

DH Einführungen

Webschnittstellen (API) und JSON

Den Aufbau von Webschnittstellen verstehen, um gezielt Daten abzufragen und weiterzuverarbeiten

DH Einführungen

- **Was sind die Digital Humanities?**
 - Interdisziplinäres Feld.
 - Bearbeitung geistes- und sozialwissenschaftlicher Fragen mit digitalen Methoden
 - Digitale Editionen, digitale Sprach- und Literaturanalyse, Georeferenzierung von Daten, Netzwerkanalysen, Linked Open Data,...
- **Digital Humanities an der RUB**
 - **Netzwerk Digital Humanities** (<https://dh-netzwerk.blogs.ruhr-uni-bochum.de/>)
 - **Digital Humanities Center** der UB (<https://dh.rub.de>)
 - **Methodenzentrum** (<https://methodenzentrum.ruhr-uni-bochum.de/>)
 - **DH Ruhr** (<https://dh-ruhr.ruhr-uni-bochum.de/>)

DH Einführungen – Ziele der Veranstaltungsreihe

- **Grundverständnis von für die Digital Humanities wichtigen Formaten, Techniken und Tools**
- **Einblick in die Möglichkeiten der jeweiligen Techniken**
- **Mehr gibt es**
 - Im Lehrplan der Institute
 - Im Angebot des Methodenzentrums
 - Selbstlernmaterialien im Internet
 - Co-working & Coding im DH Space (alle 14-Tage Donnerstags)

DH Einführungen – Ziele dieser Veranstaltung

- **Verständnis des Aufbaus von RESTful Webschnittstellen**
- **Kennenlernen des JSON Formats**
- **Anwendung einer Datenabfrage über API**
- **Kennenlernen der Möglichkeiten der Datenabfrage und Bereinigung mit OpenRefine**

Webschnittstellen (API)

- **API steht für Application Programming Interface = Anwenderschnittstelle**
 - Funktionsablauf: Anfrage → Verarbeitung → Antwort
- **REST steht für Representational State Transfer = Softwarearchitektur-Prinzipien**
 - Client – Server
 - Zustandslose Client – Server Kommunikation
 - Einheitliche Schnittstelle überträgt Informationen in Standardform

Datenabfrage über eine REST API

- **Abfrage über URL mit Parametern:**
 - z.B. `https://lobid.org/resources/search?q=digital&format=json`
- **Antwort in Form maschinenlesbarer Daten:**
Typische Datenformate:
 - HTML (Website)
 - XML, JSON (Datenaustausch zur Weiterverarbeitung)
 - IIIF (Bilder)

Aufbau einer typischen Abfrage URL

- <https://lobid.org/resources/search?q=digital&format=json>
- Basis URL: `https://lobid.org`
- Abfragespezifizierung: `/resources/search`
- Abfrageparameter: `?q=digital&format=json`
 - Suchbegriff: `q=digital`
 - Format der Antwort: `format=json`

Maskierung reservierter Zeichen in der URL

- **Problem:** ich will nicht nur nach „digital“ suchen, sondern nach „digital humanities“.
Eine **URL** darf aber **kein Leerzeichen** besitzen.
- **Lösung:** Zeichen, die in URL nicht erlaubt sind, oder eine reservierte Bedeutung haben werden **maskiert**, d.h. durch ihren Codepunkt im ASCII-Zeichensatz ersetzt.

Leerzeichen	:	"	/	+	&	%	?	\$,	=
%20	%3A	%22	%2F	%2B	%26	%25	%3F	%24	%2C	%3D

<https://lobid.org/resources/search?q=title%3A%22digital%20humanities%22&format=json>

Das JSON Format

- **JSON = JavaScript Object Notation.** Anders als der Name suggeriert, ist das Format zwar JavaScript konform, aber dennoch unabhängig von einer bestimmten Programmiersprache.
- Ist eine Sammlung von **Name/Wert Paaren**.
 - Ein Objekt ist eine ungeordnete Menge von Name/Wert Paaren
{ „Titel“ : „Mein Titel“, „Autor“ : „Ich“, ... }
 - Ein Array ist eine geordnete Liste von Werten. Ein Wert kann ein Objekt, ein Array, eine Zeichenkette, Zahl oder true/false/null sein.
[„Zeichenkette“, 42, true, [24, { „Name1“ : „Wert 1“ }], { „Name2“ : 2 }]

