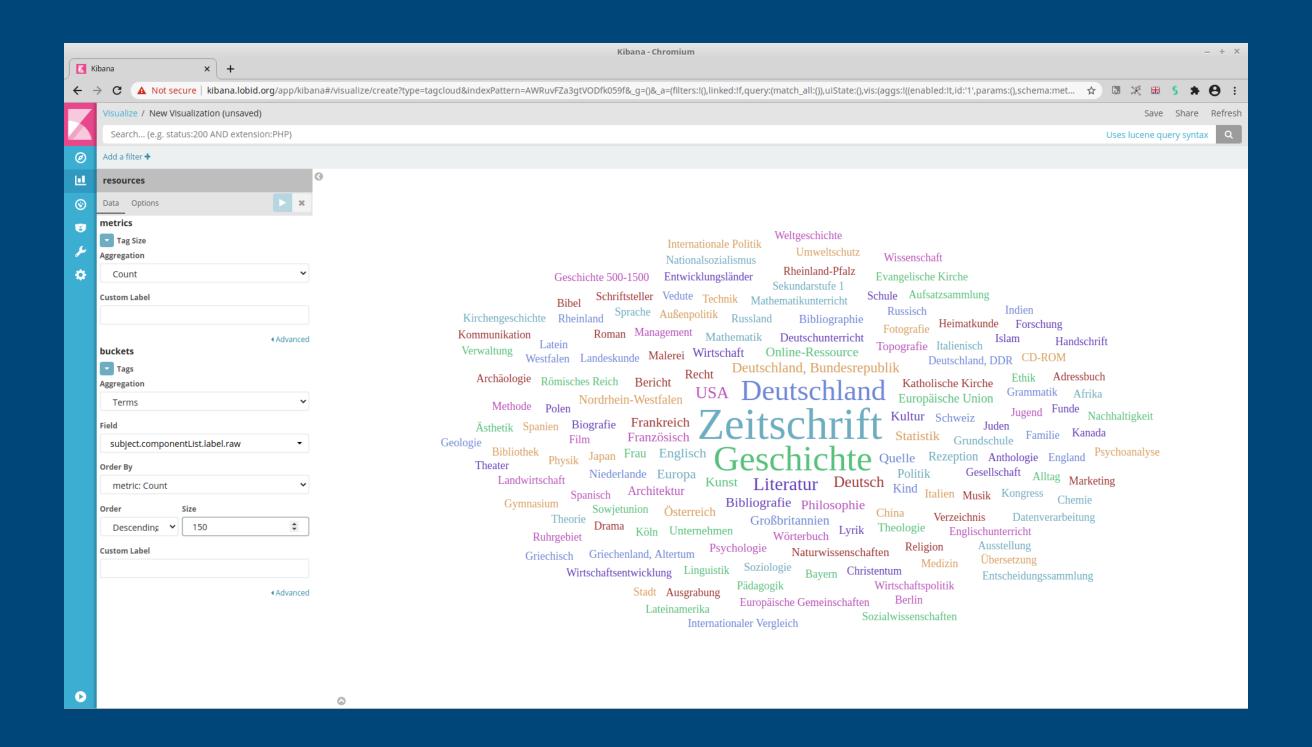
Balkendiagramm: Häufigkeit von sameAs-Links in lobid-GND



Kuchendiagramm: Bestand nach Einrichtung in lobidresources



Balkendiagramm: Publikationsdatum in lobid-resources>



Tagcloud: Top 150 GND-Schlagwörter in lobid-resources

# Teil 3: OpenRefine Reconciliation

Grundlagen	Grundlagen von OpenRefine und Reconciliation	14:00–14:30
Matching	Verbesserung des Matching auf GND-Einträge	14:30–15:00
Extension	Datenanreicherung auf Basis der Matches	15:00–15:30

F

"A powerful tool for working with messy data"

"A powerful tool for working with messy data"

"cleaning it; transforming it from one format into another; and extending it with web services and external data"

"A powerful tool for working with messy data"

"cleaning it; transforming it from one format into another; and extending it with web services and external data"

Oberfläche wie Tabellenkalkulation

"A powerful tool for working with messy data"

"cleaning it; transforming it from one format into another; and extending it with web services and external data"

Oberfläche wie Tabellenkalkulation

Läuft im Browser

Lokal, downloaden und installieren:

https://openrefine.org/download.html

Lokal, downloaden und installieren: https://openrefine.org/download.html

Gehostet, z.B. zum Ausprobieren: http://openrefine.labs.lobid.org

Gut geeigent wo UI/Tabellenkalkulation passt

Gut geeigent wo UI/Tabellenkalkulation passt

Fachliche Arbeit

Gut geeigent wo UI/Tabellenkalkulation passt

Fachliche Arbeit

Manuelle Optimierung

Gut geeigent wo UI/Tabellenkalkulation passt

Fachliche Arbeit

Manuelle Optimierung

Überschaubare Datenmengen

F

# Wofür ist OpenRefine (noch) nicht?

# Wofür ist OpenRefine (noch) nicht?

Bisher nicht geeignet für große Dateien, aber eine neue Architektur für größere Datenmengen ist fast fertig

# Wofür ist OpenRefine (noch) nicht?

Bisher nicht geeignet für große Dateien, aber eine neue Architektur für größere Datenmengen ist fast fertig

Automatisierung ohne UI nicht mit Bordmitteln, aber mit Zusatztool möglich: openrefine-batch

F

Abgleichen / Matchen eigener Daten (z.B. Namen von Personen) auf GND-Einträge

Abgleichen / Matchen eigener Daten (z.B. Namen von Personen) auf GND-Einträge

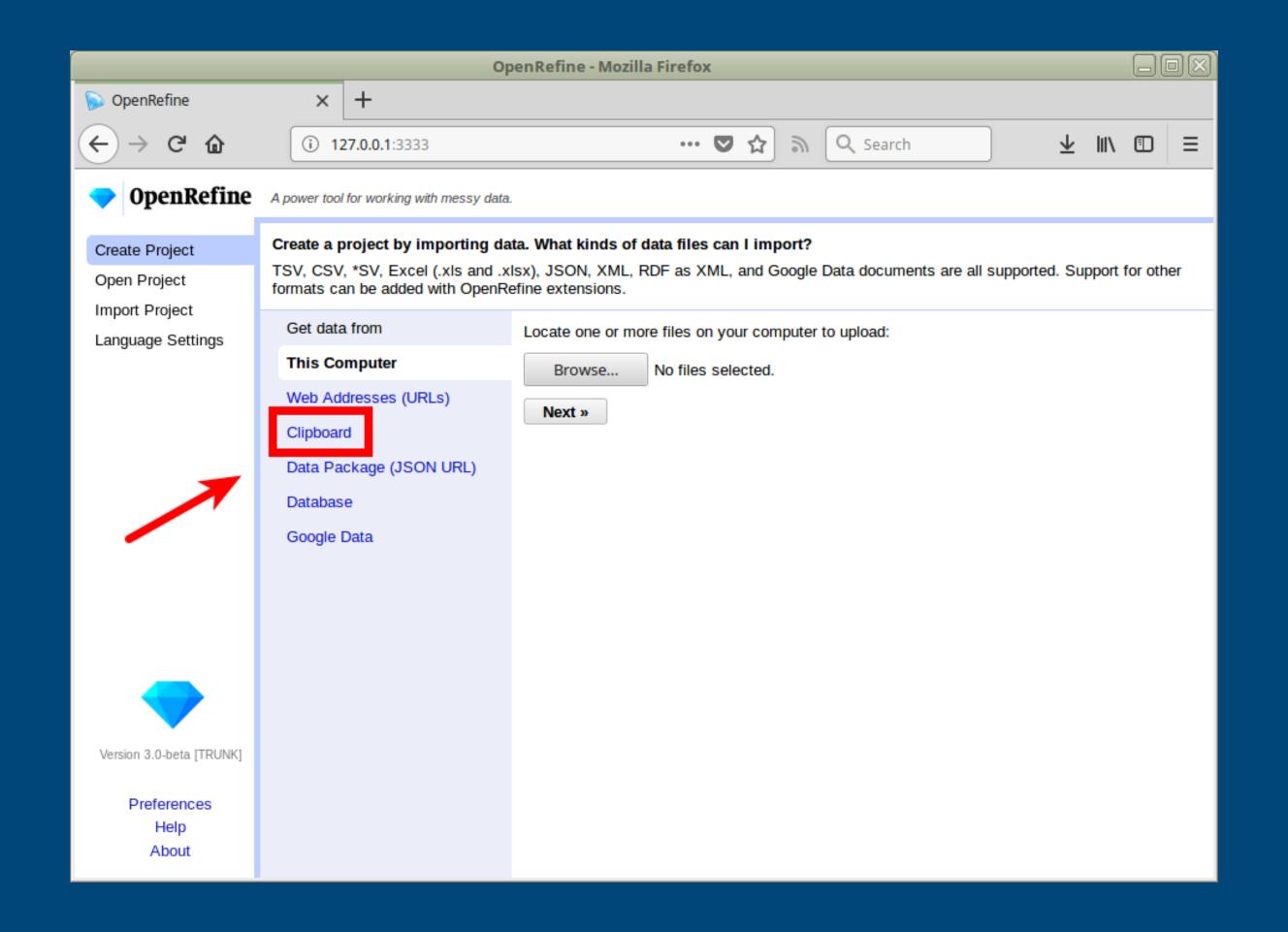
Anreicherung durch Übernahme von spezifischen Feldern der gematchten GND-Einträge

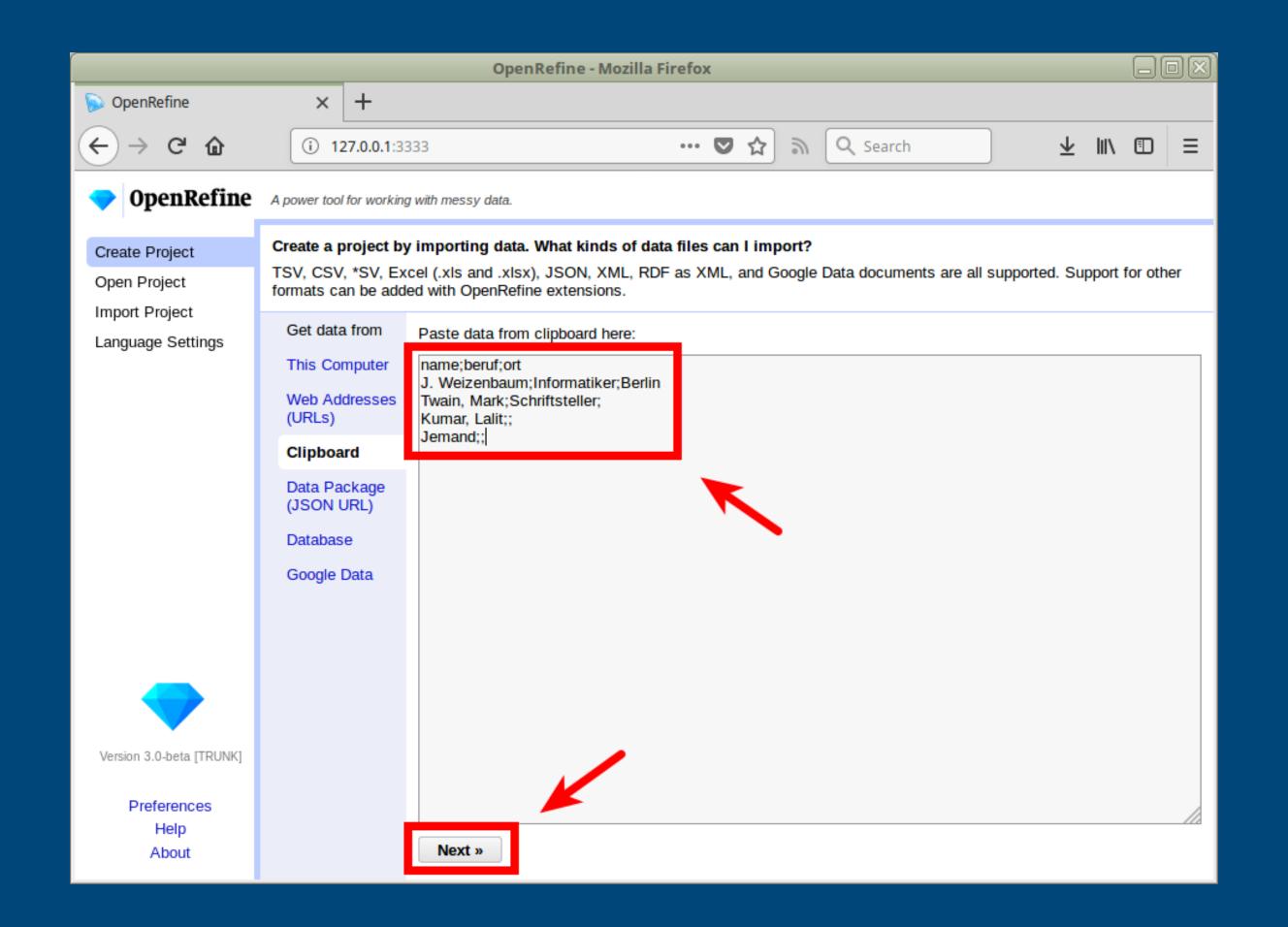
Abgleichen / Matchen eigener Daten (z.B. Namen von Personen) auf GND-Einträge

Anreicherung durch Übernahme von spezifischen Feldern der gematchten GND-Einträge

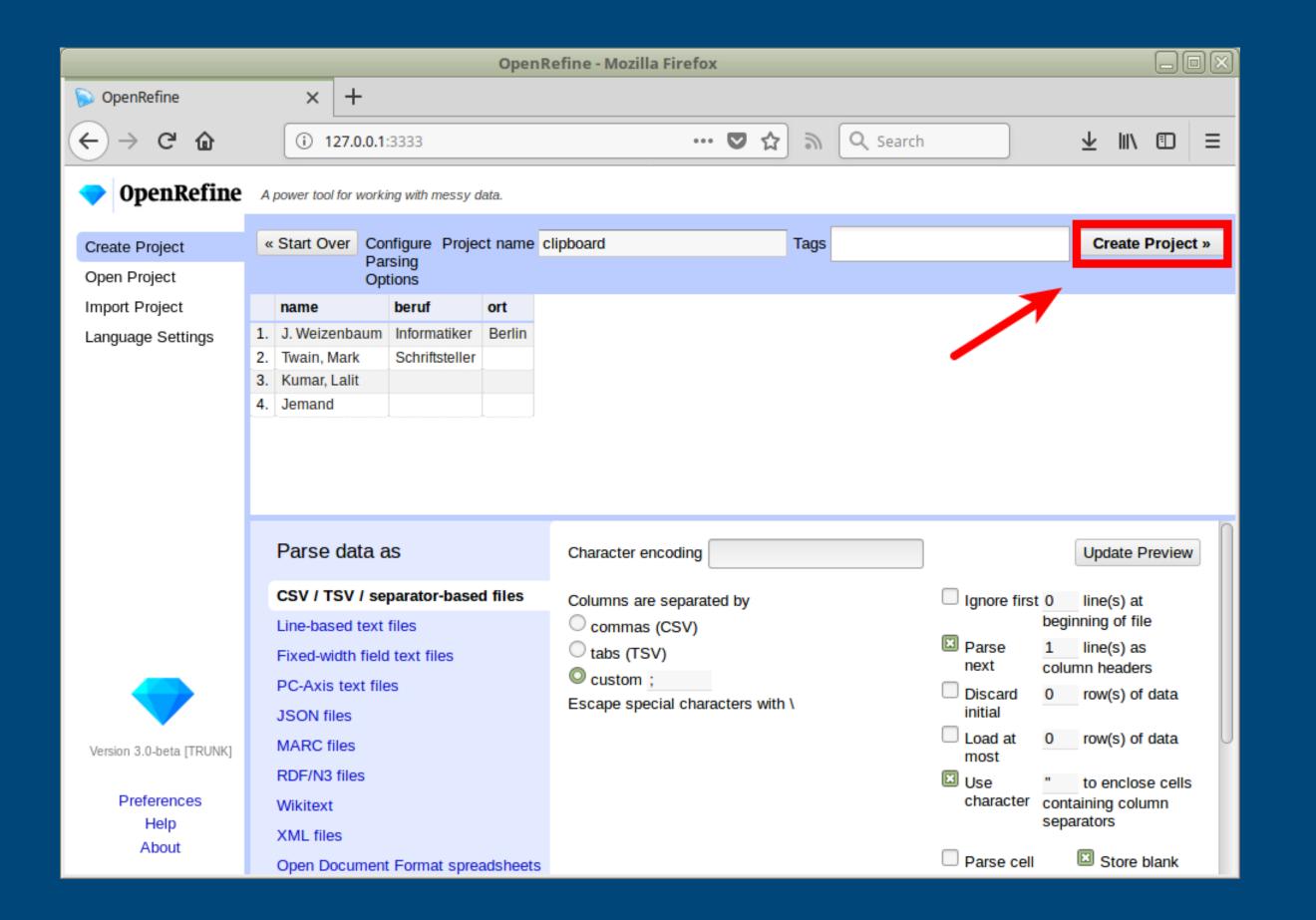
Fokus hier ist Reconciliation, auch Grundlagen von OpenRefine, aber es gibt noch viel mehr

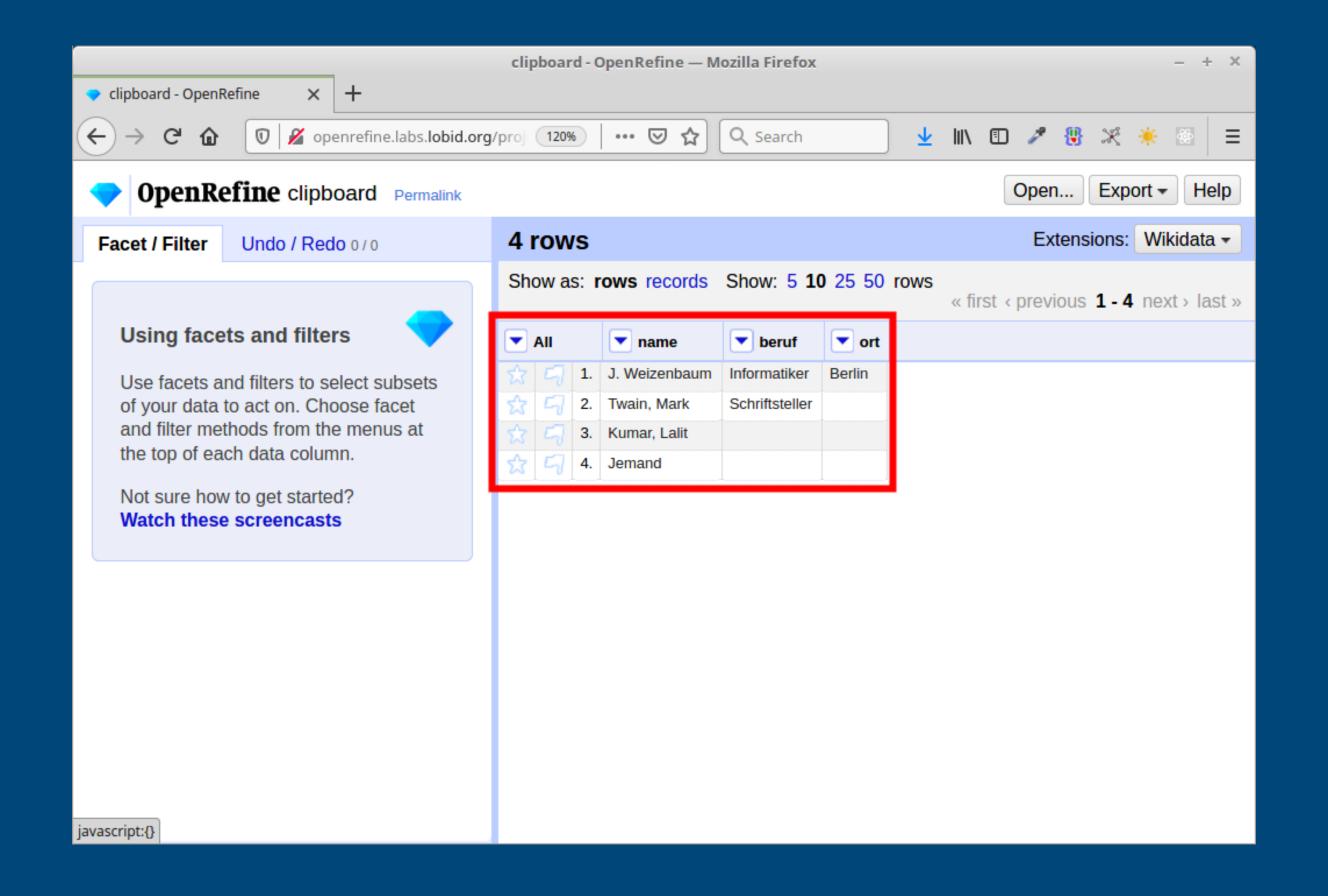
### Projekt anlegen





 $\blacksquare$ 





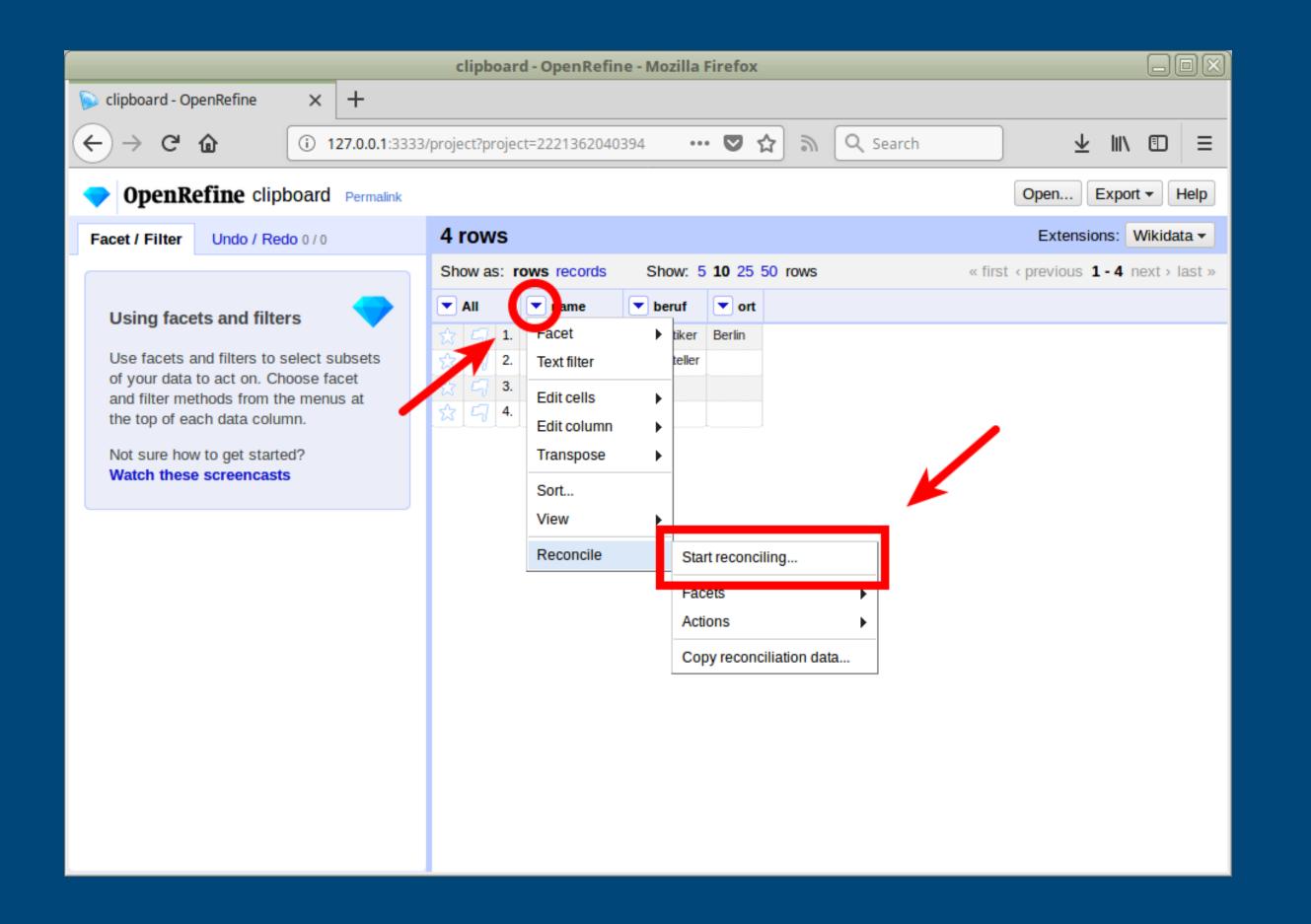
### Projekt anlegen

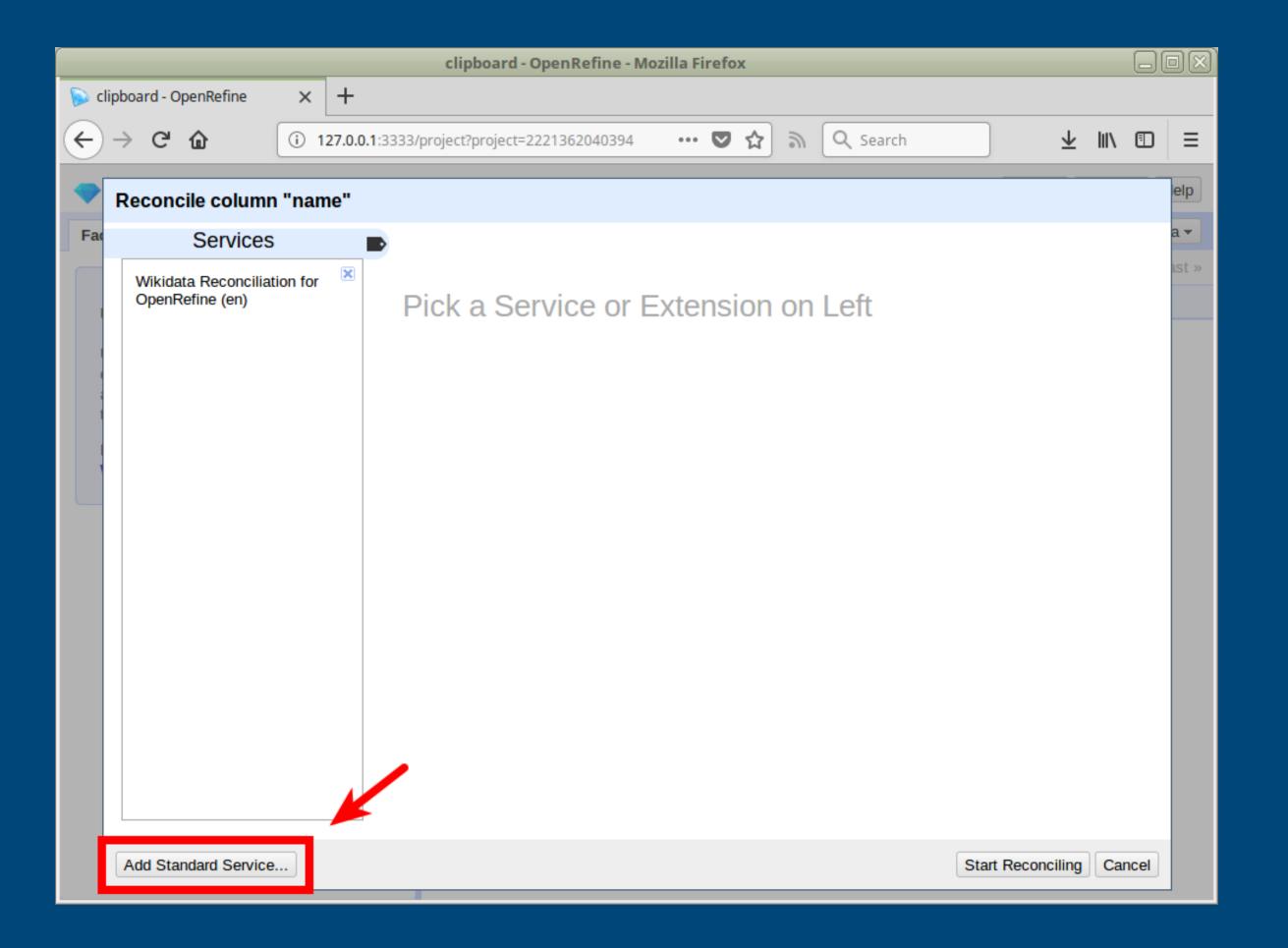
openrefine.labs.lobid.org | localhost:3333

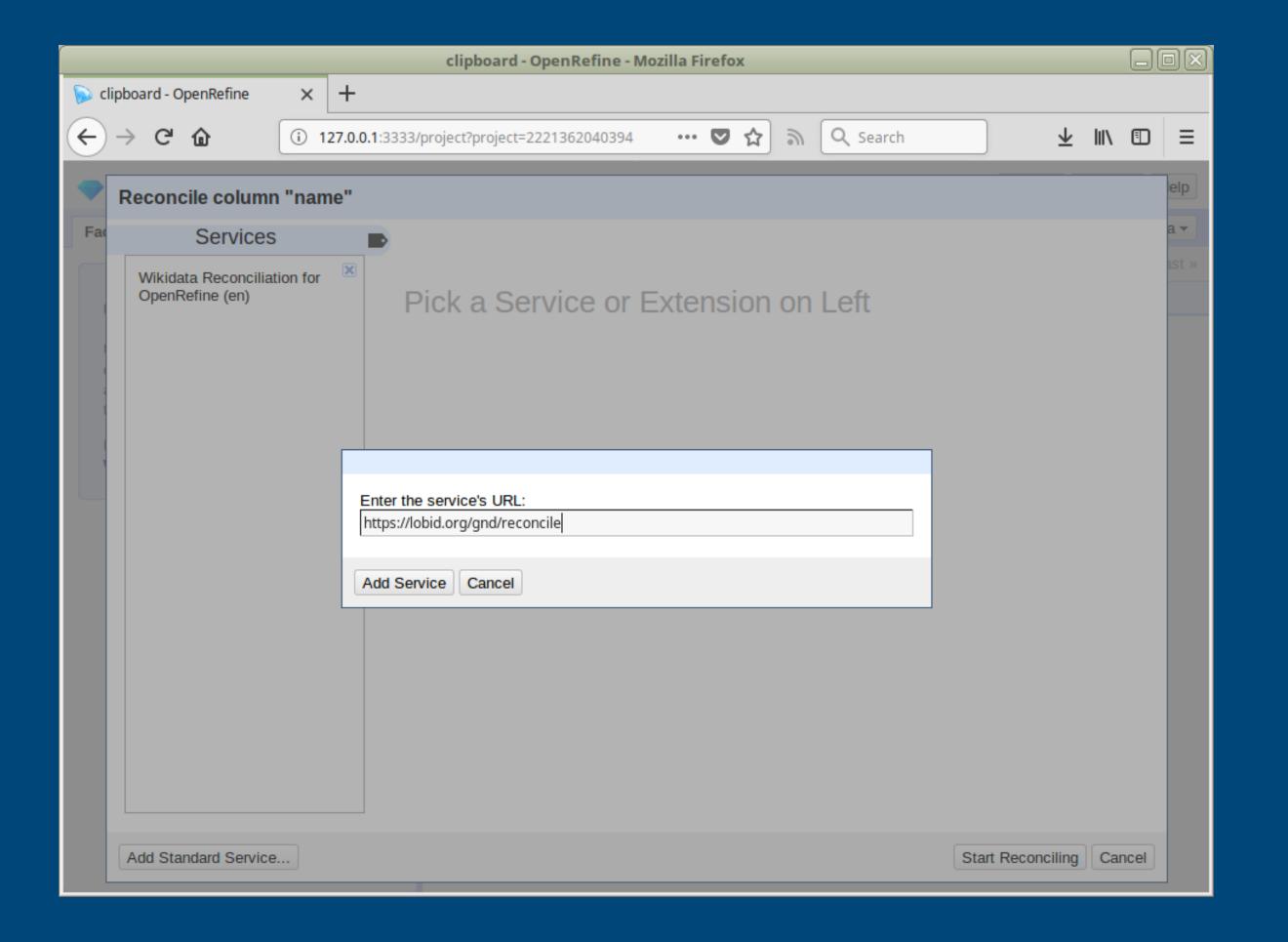
```
name;beruf;ort
J. Weizenbaum;Informatiker;Berlin
Twain, Mark;Schriftsteller;
Kumar, Lalit;;
Jemand;;
```

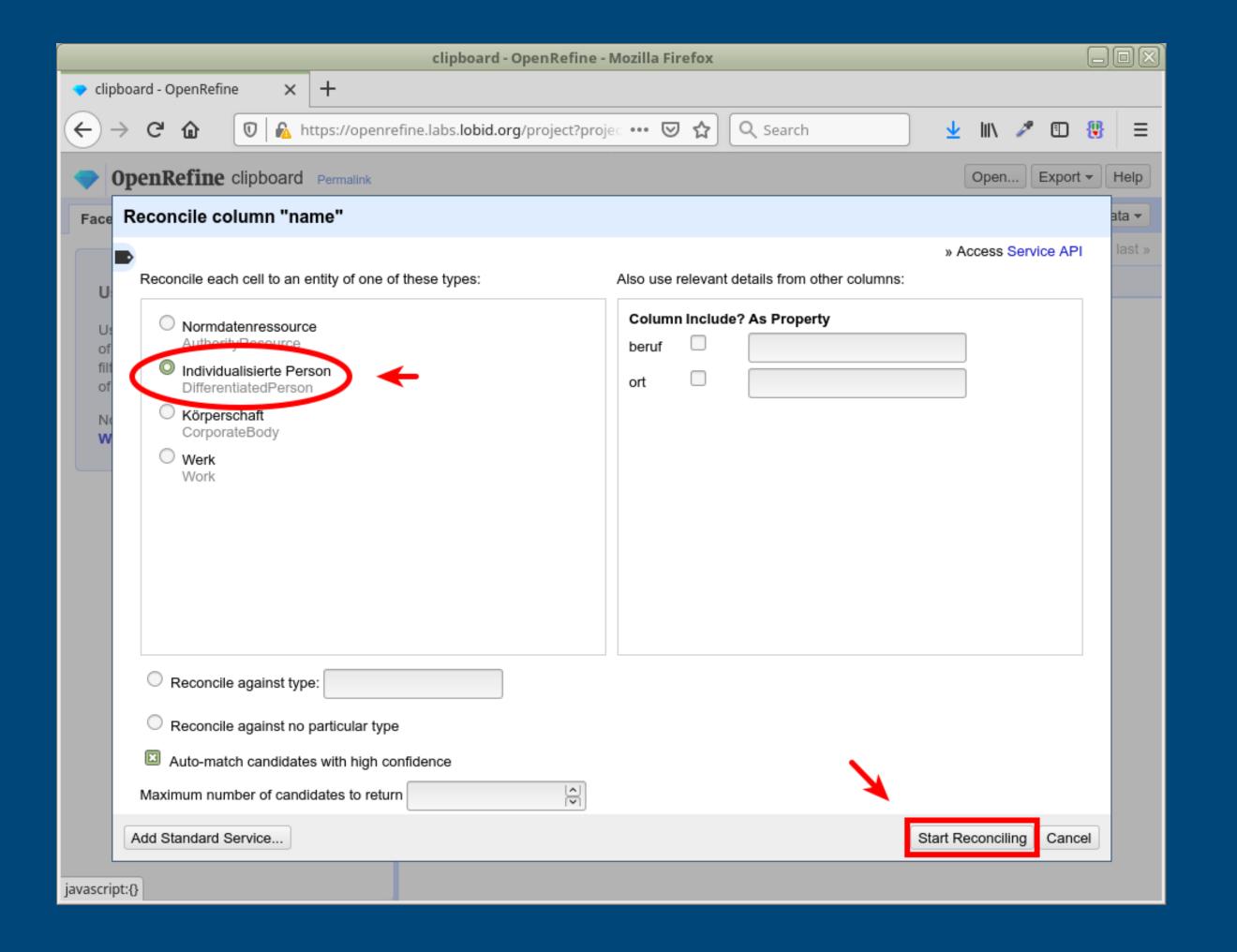
(Daten und URL per CTRL+C in die Zwischenablage kopieren)

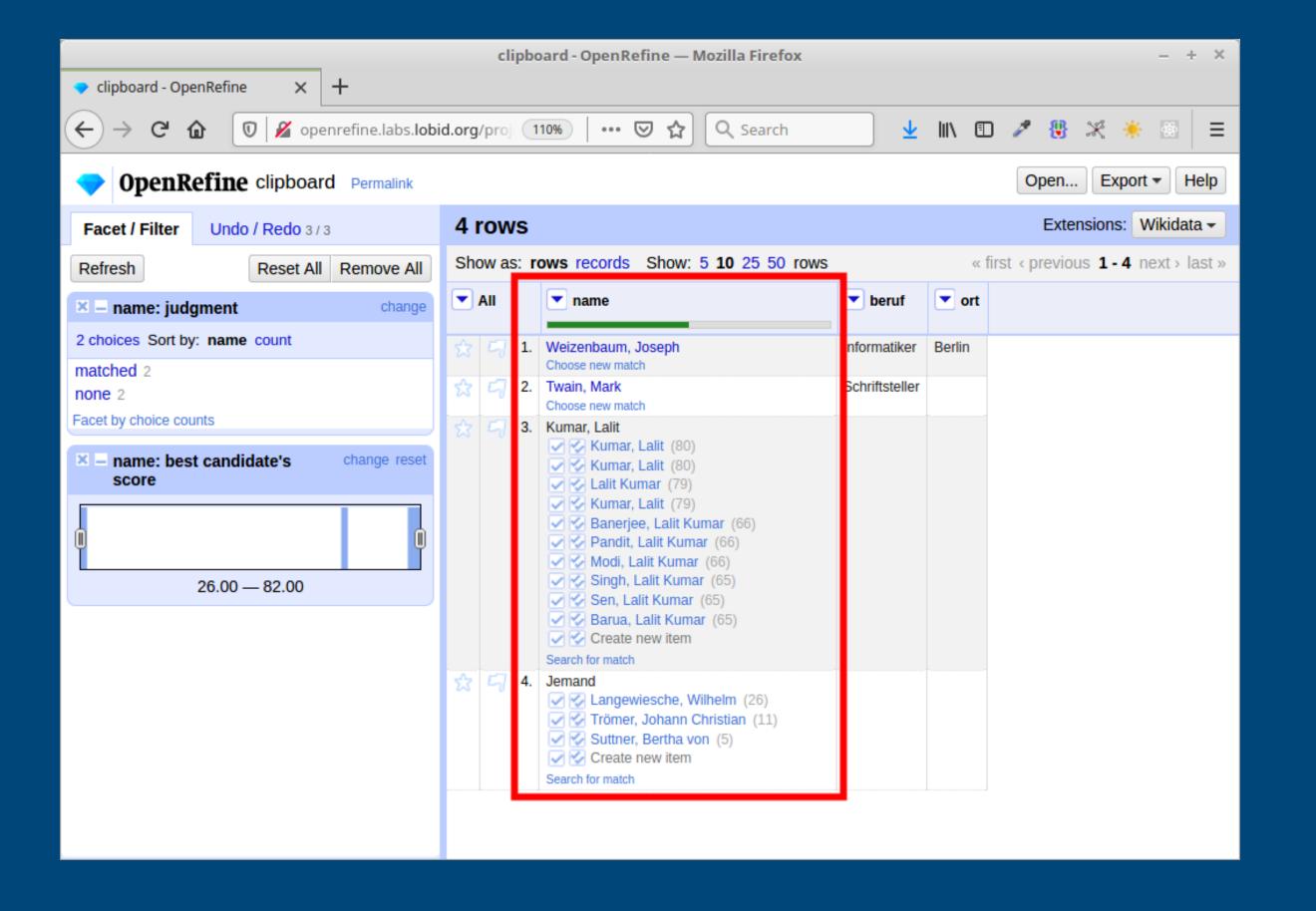
# Abgleich / Reconciliation

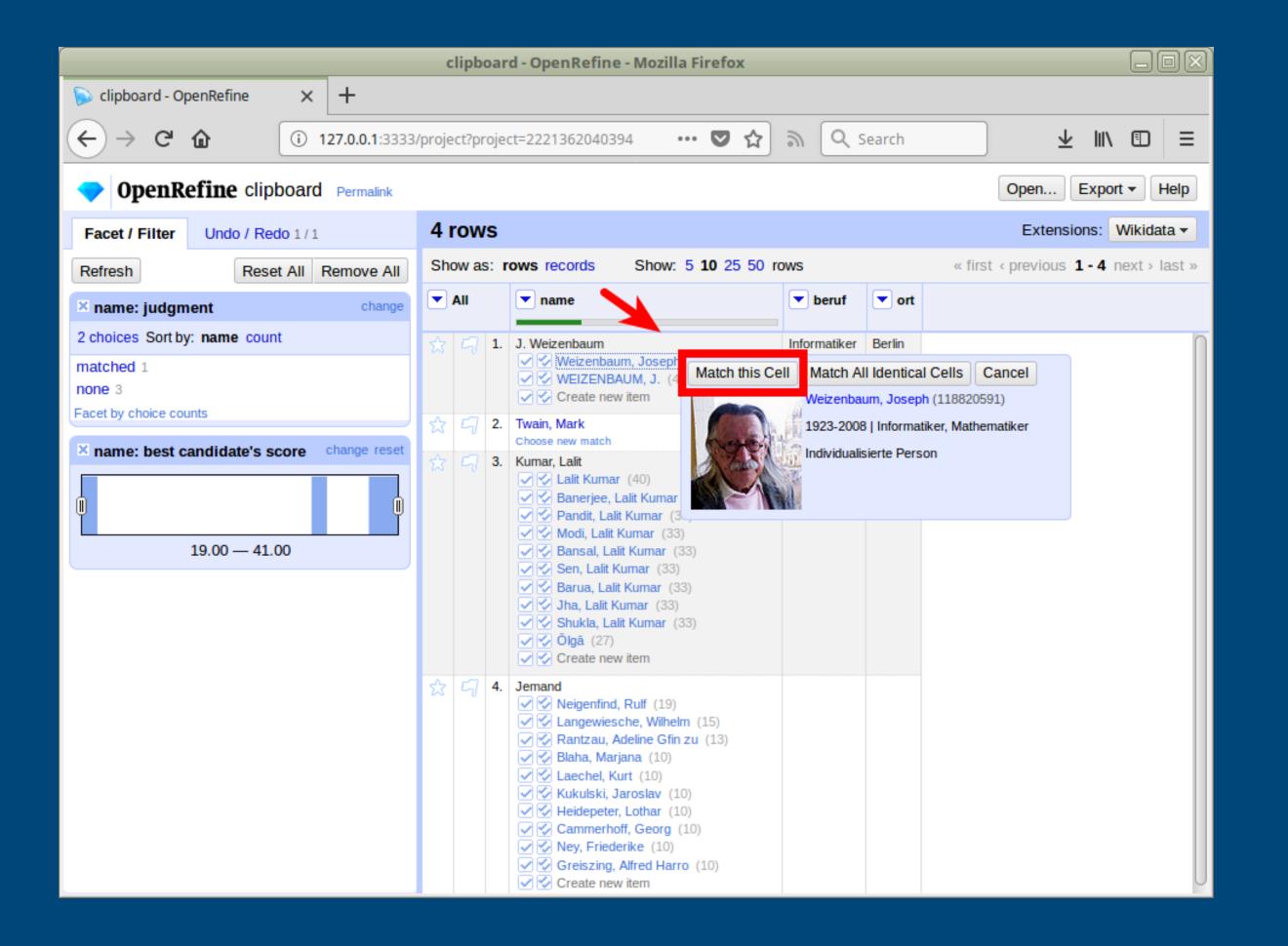












# Abgleich / Reconciliation

openrefine.labs.lobid.org | localhost:3333

https://lobid.org/gnd/reconcile

(Auf openrefine.labs.lobid.org muss der Dienst nicht ergänzt werden)