Die Antwort als JSON

- **Auszug aus dem zurück gelieferten JSON:**

```
{
  "@context": "http://lobid.org/resources/context.jsonld",
  "id": "http://lobid.org/resources/search?
q=title%3A%22digital%20humanities%22&format=json",
  "totalItems": 628,
  "member": [
    {
      "extent": "205 S. : Ill., graph. Darst.",
      "subject": [
        {
          "notation": "AK 39950",
          "source": {
            "label": "RVK (Regensburger Verbundklassifikation)",
            "id": "https://d-nb.info/gnd/4449787-8"
          },
          "type": [
            "Concept"
          ]
        }
      ],
    },
    ...
  ],
}
```

Die Antwort:

- Ist für Menschen lesbar
- Ist von Maschinen lesbar
- Erlaubt gezielt anhand der Namen nach Werten zu suchen, die uns interessieren

Die Antwort als HTML

- Manche APIs geben auch eine Webseite in HTML zurück
<https://lobid.org/resources/search?q=title%3A%22digital%20humanities%22&format=html>

lobid resources Erweiterte Suche API Merkliste

title:"digital humanities"

Sortierung: Relevanz Neueste Älteste Treffer pro Seite: 15 25 50 100

628 Treffer, zeige 1 bis 15:

Digital humanities	Gugerli, David et al.	2013
Digital Humanities eine Einführung	Jannidis, Fotis et al.	2017
Digital Humanities	Kurt Fendt et al.	
Digital humanities First edition	Carter, Bryan	2013
Digital Humanities Perspektiven der Praxis	Bockwinkel, Peggy et al.	2018
Digital Humanities Praktiken der Digitalisierung, der Dissemination und der Selbstreflexivität	Schmale, Wolfgang	2015
Digital Humanities Strukturen - Lehre - Forschung Stand: September 2016	Sauer, Marie et al.	2016
Digital humanities		2019
Digital Humanities Eine Einführung 1	Jannidis, Fotis et al.	2017

Ergebnisse eingrenzen:

▼ Erscheinungsjahr

2004-2024

▼ Mitwirkende

- ☐ Walter de Gruyter GmbH & Co. KG (10)
- ☐ Siemens, Raymond George (7)
- ☐ Clivaz, Claire (5)
- ☐ Crompton, Constance (5)
- ☐ Jannidis, Fotis (4)

+ Zeige mehr Mitwirkende

▼ Schlagwörter

- ☐ Digital Humanities (120)
- ☐ Geisteswissenschaften (26)
- ☐ Neue Medien (25)
- ☐ Digitalisierung (15)

Die Antwort

- Ist für Menschen besser lesbar
- Für Maschinen nur schwer lesbar
- Erlaubt Nutzer*innen die Interaktion mit der Schnittstelle
- Kann bei der Entwicklung zum Ausprobieren von Anfragen genutzt werden

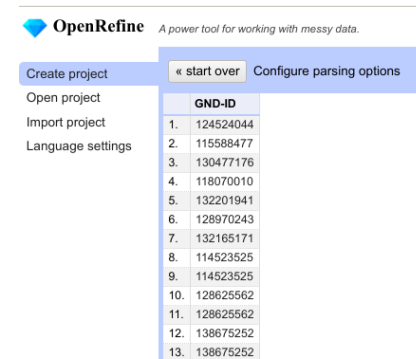
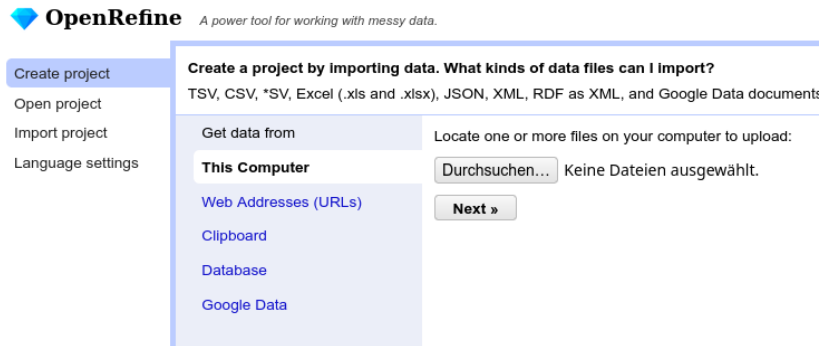
GND Abfragen mit OpenRefine

- OpenRefine Ist ein Tool zur Datenbereinigung und -anreicherung
- OpenRefine kann über ein Plug-In die Daten der GND (Gemeinsame Normdatei der Deutschen Nationalbibliothek) abfragen.
- OpenRefine muss lokal installiert werden. Download unter: <https://openrefine.org/download>

Arbeiten mit OpenRefine

- Daten können in OpenRefine ähnlich wie in Excel oder Calc als Tabelle geladen werden.
- OpenRefine ist für die Arbeit mit Textdaten optimiert, z.B. um Schreibweisen anzugleichen oder Daten mit Normdaten abzugleichen.

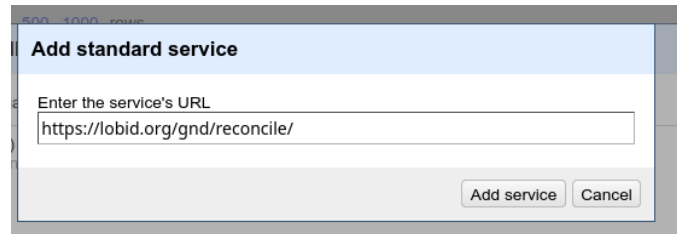
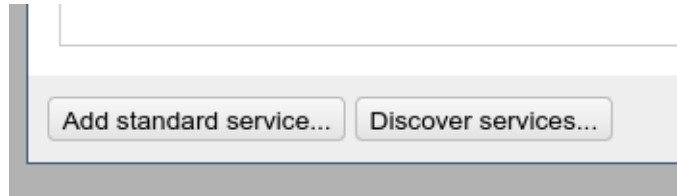
Zum Test die CSV Datei dh_autoren lokal speichern und in OpenRefine öffnen. Download URL:
https://gitlab.ruhr-uni-bochum.de/dh-space/dh-einfuehrungen/-/blob/main/api_json/dh_autoren.csv



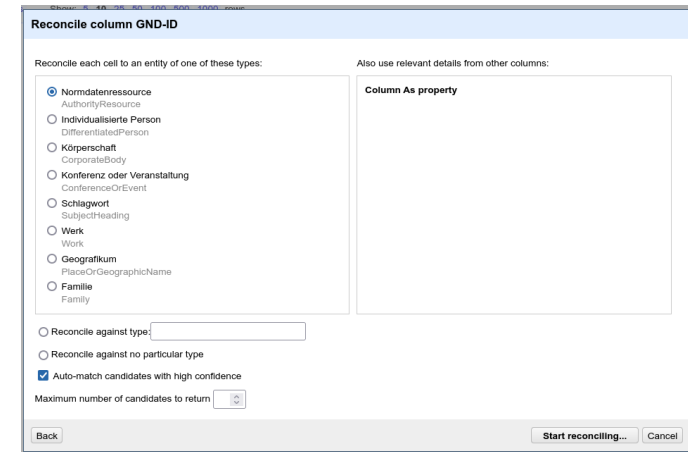
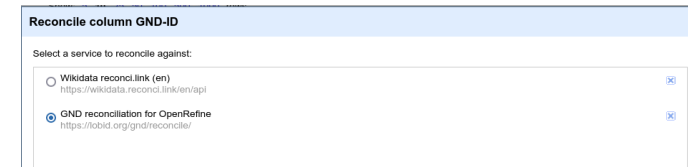
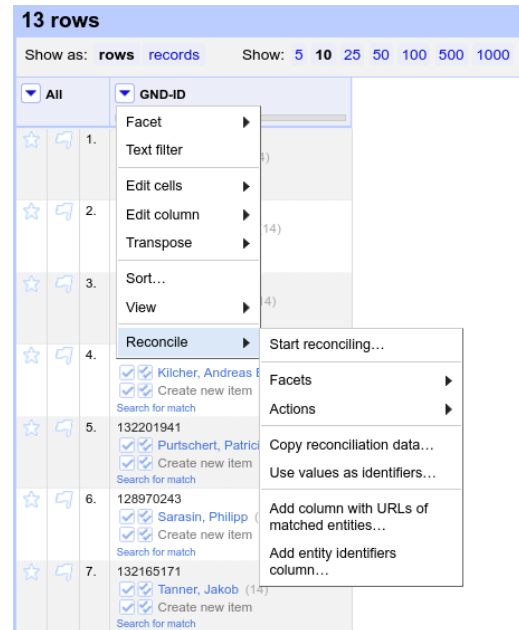
Arbeiten mit OpenRefine

1. Den Datenabgleich mit lobid einrichten:

<https://lobid.org/gnd/reconcile/>



2. Den Abgleich der Spalte GND-ID starten



Daten anreichern mit OpenRefine

In der Spalte mit abgeglichenen Daten „Edit column → Add columns from reconciled values“ auswählen und in dem Auswahlfenster neue Spalteninhalte auswählen