Namen wurden mit GND-IDs assoziiert

Namen wurden mit GND-IDs assoziiert

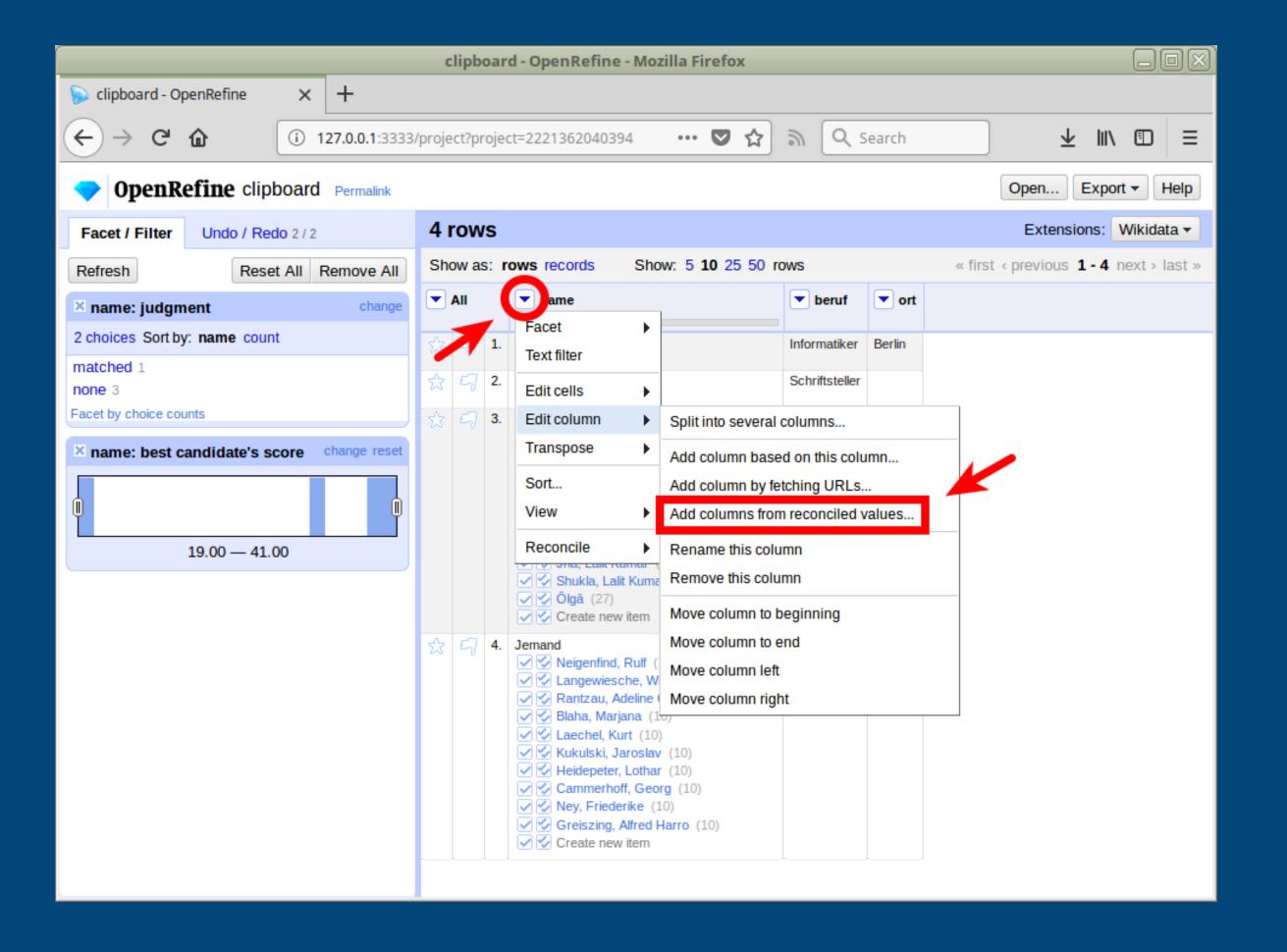
Felder der GND-Entitäten stehen so zur Verfügung

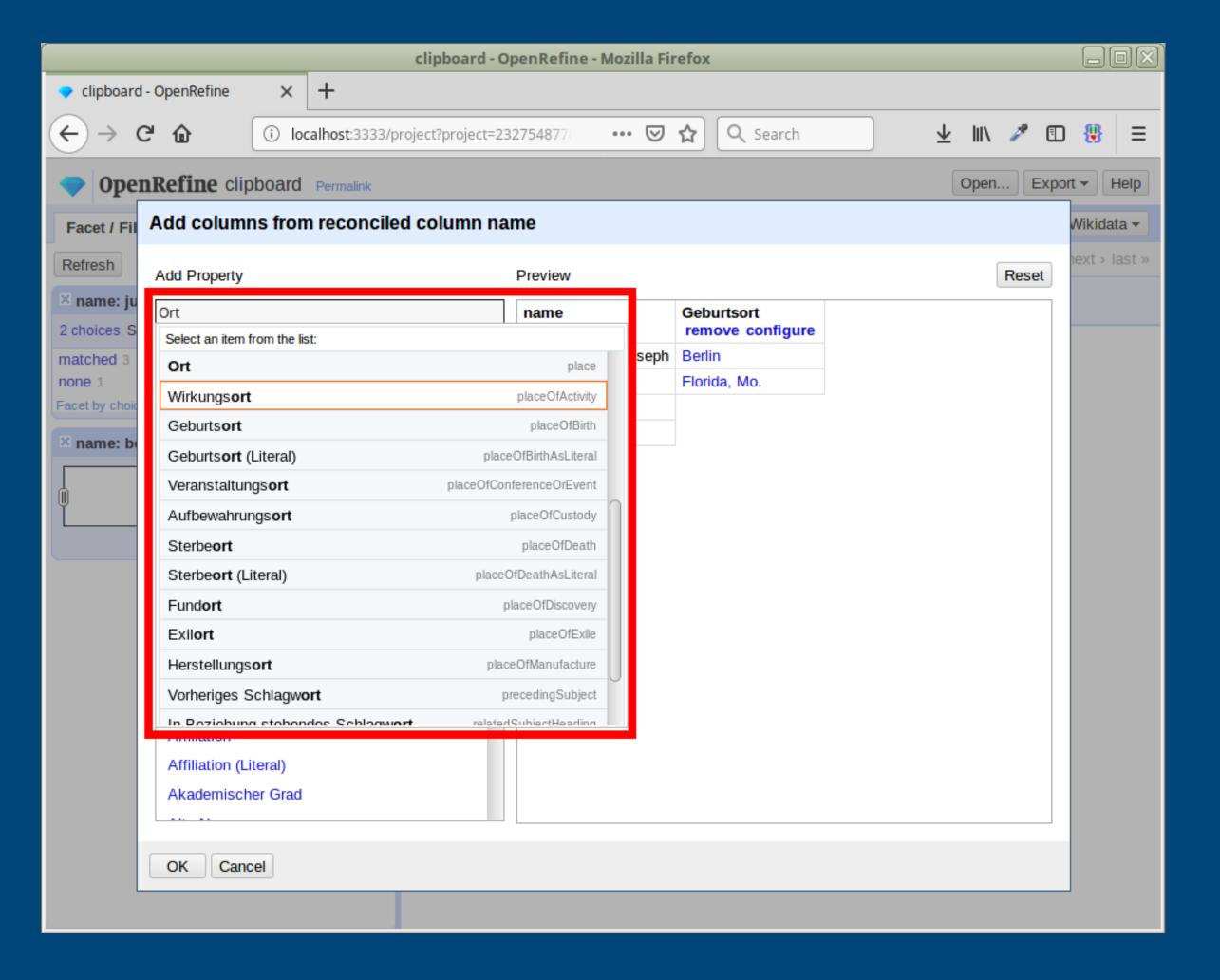
Namen wurden mit GND-IDs assoziiert

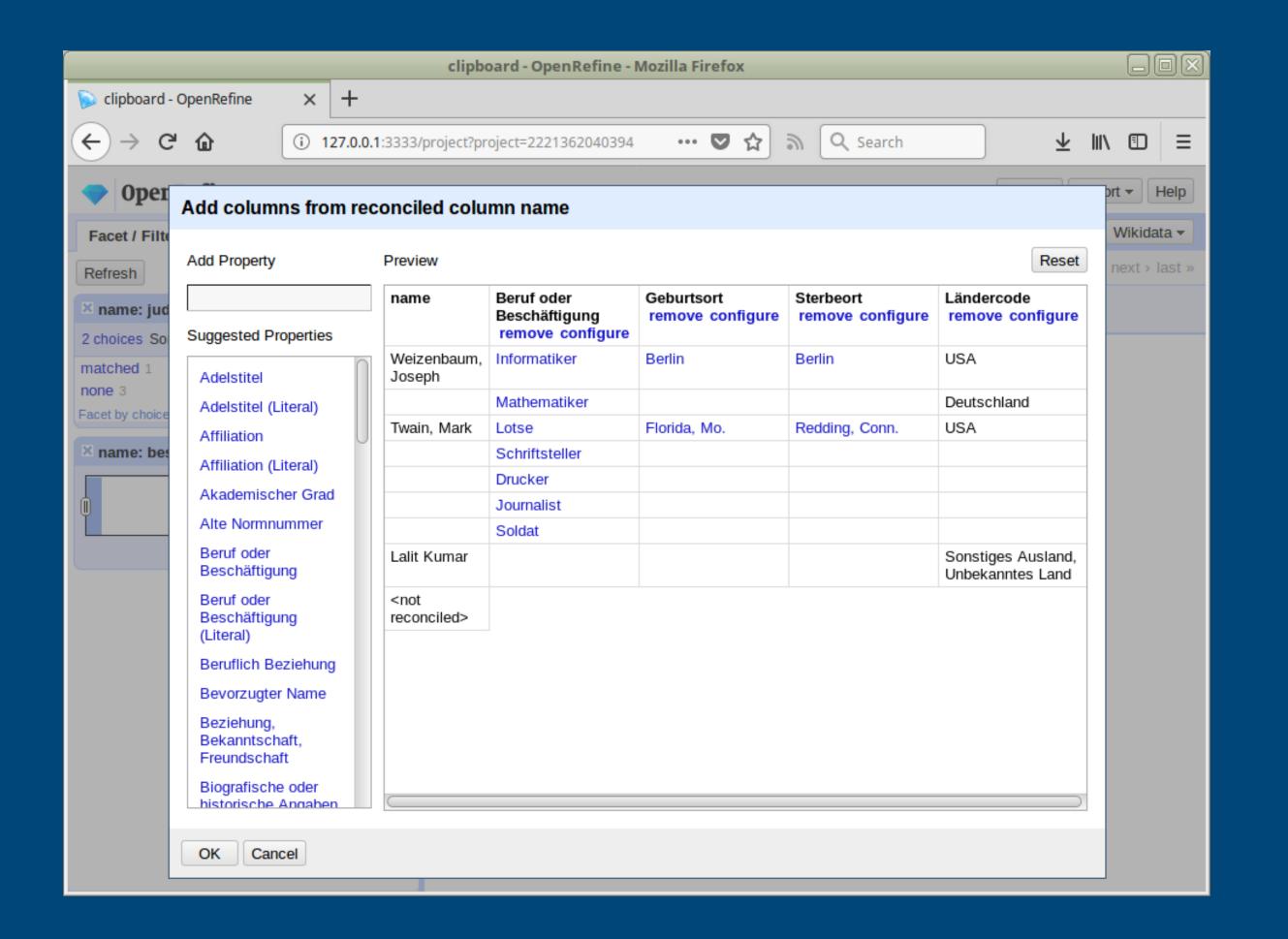
Felder der GND-Entitäten stehen so zur Verfügung

Mit diesen können wir unsere lokalen Daten anreichern

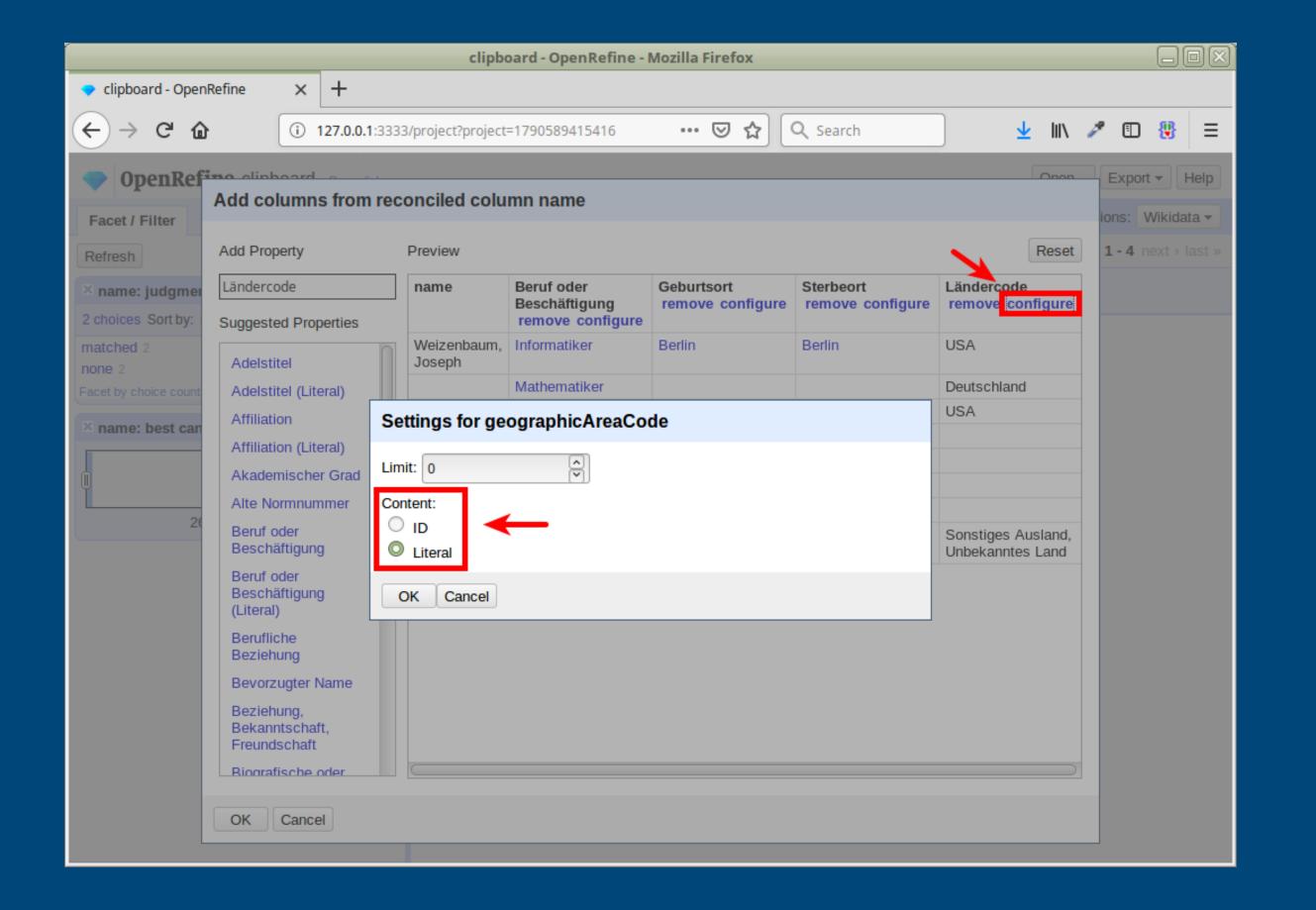
Namen wurden mit GND-IDs assoziiert
Felder der GND-Entitäten stehen so zur Verfügung
Mit diesen können wir unsere lokalen Daten anreichern
d.h. konkret: neue Spalten in unserer Tabelle anlegen







F 141



F 142

```
id, name
1, Haddaway
2, Johannes Geßner
3, Judith Kuckart
4, Rosa Luxemburg
5, Albert Einstein
6, Hazel Brugger
7, Anne Cuneo
8, Vitus Huonder
```

Projekt anlegen, Namen abgleichen und Berufe anreichern

U

Angereicherte Felder sind z.T. GND-Entitäten

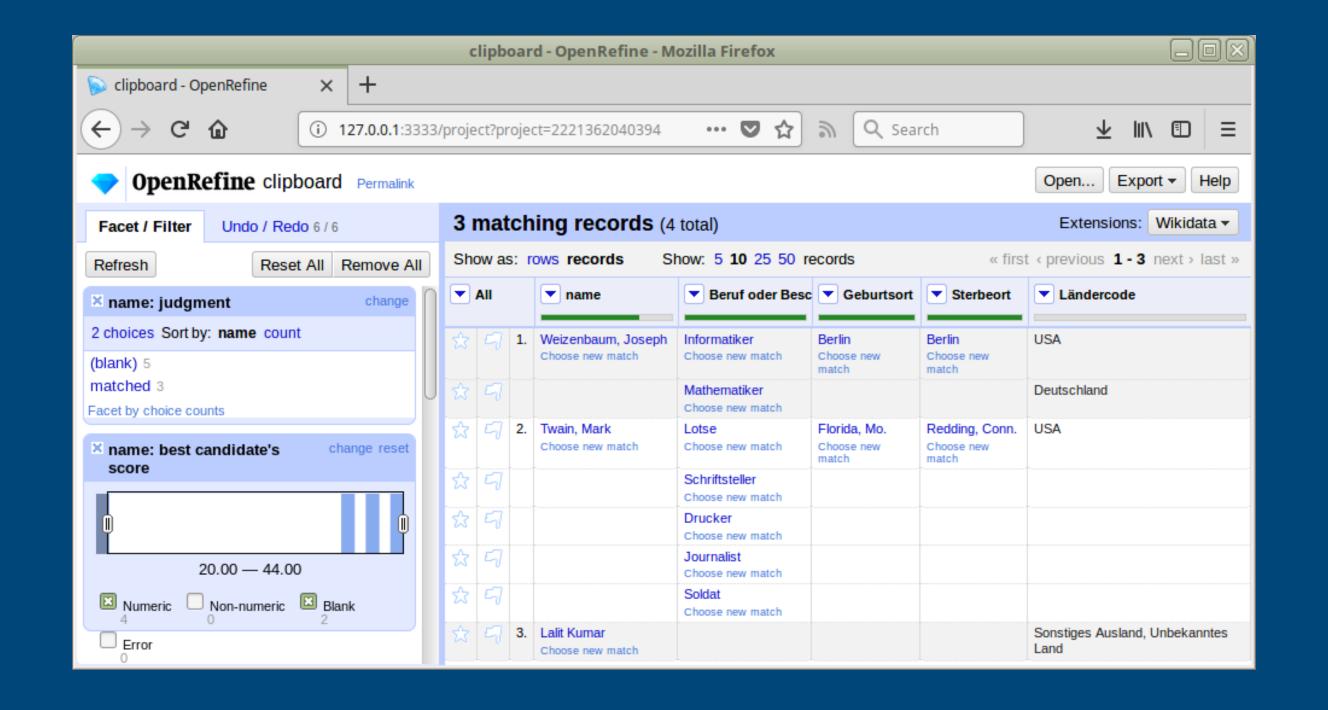
Angereicherte Felder sind z.T. GND-Entitäten

Sind automatisch abgeglichen / reconciled

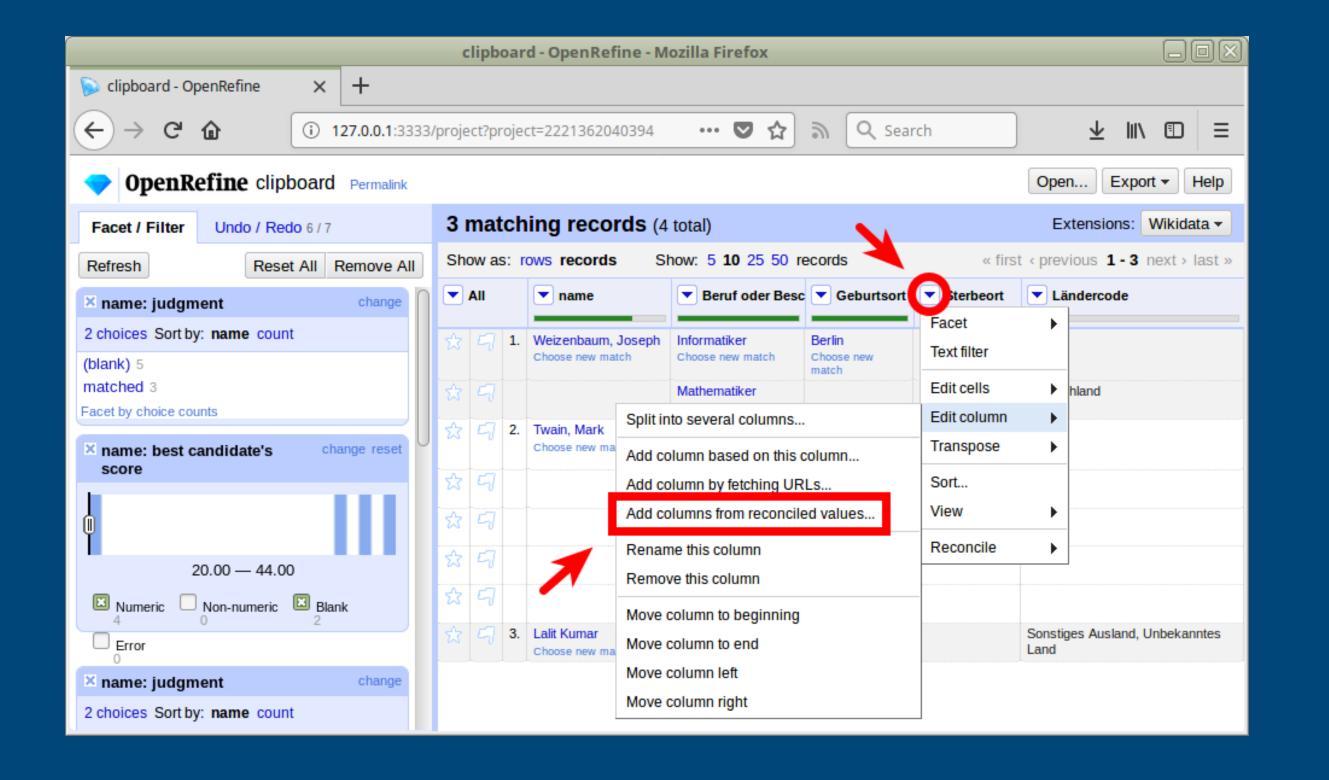
Angereicherte Felder sind z.T. GND-Entitäten

Sind automatisch abgeglichen / reconciled

Können als Grundlage für weitere Anreicherung dienen



F 145



Wirkungsort und Darstellung des Wirkungsortes anreichern

Export \* Open... Help 8 records Extensions: Wikidata -Show as: rows records Show: 5 10 25 50 records « first < previous 1 - 8 next > last » ▼ AII ▼ Beruf oder Beschäl ▼ Wirkungsort ▼ Darstellung des Wirkungsortes ▼ id ▼ name 1. 1 Haddaway Komponist Köln http://commons.wikimedia.org/wiki/Special:FilePath /DEU%20Koeln%20COA.svg Choose new Choose new match Choose new match match Sänger Kitzbühel http://commons.wikimedia.org/wiki/Special:FilePath /AUT%20Kitzb%C3%BChel%20COA.svg Choose new match Choose new match Musiker Choose new match 2. 2 Geßner, Zürich http://commons.wikimedia.org/wiki/Special:FilePath Arzt /Zurich-coat%20of%20arms.svg Johannes Choose new match Choose new match Choose new match Mathematiker Choose new match Naturwissenschaftler Choose new match 3. 3 Kuckart, Judith Schriftstellerin Berlin https://commons.wikimedia.org/wiki/Special:FilePath /Cityscape\_Berlin.jpg Choose new Choose new match Choose new match match Regisseurin http://commons.wikimedia.org/wiki/Special:FilePath Zürich /Zurich-coat%20of%20arms.svq Choose new match Choose new match Tänzerin Choose new match Choreografin Choose new match