14 rows

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

▼ All ▼ GND-ID ▼ Affiliation

1.	Facet		
2.	Text filter		
3.	Edit cells		
4.	Edit column	Split into several columns...	
5.	Transpose	Join columns...	
6.	Sort...	Add column based on this column...	
7.	View	Add column by fetching URLs...	
8.	Reconcile	Add columns from reconciled values...	
9.	128970243	Rename this column...	
10.	132165171	Remove this column	
11.	114523525	Move column to beginning	
12.	114523525	Move column to end	
13.		Move column left	
14.		Move column right	

Add columns from reconciled column GND-ID

Add property Preview Reset

Beruf

Select an item from the list:

- Beruf oder Beschäftigung professionOrOccupation
- Beruf oder Beschäftigung (Literal) professionOrOccupationAsLiteral
- Berufliche Beziehung professionalRelationship

Benutzungshinweise

Beruf oder Beschäftigung

Beruf oder Beschäftigung (Literal)

Berufliche Beziehung

Besetzung im Musikbereich

Besetzung im Musikbereich (Literal)

Besetzungsstärke

Besitzer

Beteiligte Familie

Beteiligte Körperschaft

Beteiligte Person

Beteiligtes Geografikum

ReconcilerName

GND-ID

Beruf oder Beschäftigung

Historiker

Hochschullehrer

Wissenschaftshistoriker

Arzt

<not reconciled>

130477176 Historiker

118070010 Literaturwissenschaftler

Kulturwissenschaftler

Hochschullehrer

Germanist

132201941 Philosophin

Kulturwissenschaftlerin

Geschlechterforscherin

128970243 Historiker

132165171 Herausgeber

Historiker

114523525 Literaturwissenschaftler

114523525 Literaturwissenschaftler

OK Cancel

Daten anreichern mit OpenRefine

14 rows

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

All		GND-ID	Affiliation	Geburtsort	Geburtsdatum
☆	🔍	1. 124524044 Choose new match	Eidgenössische Technische Hochschule Zürich Choose new match	Zürich Choose new match	1961
☆	🔍	2. 115588477 Choose new match	Deutsche Akademie der Naturforscher Leopoldina Choose new match		1960-01-29
☆	🔍	3. Choose new match	Eidgenössische Technische Hochschule Zürich Choose new match		
☆	🔍	4. 130477176 Choose new match	Universität St. Gallen. School of Humanities and Social Sciences Choose new match		1975
☆	🔍	5. 118070010 Choose new match	Eidgenössische Technische Hochschule Zürich Choose new match	Basel Choose new match	1963
☆	🔍	6. 132201941 Choose new match			1973
☆	🔍	7. 128970243 Choose new match			1956
☆	🔍	8. 132165171 Choose new match	Universität Zürich. Forschungsstelle für Schweizerische Sozial- und Wirtschaftsgeschichte Choose new match		1950
☆	🔍	9. 114523525 Choose new match	Julius-Maximilians-Universität Würzburg. Institut für Deutsche Philologie Choose new match	Frankfurt am Main Choose new match	1961
☆	🔍	10. 114523525 Choose new match	Julius-Maximilians-Universität Würzburg. Institut für Deutsche Philologie Choose new match	Frankfurt am Main Choose new match	1961

Die neuen Spalten werden der Tabelle hinzugefügt und automatisch mit den gefundenen Werten ausgefüllt.

Ressourcen zum Weiterlesen

- **Einführungen und Tutorials zur Nutzung von APIs und OpenRefine:**
 - <http://programminghistorian.org/>
 - <https://fdmlab.landesarchiv-bw.de/workshop/openrefine-einsteiger>
 - <https://lobid.org/resources/api>
 - <https://www.json.org/json-de.html>
 - <https://de.wikipedia.org/wiki/URL-Encoding>

Unterstützungsangebote an der RUB

- Digital Humanities Center der UB
<https://dh.rub.de/>
- Methodenzentrum
<https://methodenzentrum.ruhr-uni-bochum.de/>
- DH Ruhr Projekt
<https://dh-ruhr.ruhr-uni-bochum.de/>

Vernetzungsangebote an der RUB

- **Netzwerk Digital Humanities**
<https://dh-netzwerk.blogs.ruhr-uni-bochum.de/>
- **Mailingliste DH an der RUB**
<https://lists.ruhr-uni-bochum.de/mailman/listinfo/digitalhumanities>
- **Element Chat**
<https://element.rub.de/#/room/#dh:ruhr-uni-bochum.de>