### Verbesserung des Matching auf GND-Einträge

# Ergebnisse verbessen

### Ergebnisse verbessern

Einschränkung der Typen der Zielentitäten

# Ergebnisse verbessen

Einschränkung der Typen der Zielentitäten

Verwendung zusätzlicher lokaler Daten

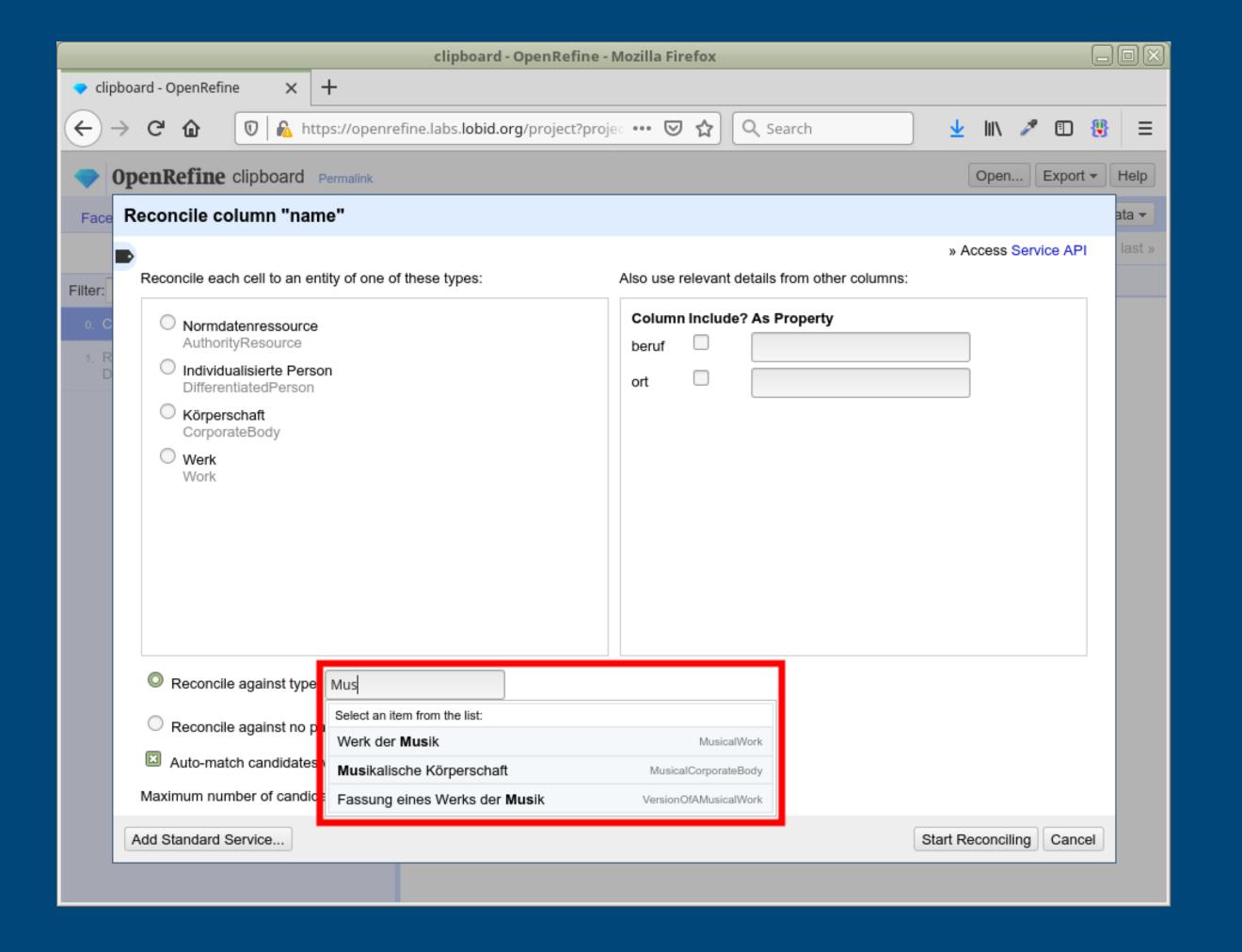
# Ergebnisse verbessen

Einschränkung der Typen der Zielentitäten

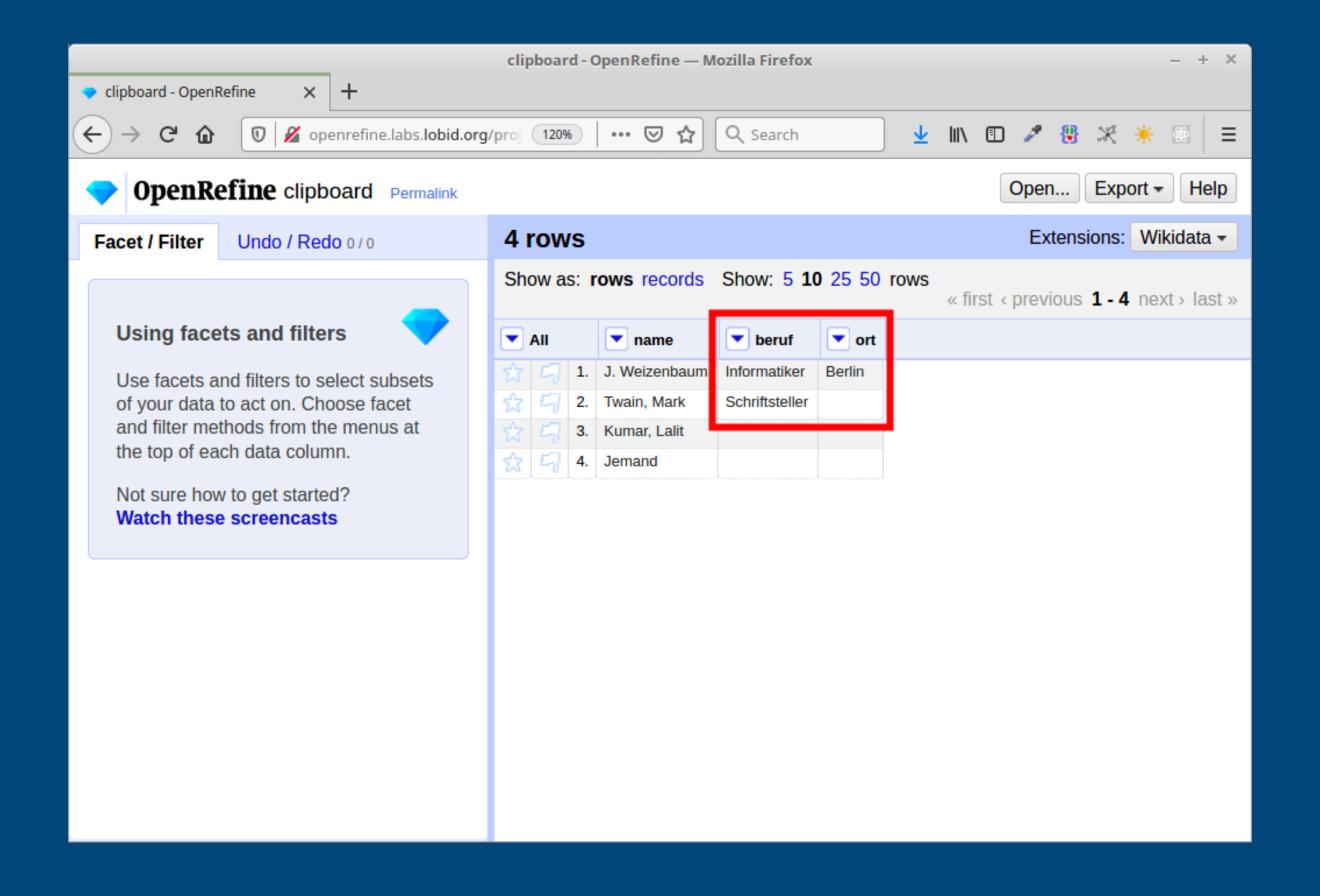
Verwendung zusätzlicher lokaler Daten

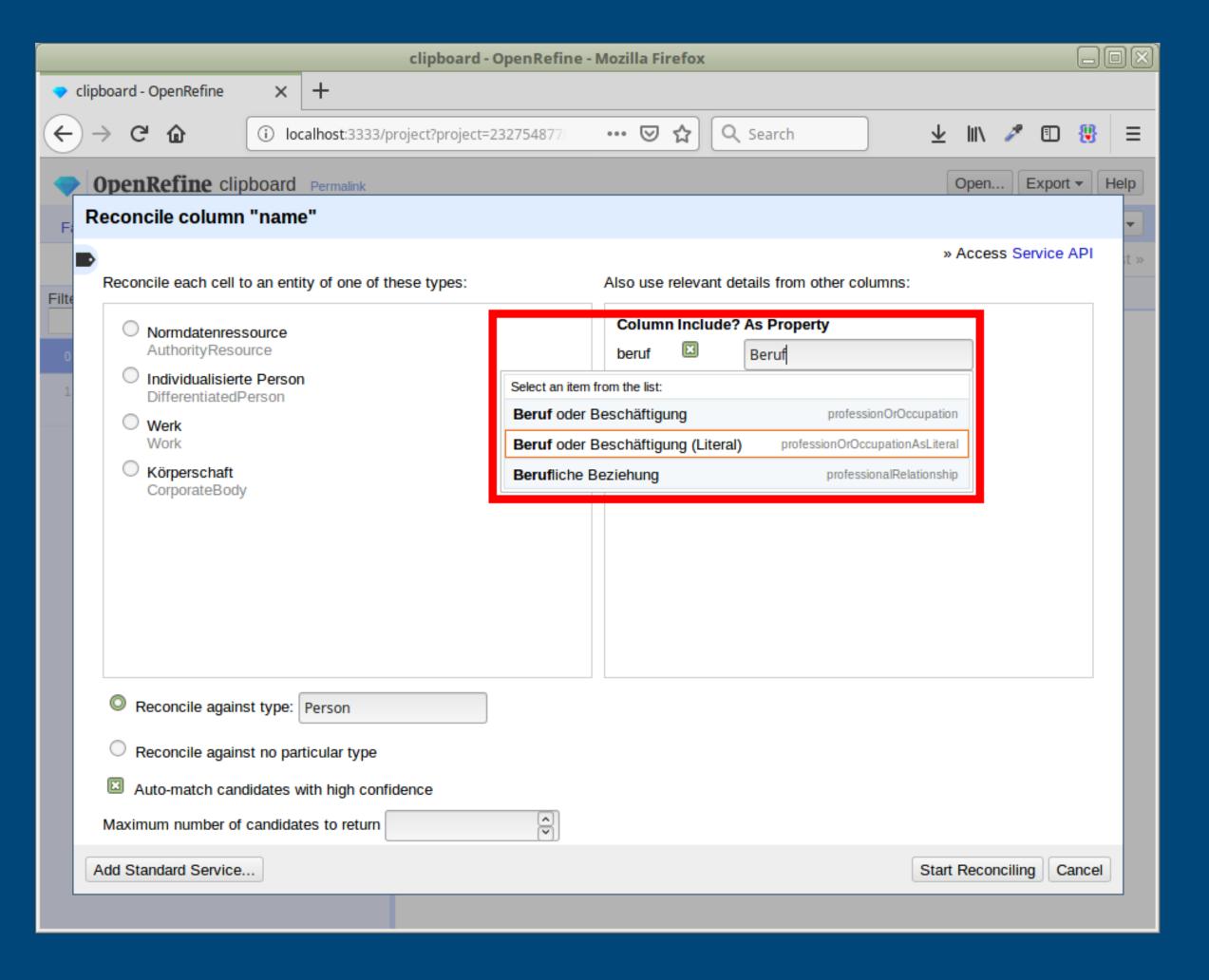
Gezielte Suche nach Treffern

#### Einschränkung der Entitätstypen

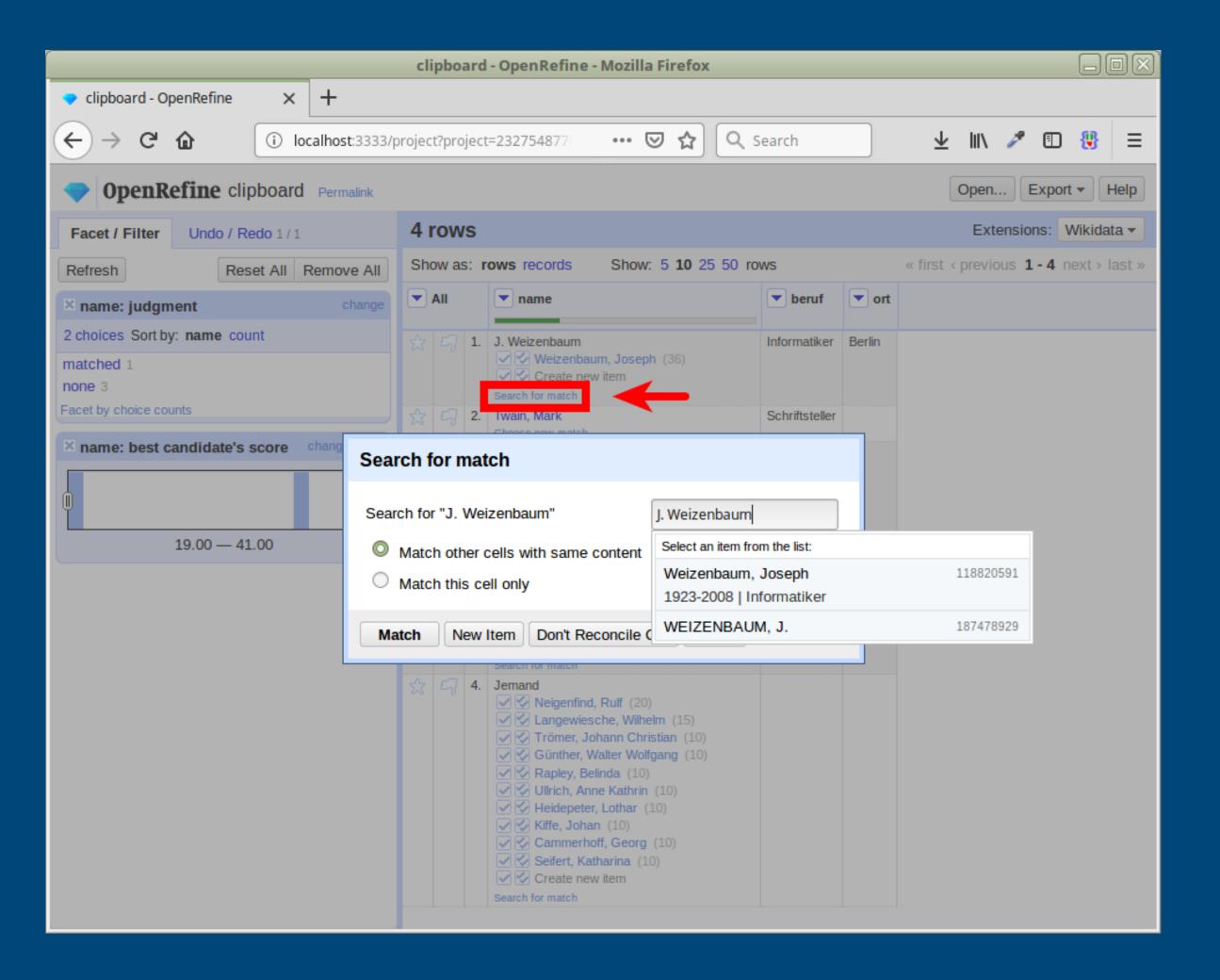


#### Verwendung zusätzlicher lokaler Daten





#### Gezielte Suche nach Treffern



Verwendung zusätzlicher lokaler Daten

Verwendung zusätzlicher lokaler Daten

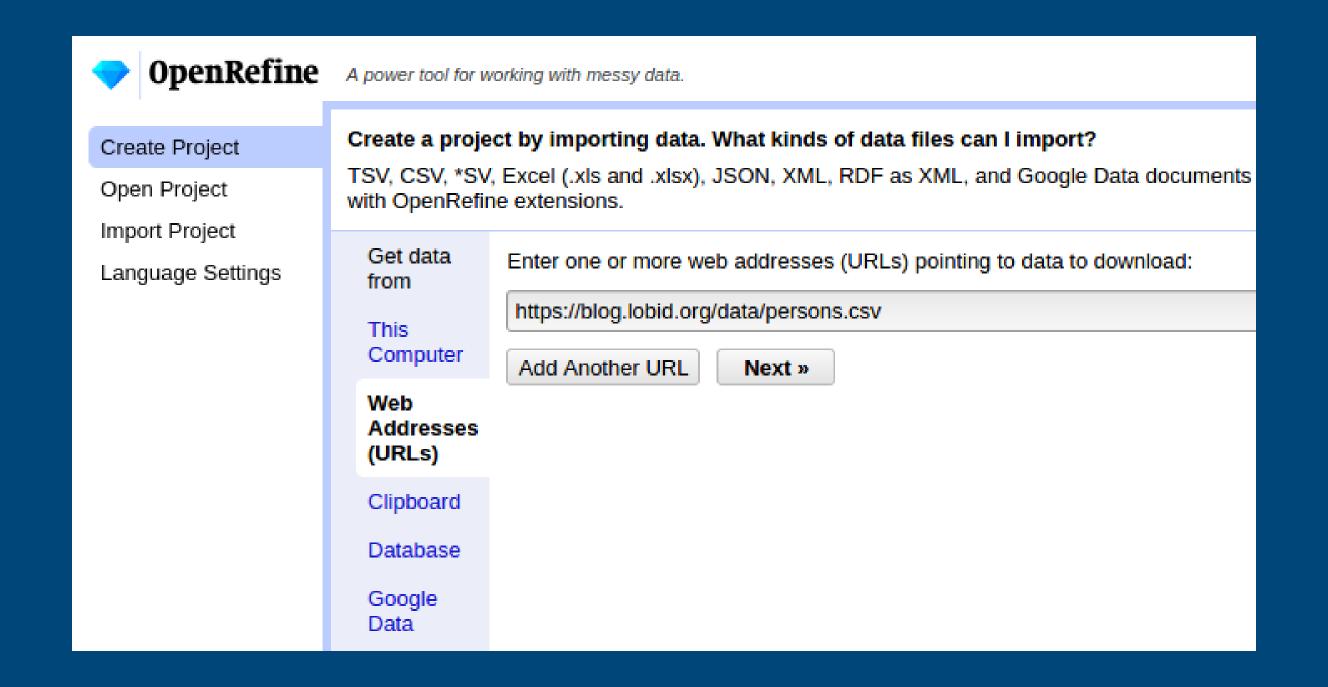
1. Beispiel: Lebensdaten

Verwendung zusätzlicher lokaler Daten

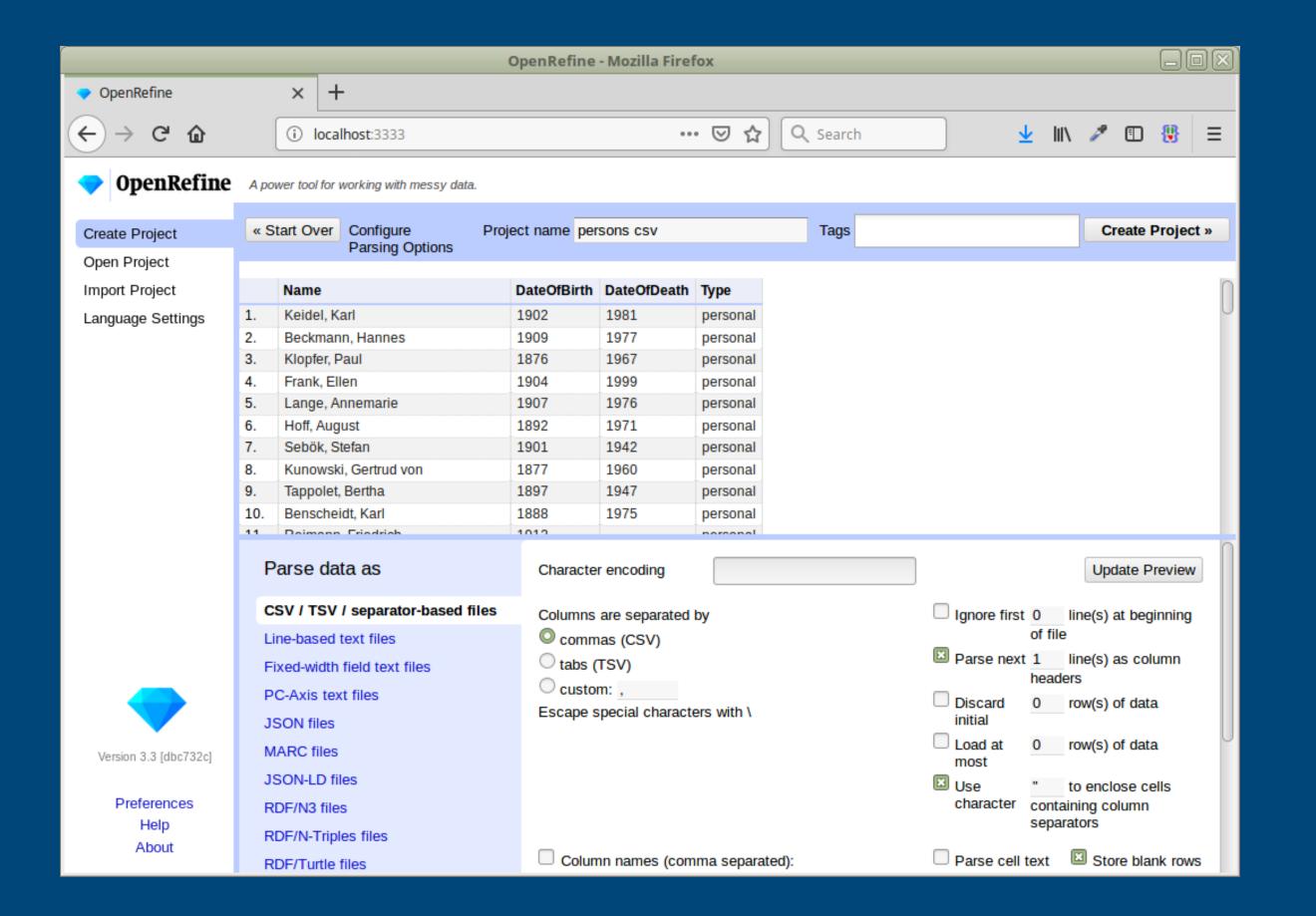
1. Beispiel: Lebensdaten

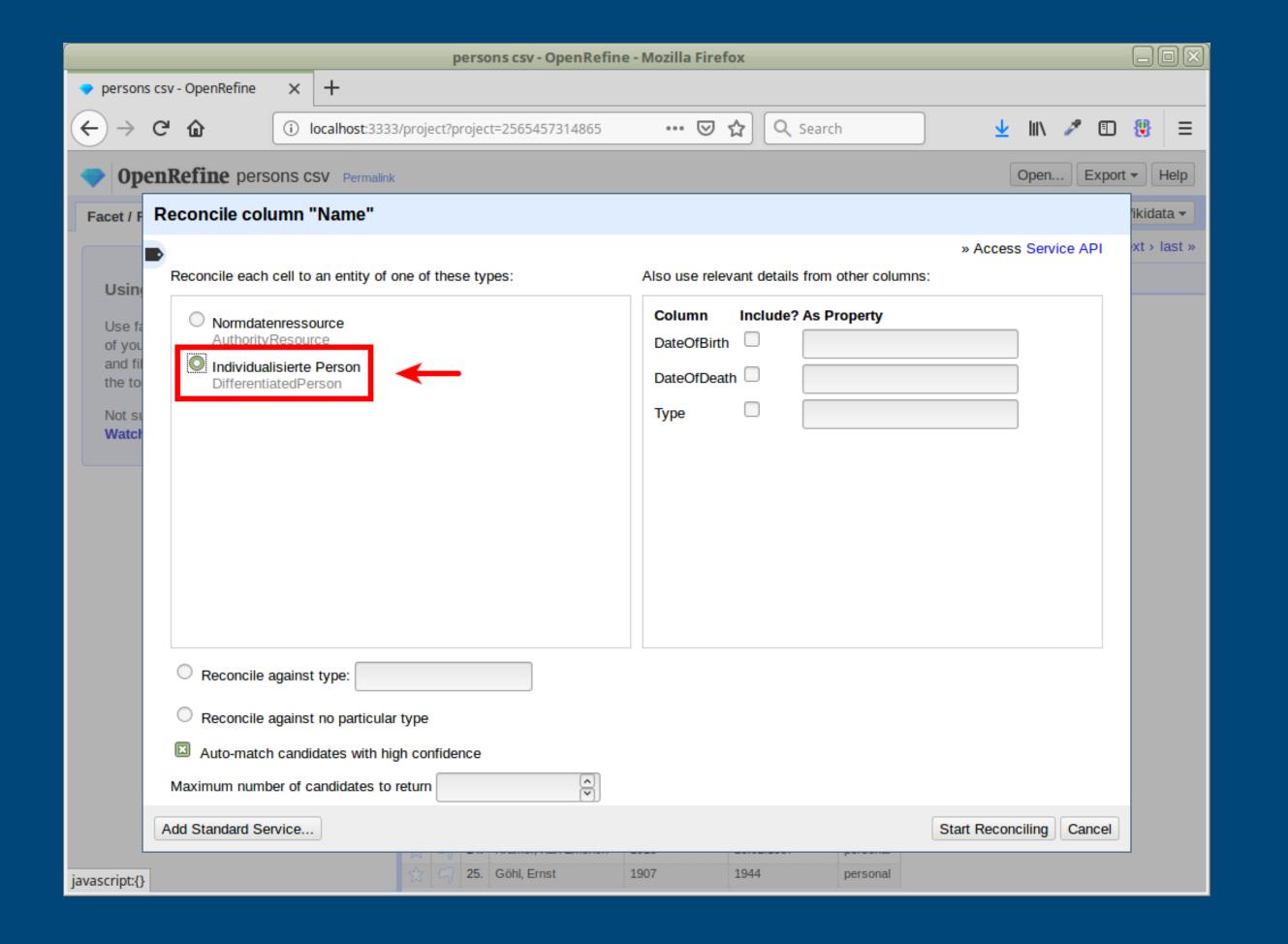
2. Beispiel: Beruf & Parteizugehörigkeit

### 1. Beispiel: Lebensdaten



CSV lokal speichern oder direkt per URL: https://blog.lobid.org/data/persons.csv

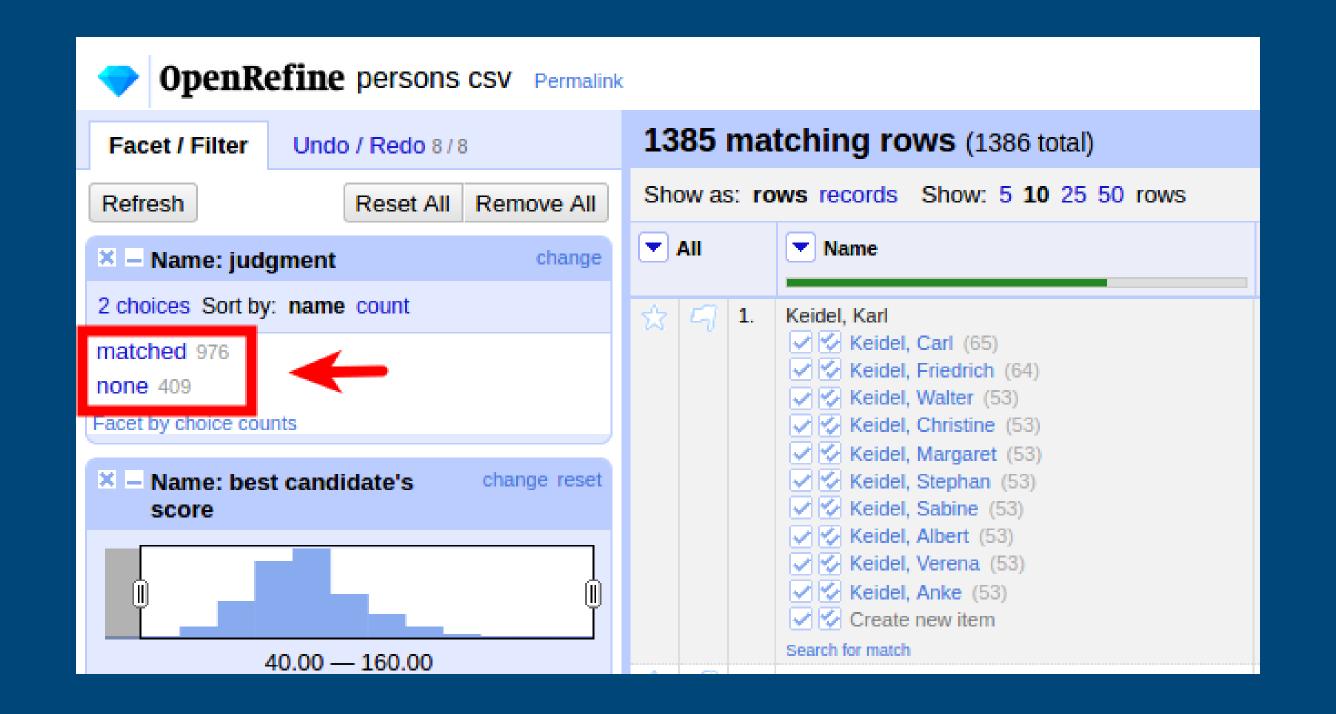




## Project anlegen und Namen abgleichen

openrefine.labs.lobid.org | localhost:3333

persons.csv



### Ein erstes Ergebnis: 70% automatisch abgeglichen

## Wir wollen die Lebensdaten zur Disambiguierung nutzen

#### Wir wollen die Lebensdaten zur Disambiguierung nutzen

In unseren abzugleichenden Daten z.B. "1904"

# Wir wollen die Lebensdaten zur Disambiguierung nutzen

In unseren abzugleichenden Daten z.B. "1904"

Womit gleichen wir ab? z.B. Ellen Frank

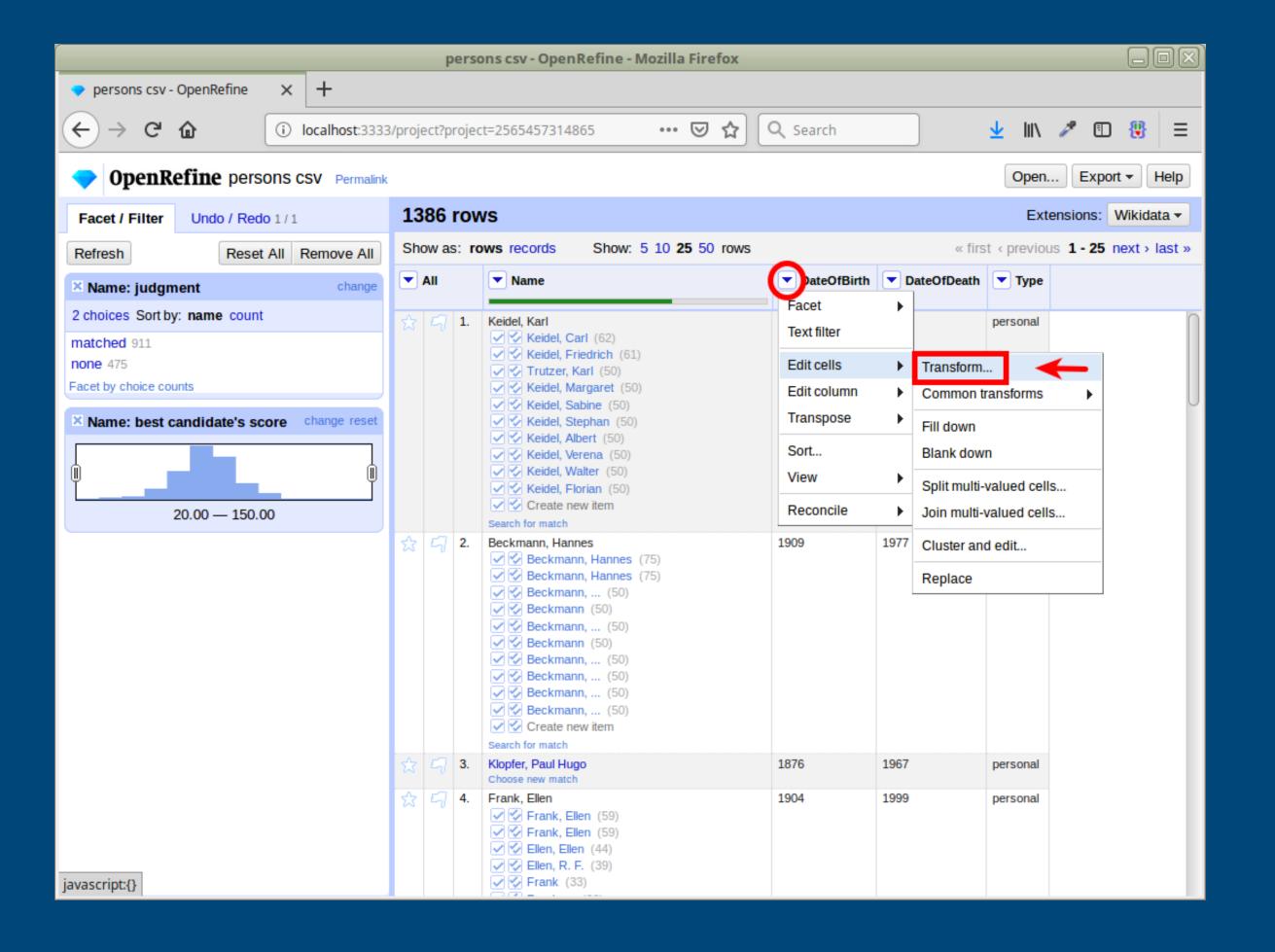
# Wir wollen die Lebensdaten zur Disambiguierung nutzen

In unseren abzugleichenden Daten z.B. "1904"

Womit gleichen wir ab? z.B. Ellen Frank

```
"dateOfDeath": [
    "1999-09-17"
],
"dateOfBirth": [
    "1904-03-09"
]
```

F



Für die Transformation verschiedene Sprachen

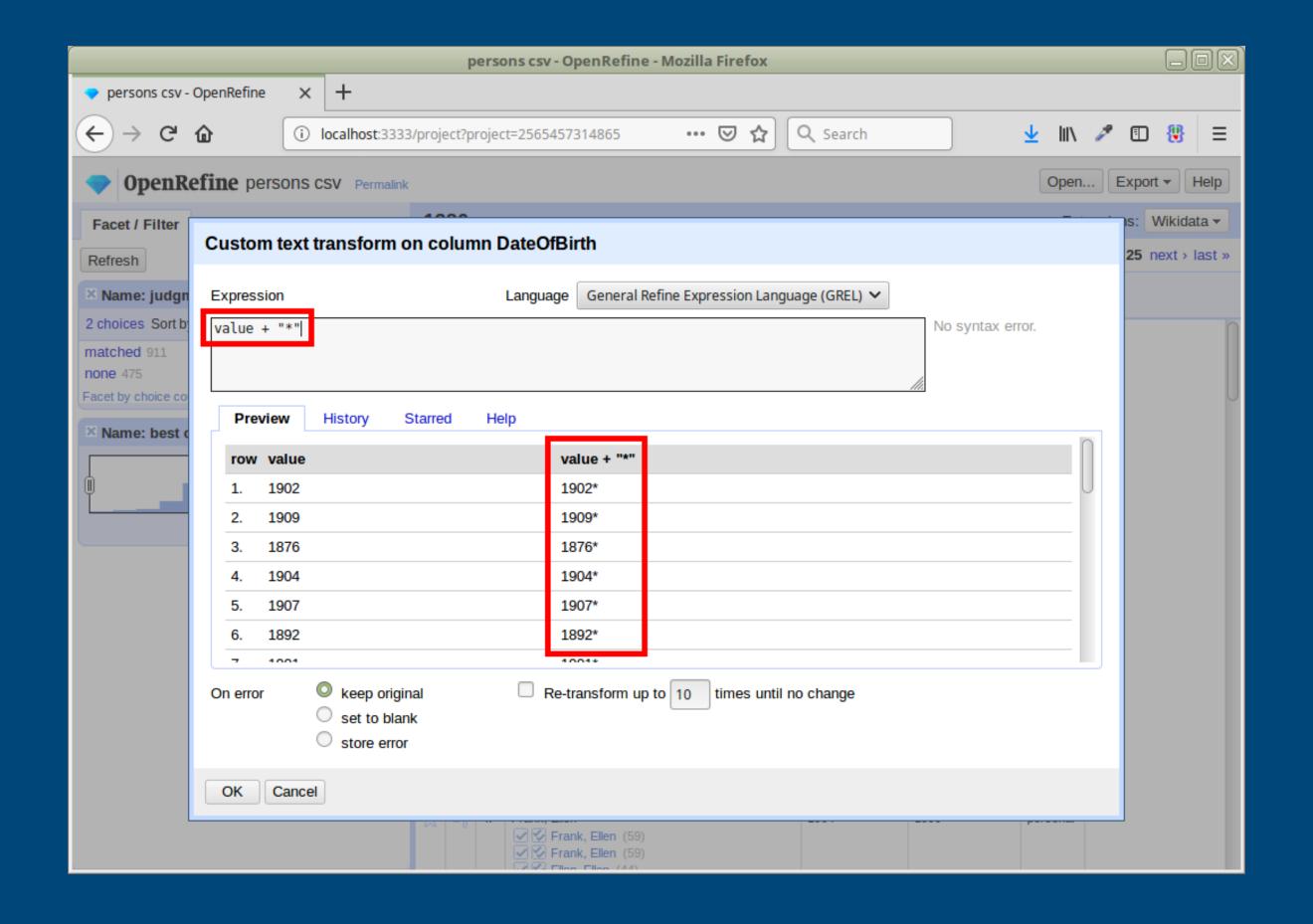
Für die Transformation verschiedene Sprachen

General Refine Expression Language (GREL)

Für die Transformation verschiedene Sprachen

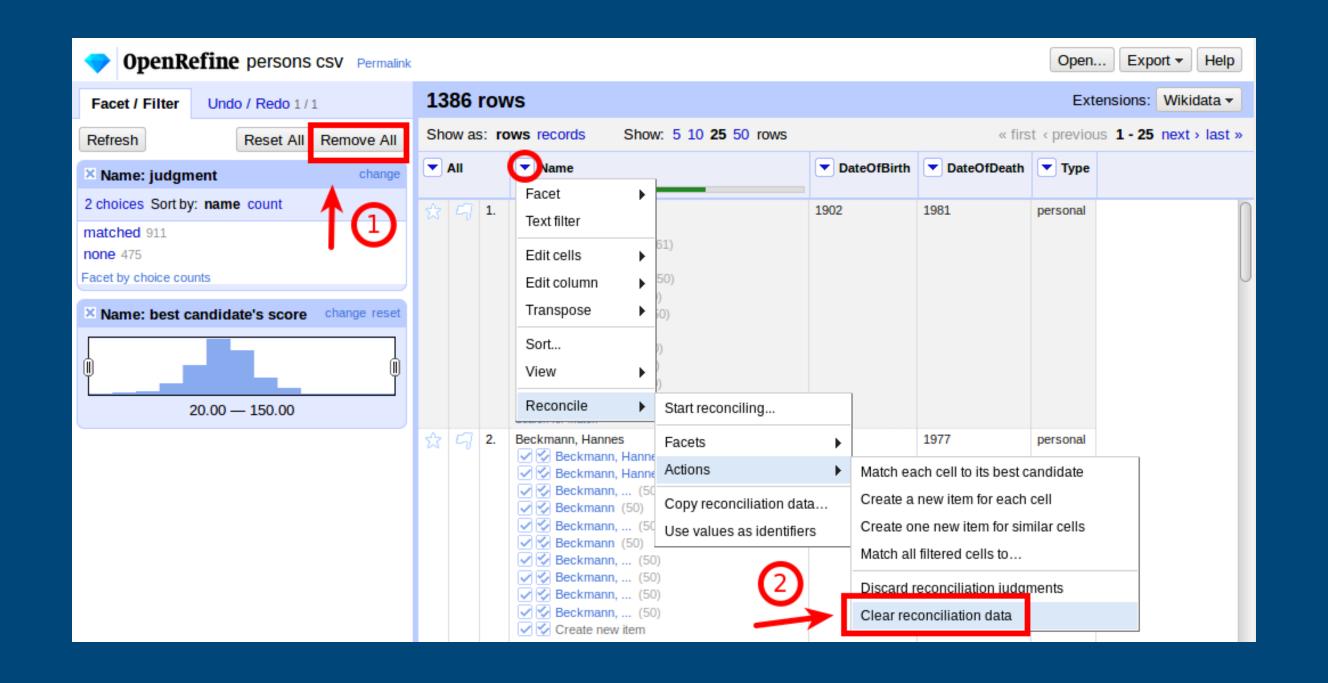
General Refine Expression Language (GREL)

Python/Jython & Clojure



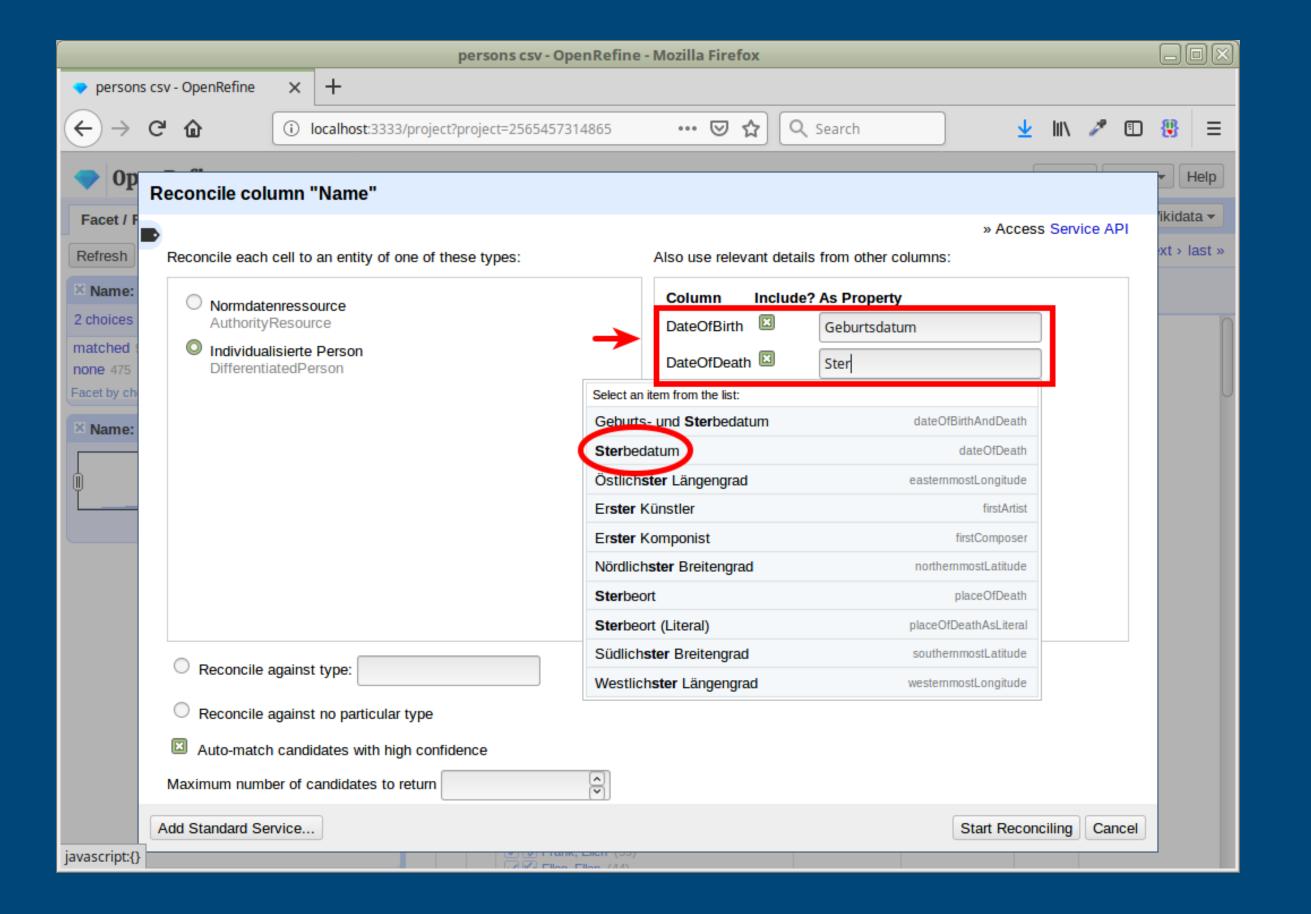
Wildcards: ? für ein Zeichen, \* für beliebig viele Zeichen

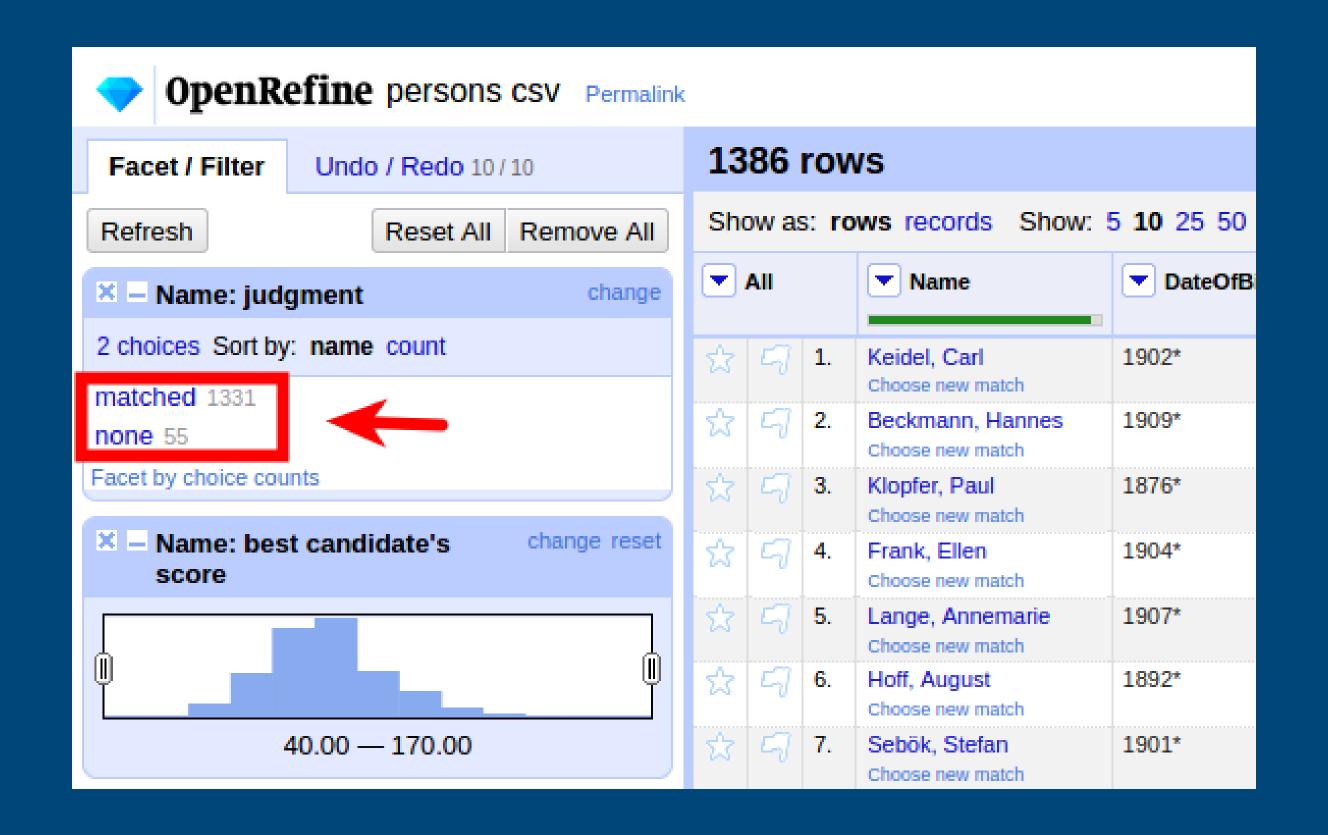
#### Daten bereit für neuen, verbesserten Abgleich



Vorherigen Abgleich zurücksetzen

170

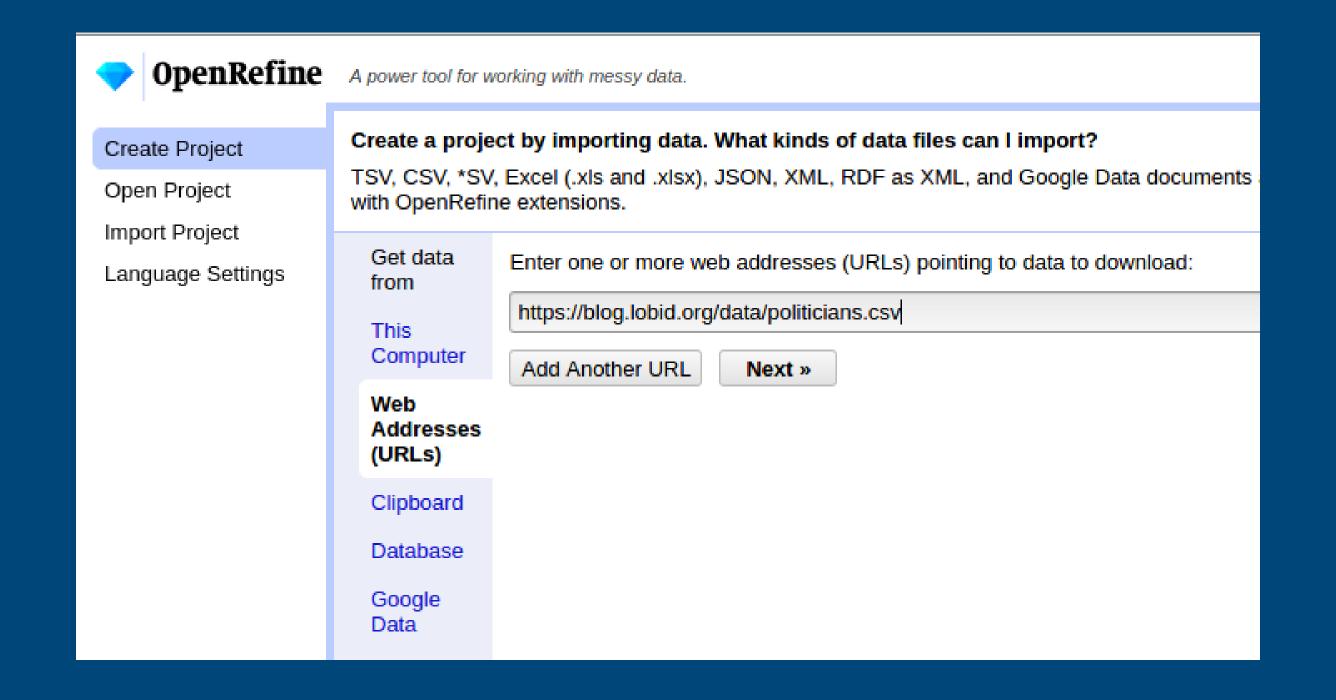




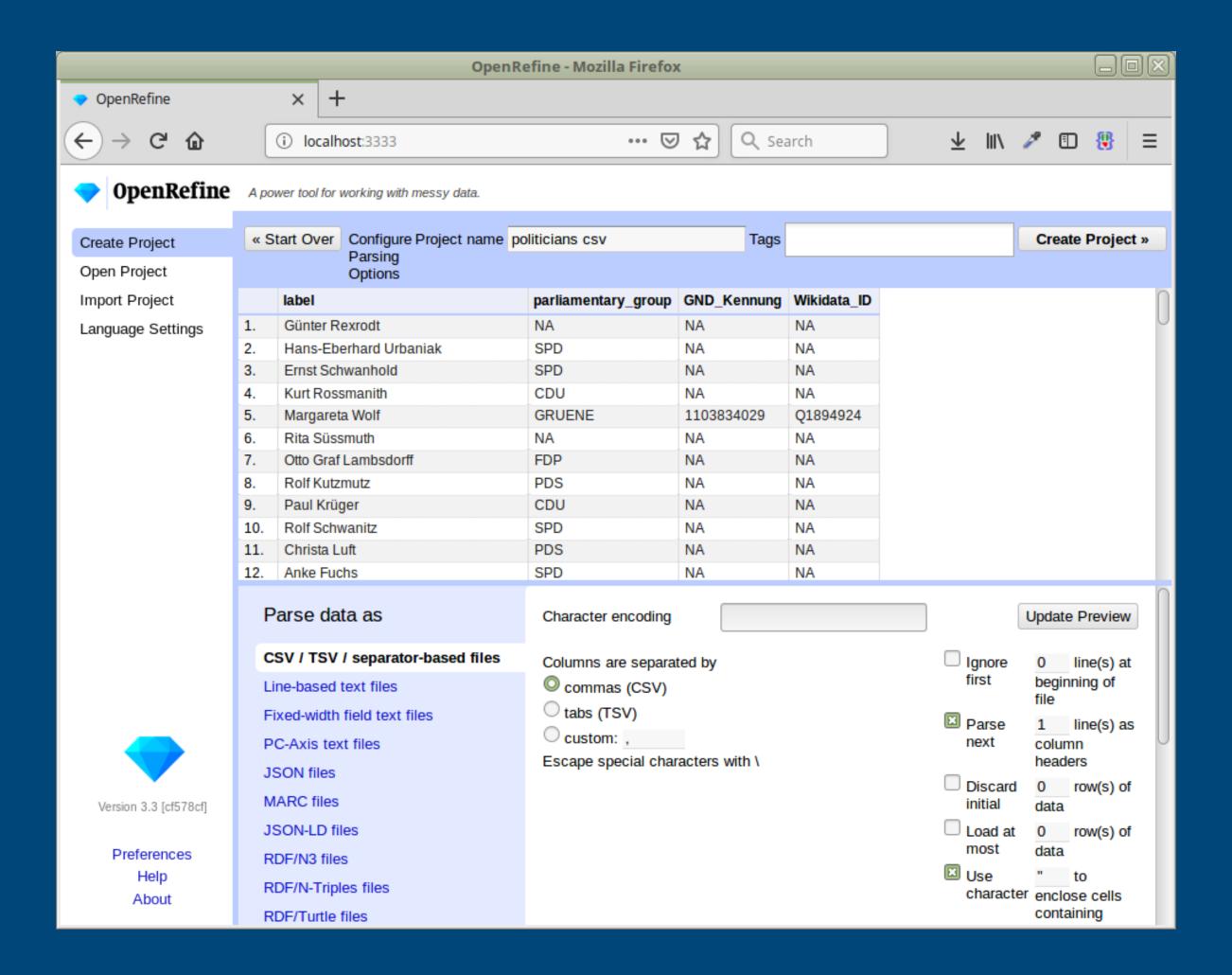
#### >95% automatisch abgeglichen

1/2

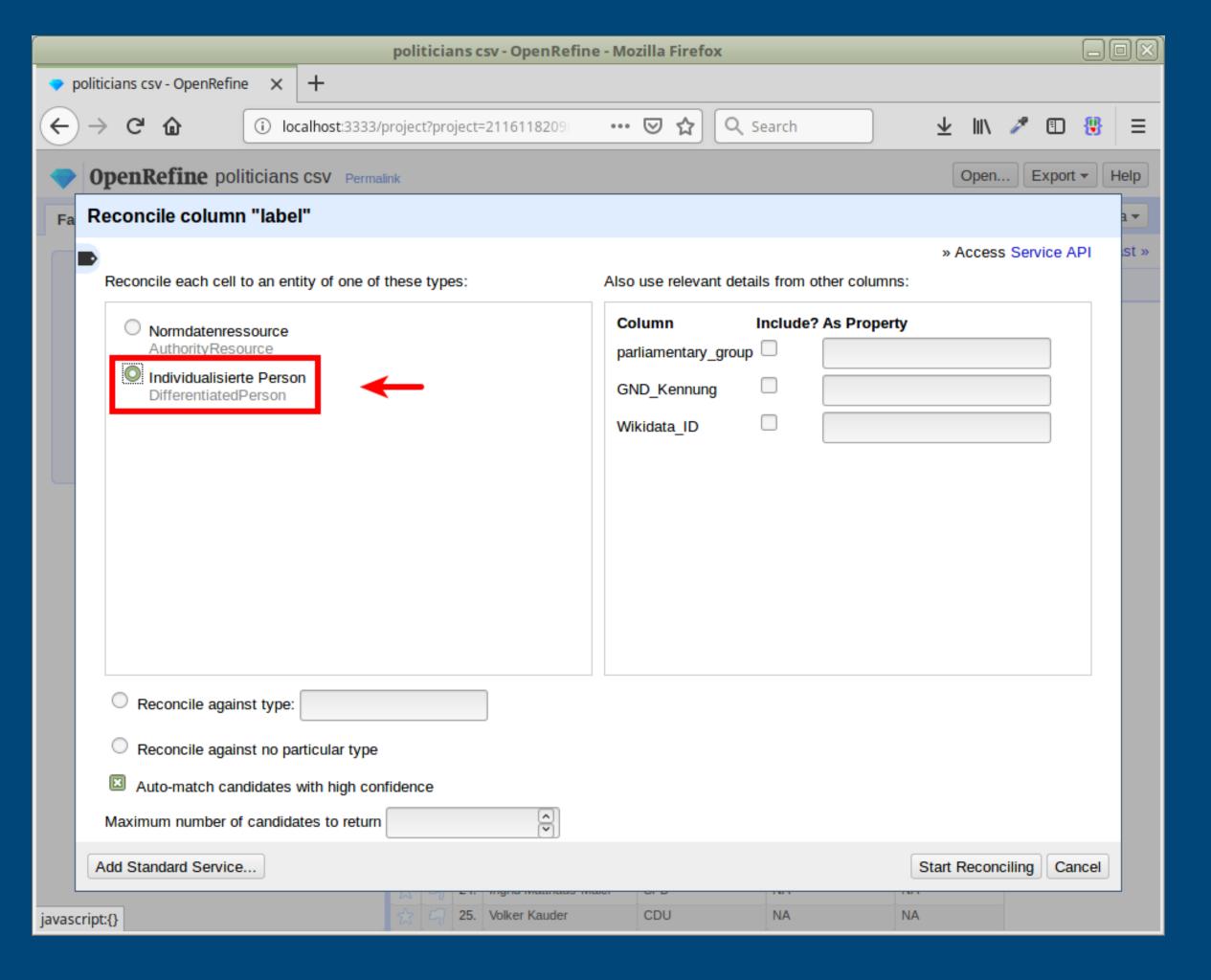
# 2. Beispiel: Beruf & Parteizugehörigkeit



CSV lokal speichern oder direkt per URL: https://blog.lobid.org/data/politicians.csv

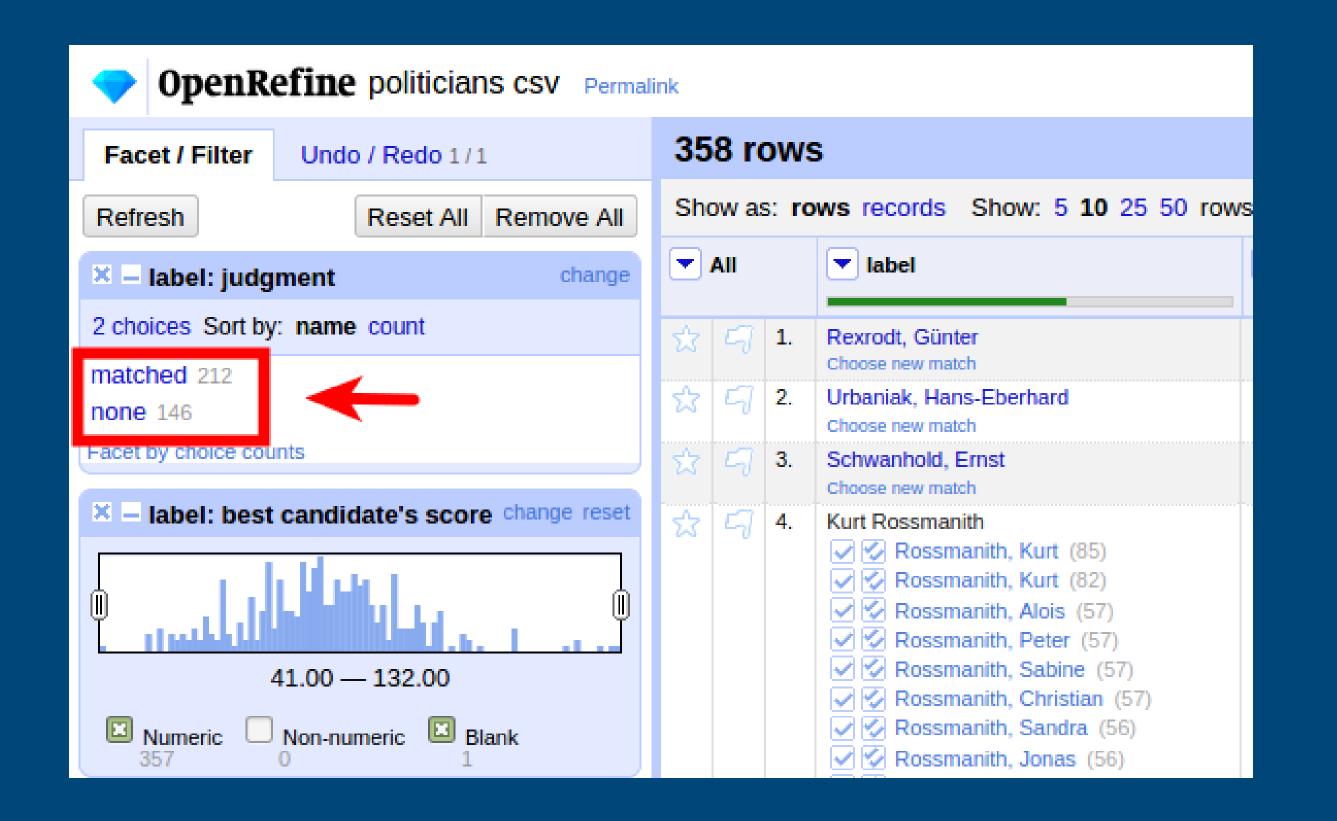


F 175



## Project anlegen und Namen abgleichen

openrefine.labs.lobid.org | localhost:3333 politicians.csv



### Ein erstes Ergebnis:

< 60% automatisch abgeglichen

⊢

## Wir wollen die Parteizugehörigkeit nutzen

### Wir wollen die Parteizugehörigkeit nutzen

In unseren abzugleichenden Daten z.B. SPD

## Wir wollen die Parteizugehörigkeit nutzen

In unseren abzugleichenden Daten z.B. SPD

Womit gleichen wir ab? z.B. Willy Brandt

## Wir wollen die Parteizugehörigkeit nutzen

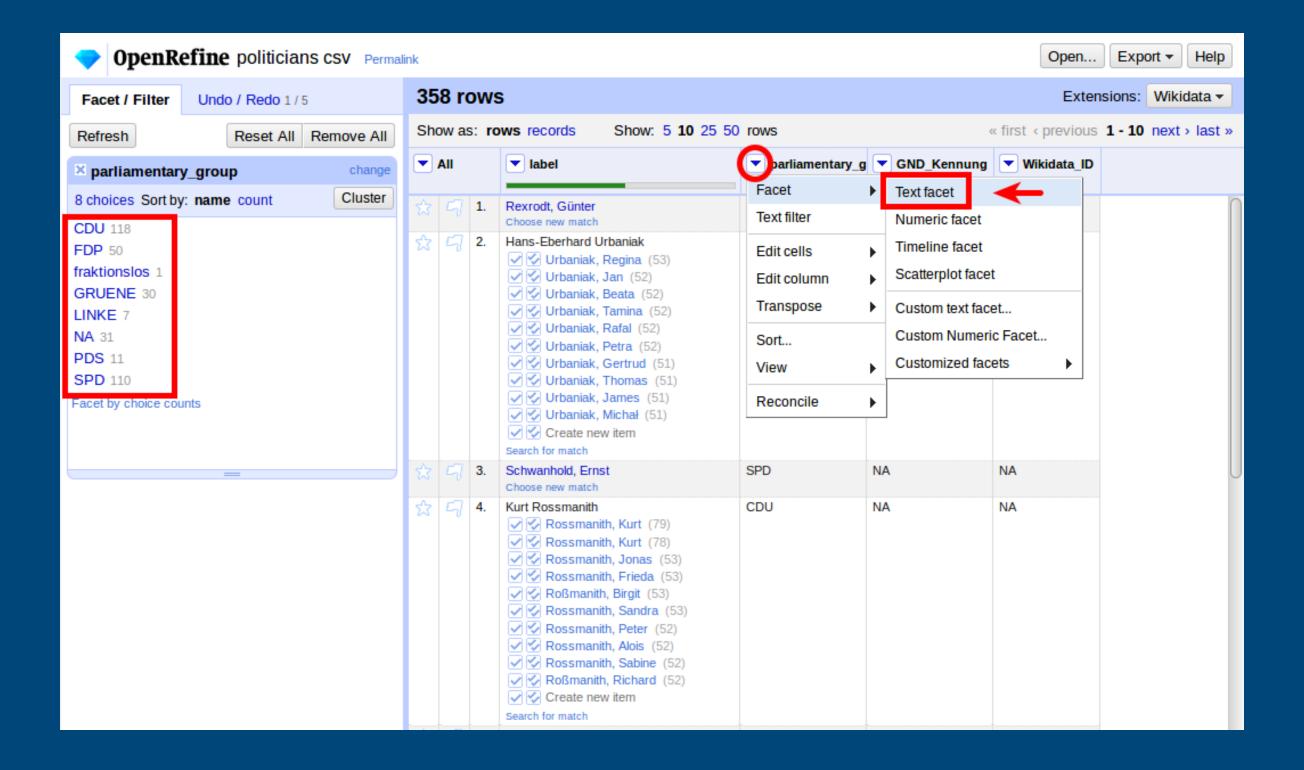
In unseren abzugleichenden Daten z.B. SPD

Womit gleichen wir ab? z.B. Willy Brandt

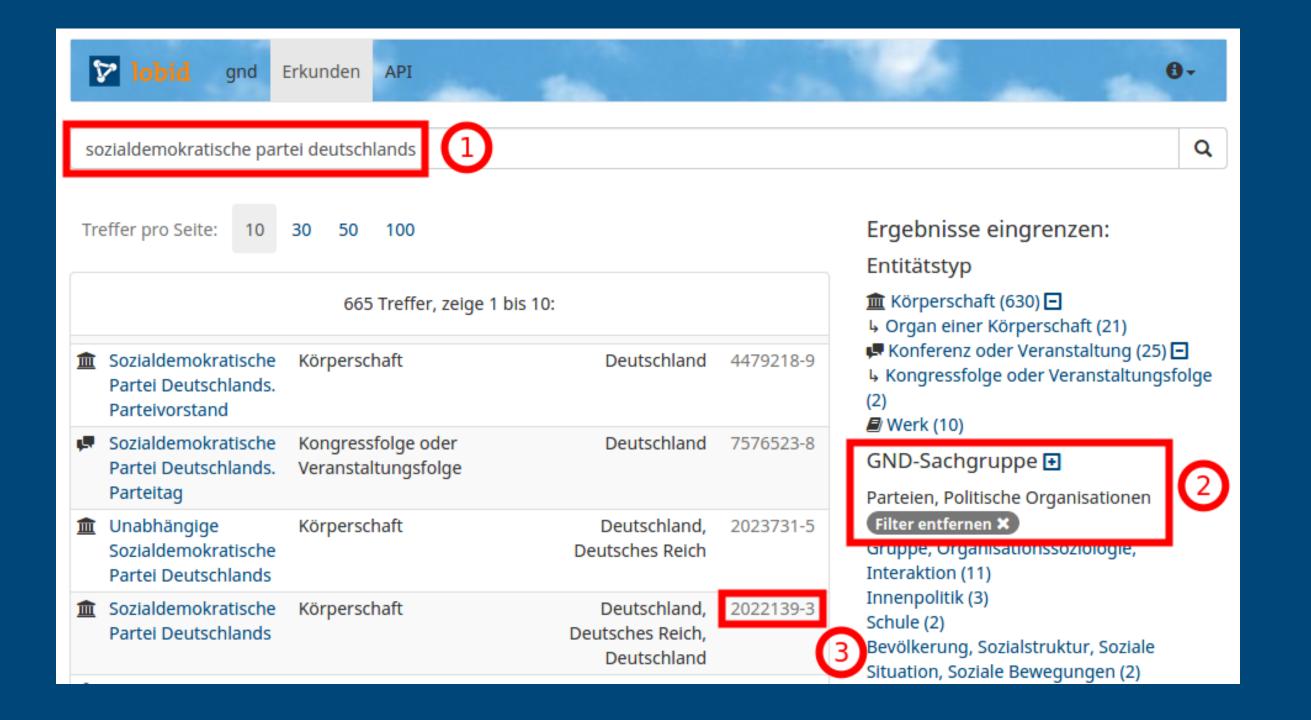
```
"affiliation": [
     {
        "id": "https://d-nb.info/gnd/2022139-3",
        "label": "Sozialdemokratische Partei Deutschlands"
     }
]
```

## Abkürzungen auf eindeutige Identifier mappen

### Welche Abkürzungen kommen in den Daten vor?



#### Wie finden wir die GND-Nummern? z.B. SPD:



### Gesuchte Mappings

- SPD → https://d-nb.info/gnd/???
- CDU → https://d-nb.info/gnd/???
- GRUENE → https://d-nb.info/gnd/???
- FDP → https://d-nb.info/gnd/???
- LINKE → https://d-nb.info/gnd/???
- PDS → https://d-nb.info/gnd/???

### Mappings

- SPD → https://d-nb.info/gnd/2022139-3
- CDU → https://d-nb.info/gnd/7230-8
- GRUENE → https://d-nb.info/gnd/2124337-2
- FDP → https://d-nb.info/gnd/37037-X
- LINKE → https://d-nb.info/gnd/10173685-X
- PDS → https://d-nb.info/gnd/5010217-5

### In der Text-Facette können wir alle Vorkommen der Werte entsprechend unseres Mappings editieren:



### Diese manuelle Schritte können in der Undo/Redo-History extrahiert werden:



U

#### Wir bekommen die Editierhistorie als JSON:

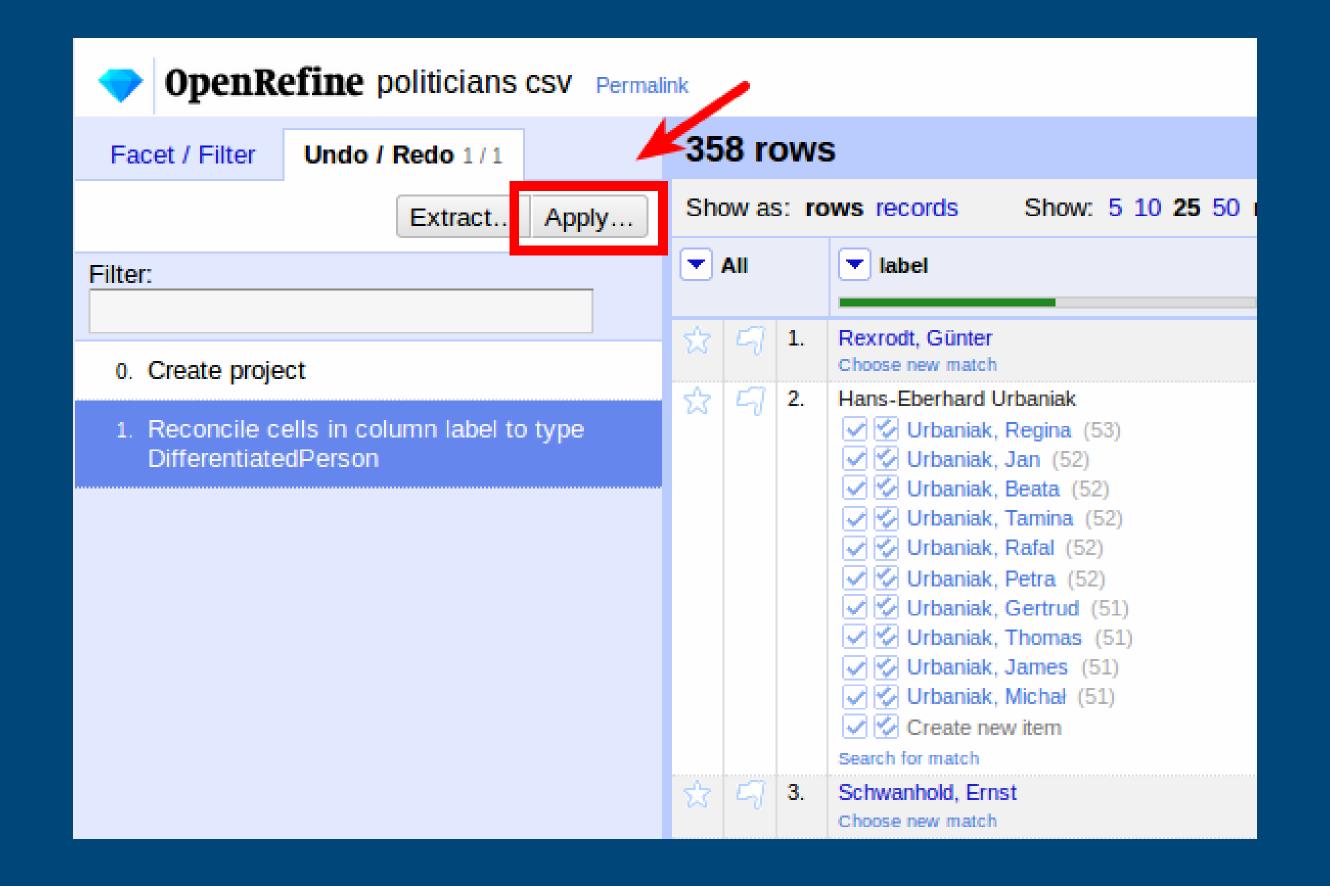
```
"op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    "baseColumnName": "parliamentary_group",
    "expression": "grel:value",
    "onError": "set-to-blank",
    "newColumnName": "affiliation",
    "columnInsertIndex": 2,
    "description": "Create column affiliation at index 2 based
on column parliamentary_group using expression grel:value"
```

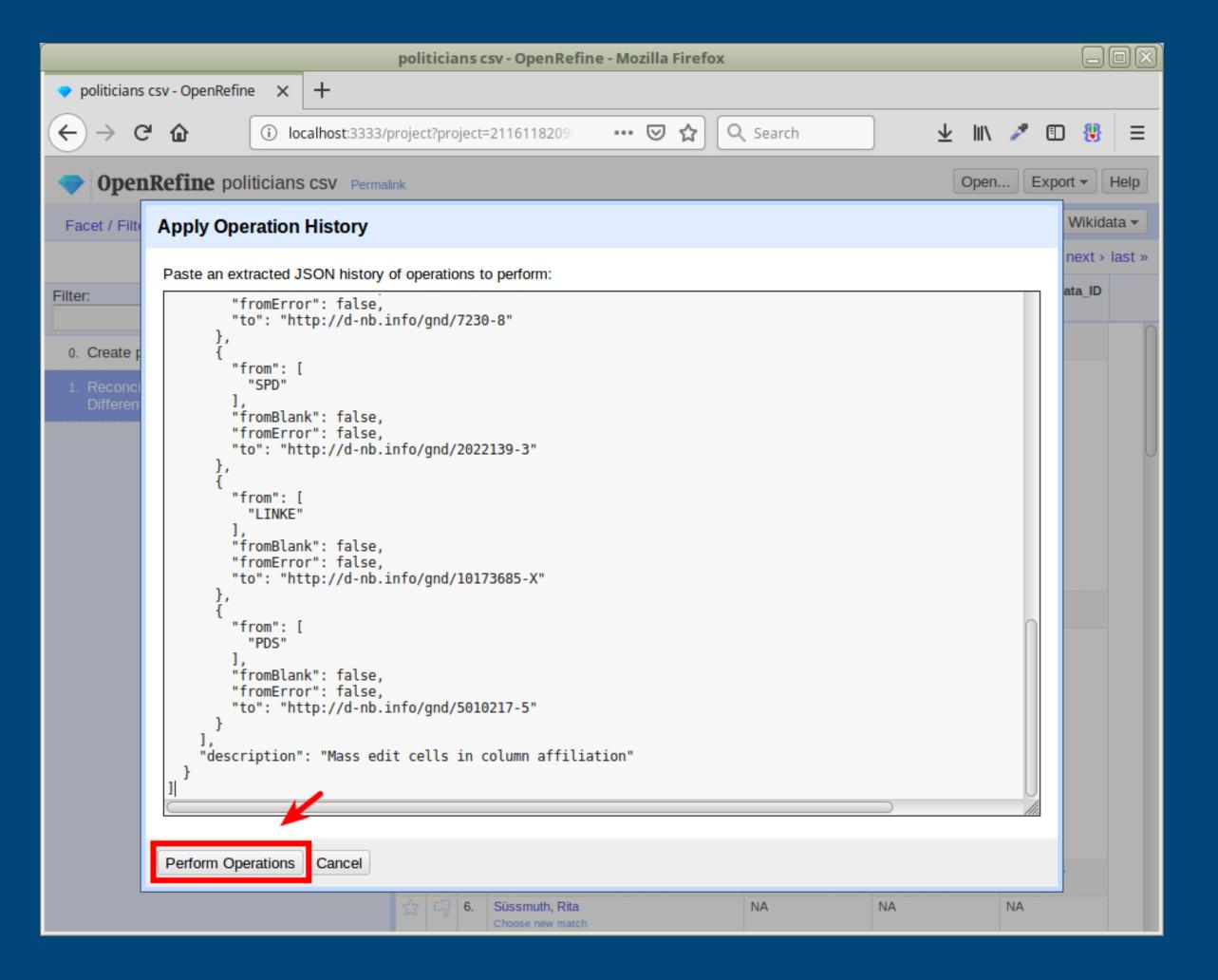
#### Wir bekommen die Editierhistorie als JSON:

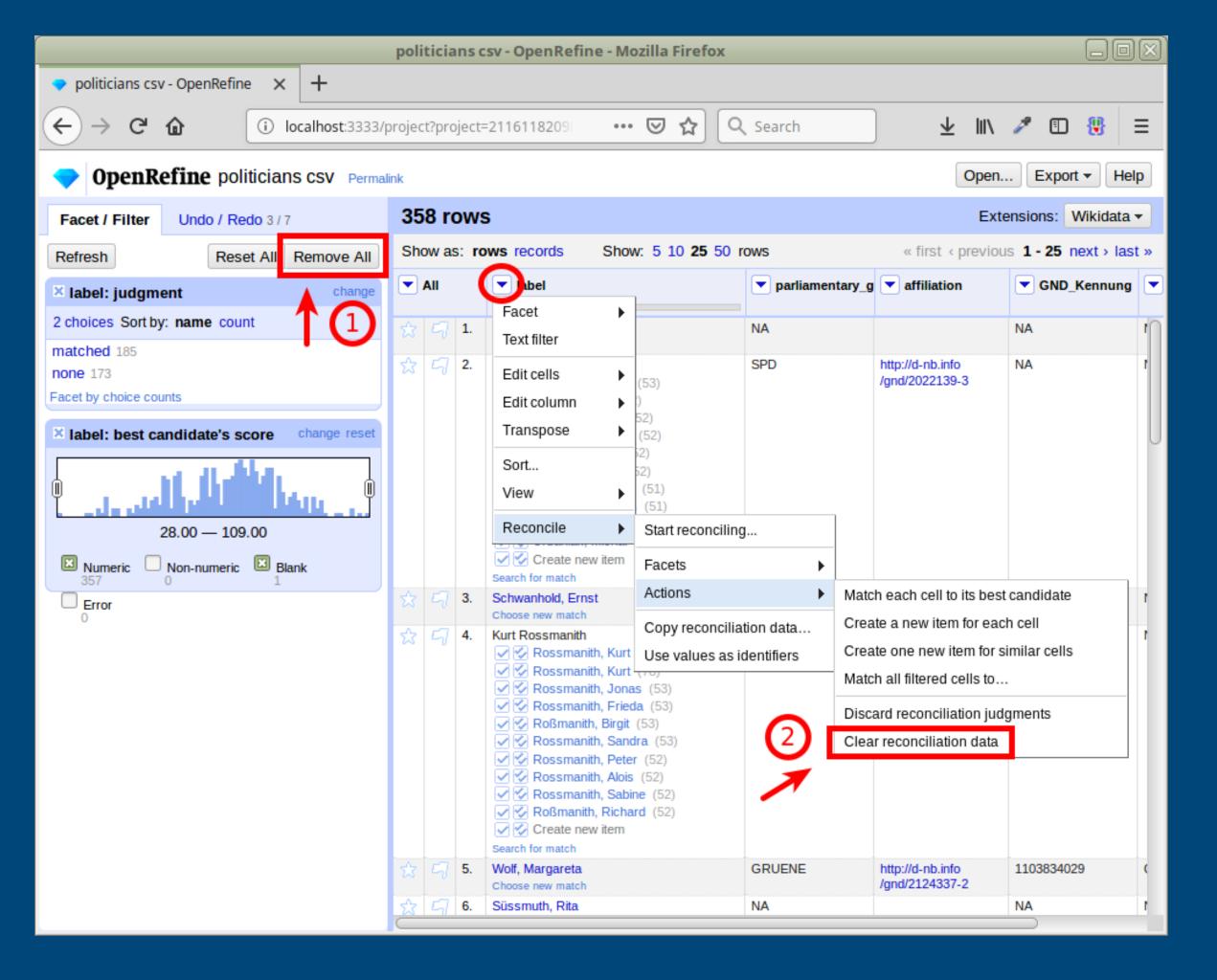
```
"op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    "baseColumnName": "parliamentary_group",
    "expression": "grel:value",
    "onError": "set-to-blank",
    "newColumnName": "affiliation",
    "columnInsertIndex": 2,
    "description": "Create column affiliation at index 2 based
on column parliamentary_group using expression grel:value"
```

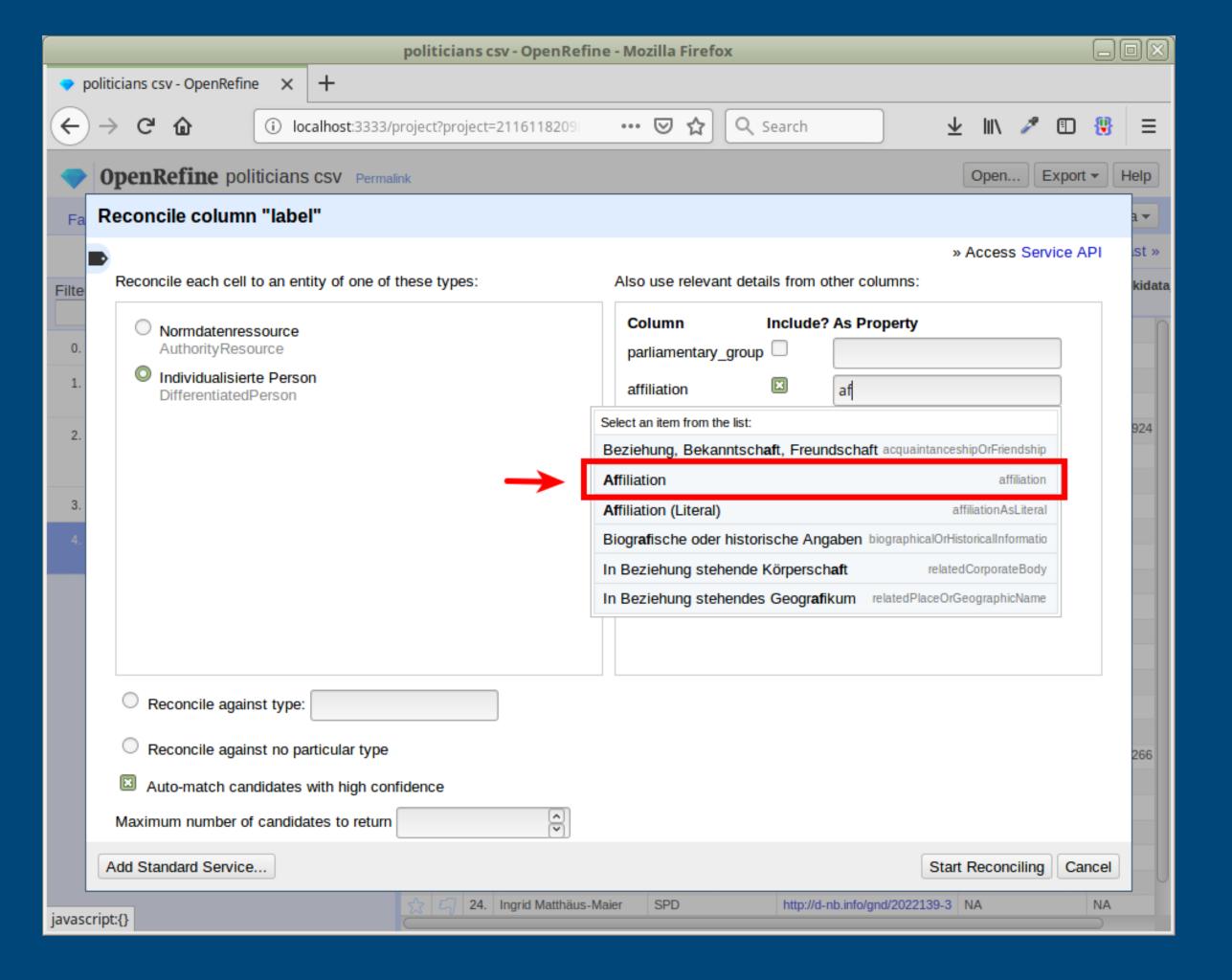
(Hier setzt auch die erwähnte Möglichkeit zur Automatisierung an: die manuell von der Fachebene in der Oberfläche erstellten Schritte können so ohne Oberfläche automatisiert werden)

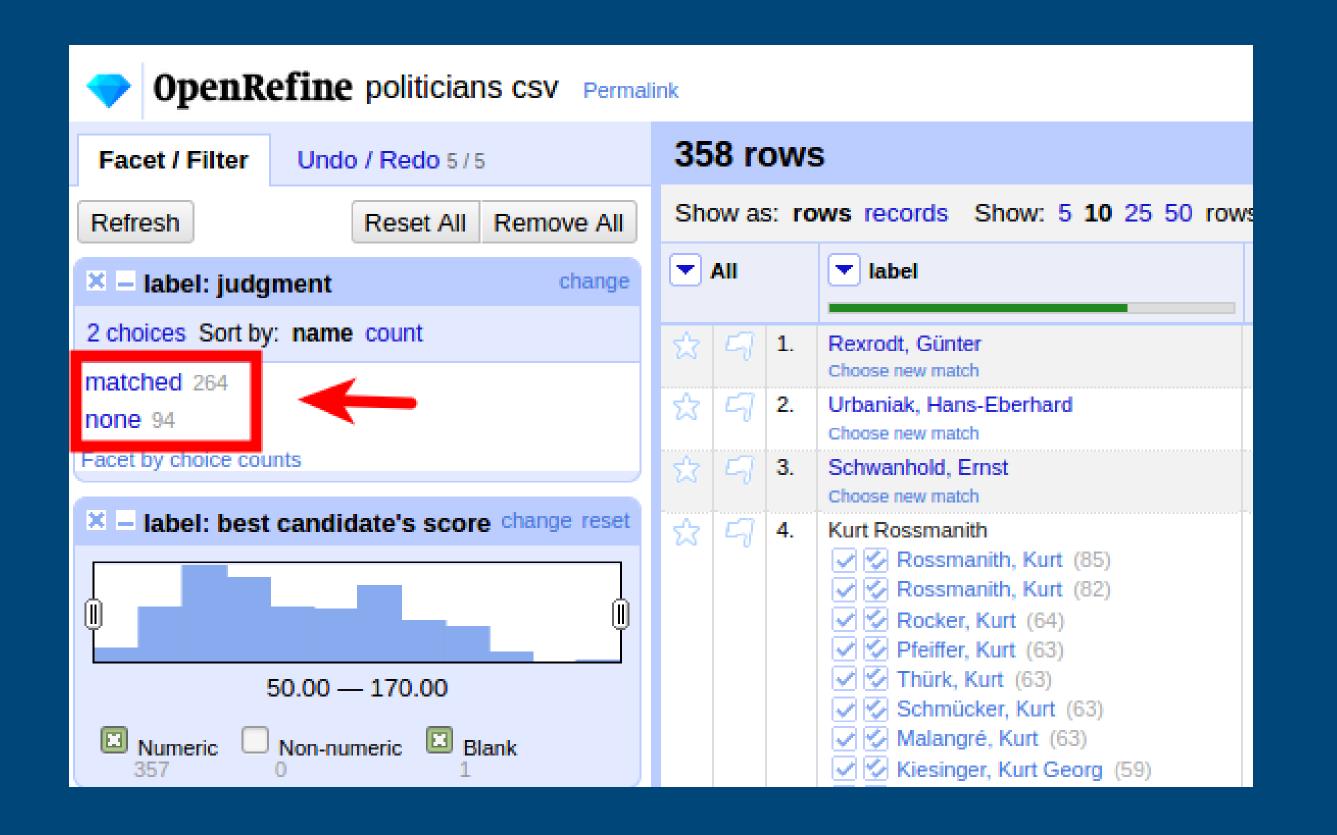
#### Operationen können angepasst und reproduziert werden:











Zwischenergebnis: > 70% automatisch abgeglichen

 $\mathsf{F}$ 

Hintergrundwissen, das so gar nicht in den Daten ist

Hintergrundwissen, das so gar nicht in den Daten ist

Formal ergänzen für die Disambiguierung

Hintergrundwissen, das so gar nicht in den Daten ist

Formal ergänzen für die Disambiguierung

Hier z.B. wissen wir, dass es sich um Politiker:innen handelt

D.h. wir ergänzen eine neue Spalte mit dem Beruf

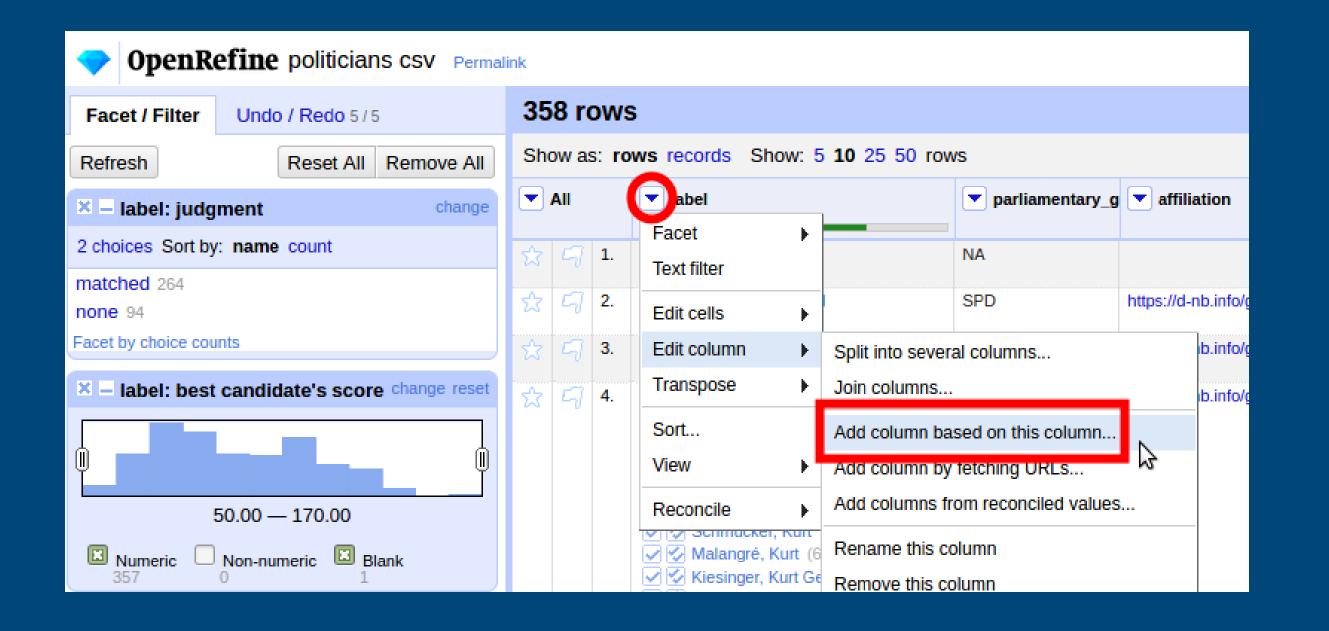
D.h. wir ergänzen eine neue Spalte mit dem Beruf

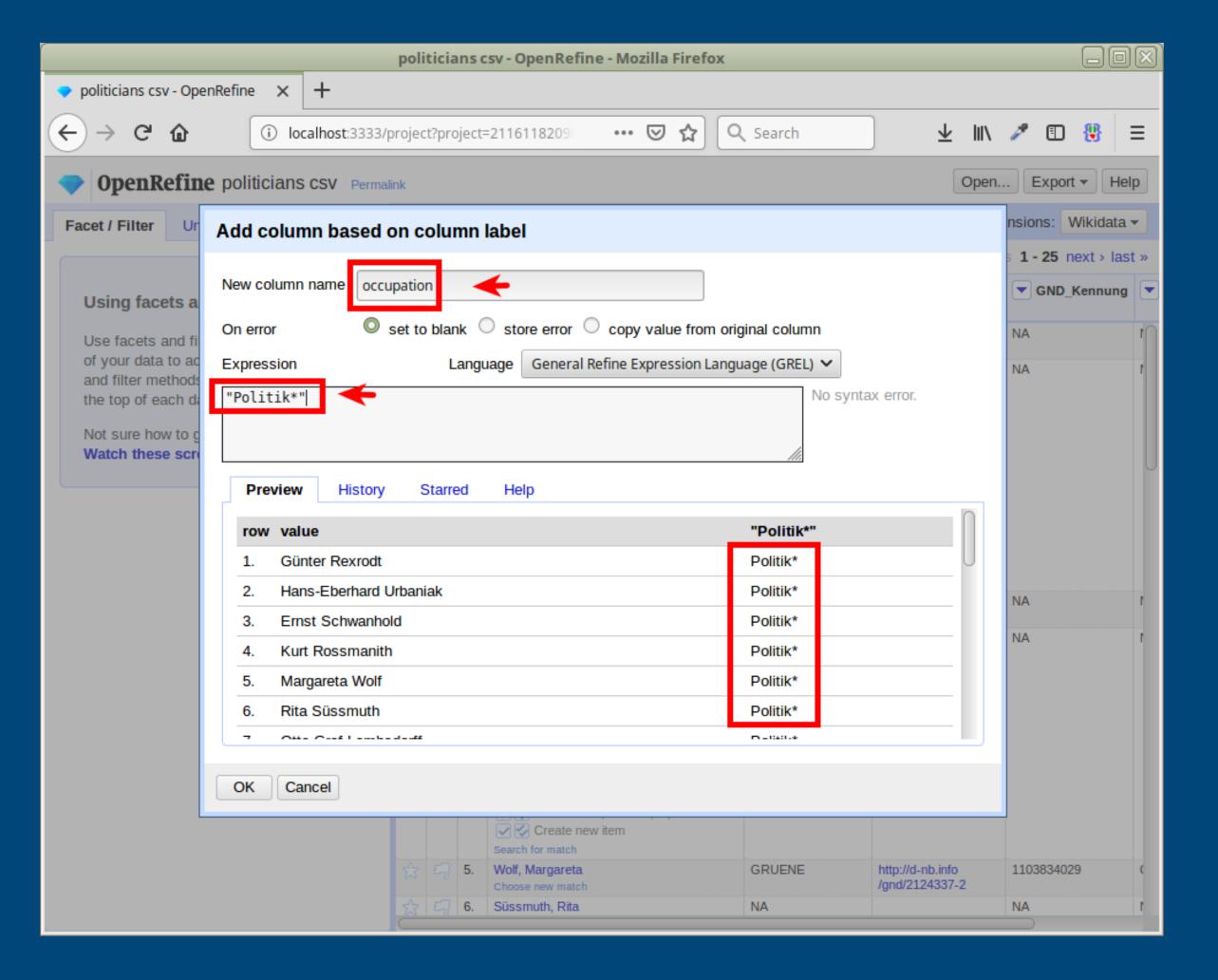
Mit dem Wert Politik\*

D.h. wir ergänzen eine neue Spalte mit dem Beruf

Mit dem Wert Politik\*

Über die Wildcard decken wir mehrere Formen ab: Politiker, Politikerin, Politikwissenschaftler, Politikwissenschaftlerin

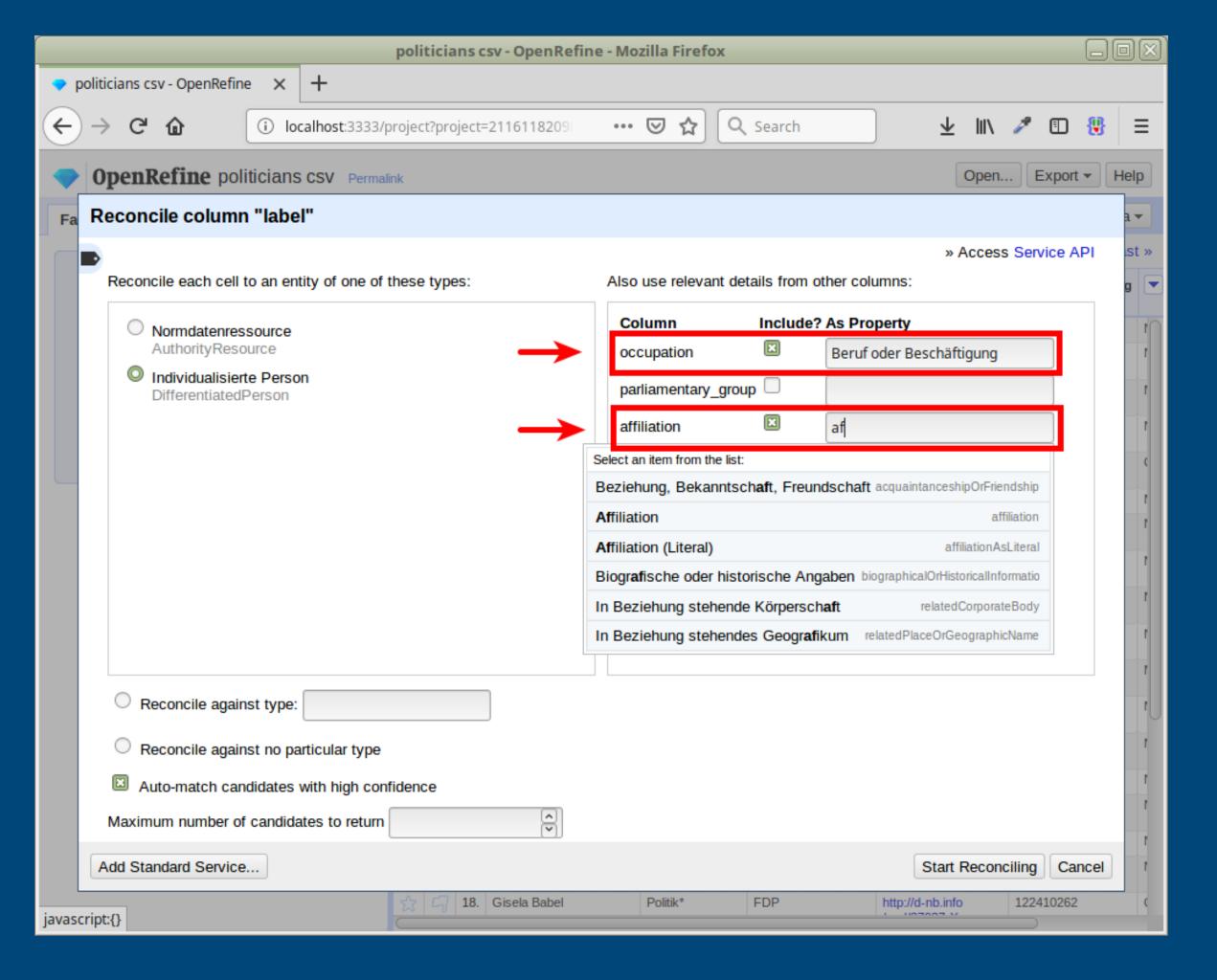


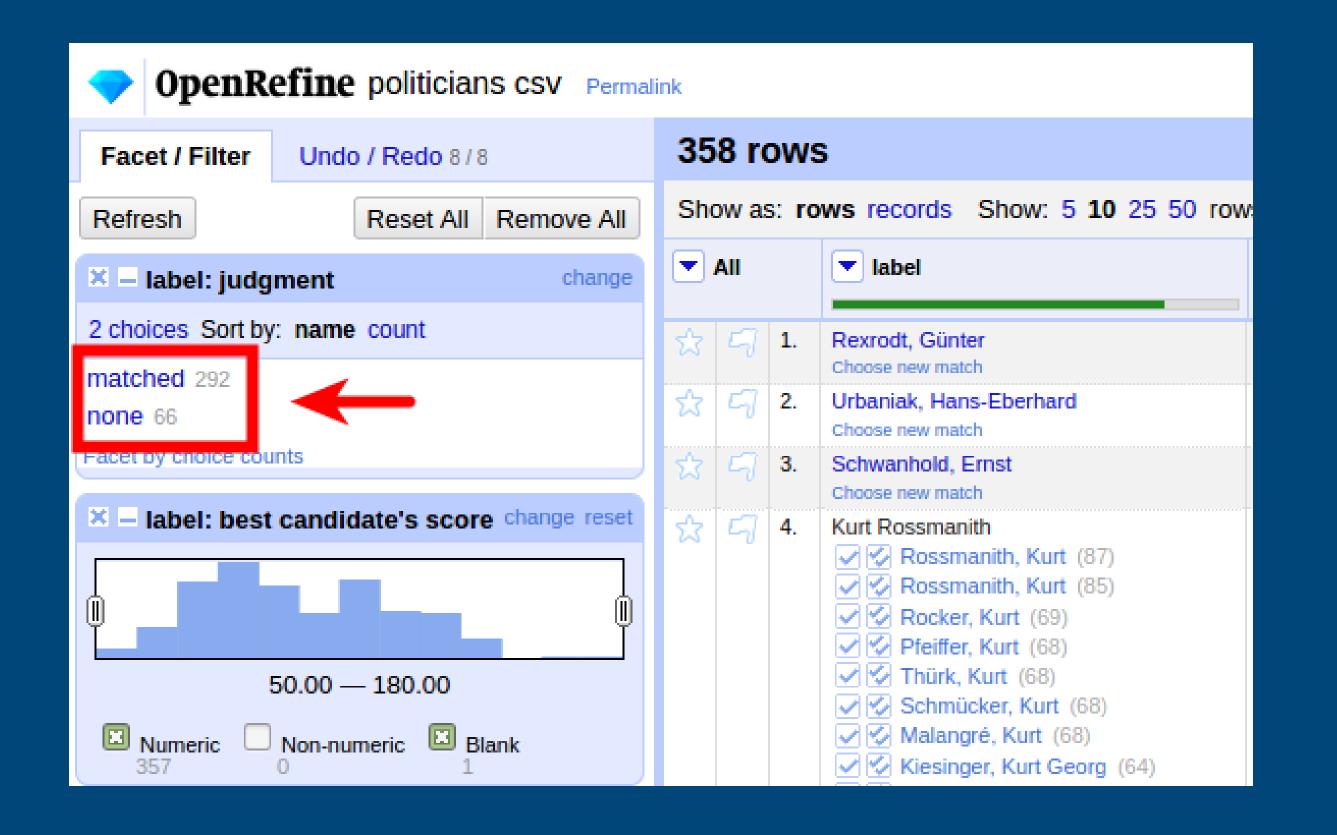


Vorherigen Abgleich zurücksetzen

Vorherigen Abgleich zurücksetzen

Beruf & Affiliation mitschicken





> 80% automatisch abgeglichen

Fertige Daten können flexibel exportiert werden

Fertige Daten können flexibel exportiert werden

https://slides.lobid.org/2021-kim-reconcile/#/99

Fertige Daten können flexibel exportiert werden

https://slides.lobid.org/2021-kim-reconcile/#/99

https://lobid.org/gnd/reconcile

### Weiterführende Informationen – Rund um lobid und Linked Data

- lobid-Blog, Präsentationsfolien, lobid auf Twitter und Mastodon
- Kreutzer, Till (2011): Open Data Freigabe von Daten aus Bibliothekskatalogen. Hg. v. Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (PDF)
- Pohl, Adrian / Steeg, Fabian / Christoph, Pascal (2018): lobid –
   Dateninfrastruktur für Bibliotheken. In: Informationspraxis 4(1).
   https://doi.org/10.11588/ip.2018.1.52445
- Steeg, Fabian / Pohl, Adrian / Christoph, Pascal (2019): lobid-gnd Eine Schnittstelle zur Gemeinsamen Normdatei für Mensch und Maschine. In: Informationspraxis 5(1). https://doi.org/10.11588/ip.2019.1.52673
- Steeg, Fabian / Pohl, Adrian (2021): Ein Protokoll für den Datenabgleich im Web am Beispiel von OpenRefine und der Gemeinsamen Normdatei (GND). In: Qualität in der Inhaltserschließung. https://doi.org/10.1515/9783110691597-013

## Weiterführende Informationen – Nutzung von lobid-gnd

- Stefan Dumont: correspSearch: Neue Tools, neue Daten und eine Auszeichnung https://dhd-blog.org/?p=10632
- Harald Lordick: Die Öffnung der GND und die historische Forschung: https://djgd.hypotheses.org/1511
- Harald Lordick & Beata Mache: Annotationen anhand der Gemeinsamen Normdatei aus einer anwendungsorientierten Perspektive historischer Forschung https://zenodo.org/record/1188230
- Christian Thomas: Humboldts ›Bekenntnisse‹ von 1801 nach der Original-Handschrift aus den Amerikanischen Reisetagebüchern im Deutschen Textarchiv der BBAW http://www.avhumboldt.de/?p=14447
- Christian Erlinger-Schiedlbauer et al.: Semantische Suchabfragen mit der Linked Open Data Cloud generieren https://www.societybyte.swiss/2018/10/24/semantische-suchabfragen-mit-derlinked-open-data-cloud-generieren/