

From String to Thing

Wikidata based query expansion

Bernd Uttenweiler

IT Project Manager ETH Library

SWIB21 3. Dec 2021



Can we use Wikidata to improve our discovery system?

Context and preliminaries

- As part of the Swiss Wikidata Hackdays 2019, we were able to address several questions that we have encountered time and time again.
- In terms of resources, it was a smaller project. But it was all the more important to create concrete benefits for customers.
- Part of ETH team were Petra Maier, Michael Gasser and Bernd Uttenweiler.
- Further supporters were Annette Guignard, Noémie Ammann and Germano Giuliani.
- To say it in advance: We have dealt with the topics "person" and "place". I will focus on "person" here.
- Let's start with different questions that at first glance have nothing to do with each other!



A lot of data is organized by “person”

Here you can see "Hans Conrad Escher" on some pages of the partners of the Swiss project [Metagrid](#).

The diagram illustrates how data about Hans Conrad Escher is organized across various platforms. It features three main screenshots with red arrows indicating connections:

- Burgerbibliothek Bern:** A screenshot of the library's website showing a profile for "Escher, Hans Conrad (1767-1823)". It includes his birth and death dates, profession (Geologe, Ingenieur, Politiker), and a link to documents in the catalog.
- hallerNet:** A screenshot of the hallerNet website showing a profile for "Escher von der Linth, Hans Conrad (1767-1823)". It includes a silhouette, ID number 22695, and a dropdown menu showing 4 persons.
- dodis.ch:** A screenshot of the dodis.ch website showing a profile for "Escher de la Linth, Konrad". It includes his birth and death dates, additional names, and a list of further information sources (HDS, GND, Helveticat, VIAF, BSH, Escher, ETH Library, Sikart, hallerNet, HFLS, Burgerbibliothek).
- Alfred Escher Stiftung:** A screenshot of the Alfred Escher Stiftung website showing a profile for "Hans Conrad Escher von der Linth | 1767-1823". It includes a list of mentions in letters and overview comments.

Red arrows indicate the flow of information and connections between these platforms, showing how data is organized by person across different sources.

Q 1: How to link persons?

- We have a lot of rare resources in our archives and collections.
How can people find them if they do not use our library system?
- Solution: Linking of different systems/sites/databases/portals, even if they are structured by person.
- But where should the link in our Discovery System Primo VE lead to?
- We need something that represents a person and is linkable.



Q 2: Disambiguation: Which is the „Staudinger“ I am looking for?

1



IMAGE

Beschriftung: "H. T. Staudinger". Handschriftliches Exlibris

Staudinger, H. T.

Gerardus Joannes Vossius: (...) De theologia gentili et physiologia Christiana : sive de origine ac progressu idololatriae, ad veterum gesta, ac rerum naturam reductae, deque naturae mirandis, quibus homo adducitur ad Deum (...). Francofurti: sumptibus Casparis Waechtleri typis Iohannis Nicolai Hummii, 1668. Eingangsdatum nicht ermittelbar.
Nicht ermittelbar

 OPEN ACCESS

 [Available Online](#) >

2



BOOK

Catalog der Lepidopteren Europa's und der angrenzenden Länder = Catalogue des lépidoptères d'Europe et des pays limitrophes

Staudinger, Otto 1830-1901 (creator)

1861

 Not available at ETH Zurich
[Check for available services at other swisscovery libraries](#) >

3



BOOK

Point par point : techniques de bandages de compression "point par point" et leurs bases théoriques

Staudinger, Peter (creator)

[2. éd remaniée.] - 2007

 Not available at ETH Zurich
[Check for available services at other swisscovery libraries](#) >

4



MANUSCRIPT

1 Briefkopie an Hermann Staudinger

Ganz, Emil (author)

21. 11. 1858



Q 3: How to find all resources of a person?

Wu Wen-tsin

swisscovery without CDI

Advanced Search

0 selected 3 Results

1 ARCHIVAL MATERIAL
Wen-tsin Wu, Chinesischer Mathematiker, an Heinz Hopf (1894-1971)
01.06.1949
Check for available services >

2 ARCHIVAL MATERIAL
Wen-tsin Wu, Chinesischer Mathematiker, an Heinz Hopf (1894-1971)



Because of „other names“ / „Also Known as“ ...

- ... often different name variants are included in the metadata
- ... and mostly not all name variants are included in the metadata

Wikidata

Wu Wenjun (Q926125)

Chinese mathematician

Wenjun Wu 0003 | Wen-Tsun Wu

[edit](#)

▼ In more languages

Configure

Language	Label	Description	Also known as
English	Wu Wenjun	Chinese mathematician	Wenjun Wu 0003 Wen-Tsun Wu
German	Wu Wenjun	chinesischer Mathematiker	Wu Wen-Tsün Wenjun Wu 0003 <u>Wu Wen-tsin</u>
Estonian	No label defined	Hiina matemaatik	
Spanish	Wu Wenjun	matemático chino	

GND

Person	Wu, Wenjun
Geschlecht	männlich
Andere Namen	Wu, Wen-zun Wu, Wen-tsun Wu Wen-tsün Wu Wenjun Wu, Wen-tsün Wen-tsün Wu



Q 4: How to use graphs in a discovery system?

- A library system like Primo VE offers only few options for flexibly linking records. A system like Wikidata is based on graphs. Can we use Wikidata graphs for our discovery system (Primo VE)?
- Could the user's search query be pre-processed to provide additional offers in Primo VE?
- What does the user search for when entering a search term (semantic interpretation instead of lexical search)?
- Or should there be some sort of advanced search whose fields and selections are then converted to SPARQL queries?
- Or can we identify some search patterns that we can implement in SPARQL queries and offer their results to the user as an option? E.g. Is the user looking for a person or a place?



Approach: offer a list of persons with the search results

When the user enters a search term, we offer him a list of persons, if possible.

The screenshot illustrates the search interface of the ETH Library. At the top, a search bar contains the term "staudinger". Below the search bar, a "Tweak your results" section is visible, featuring a "Linked Data Search" button with a "PLUS" icon. A blue arrow points from the search term "staudinger" to the "Linked Data Search" button. Another blue arrow points from the "PLUS" icon to the "Beschriftung: 'H. T. Staudinger'" label in the search results. The search results are displayed in a list format, showing the following entries:

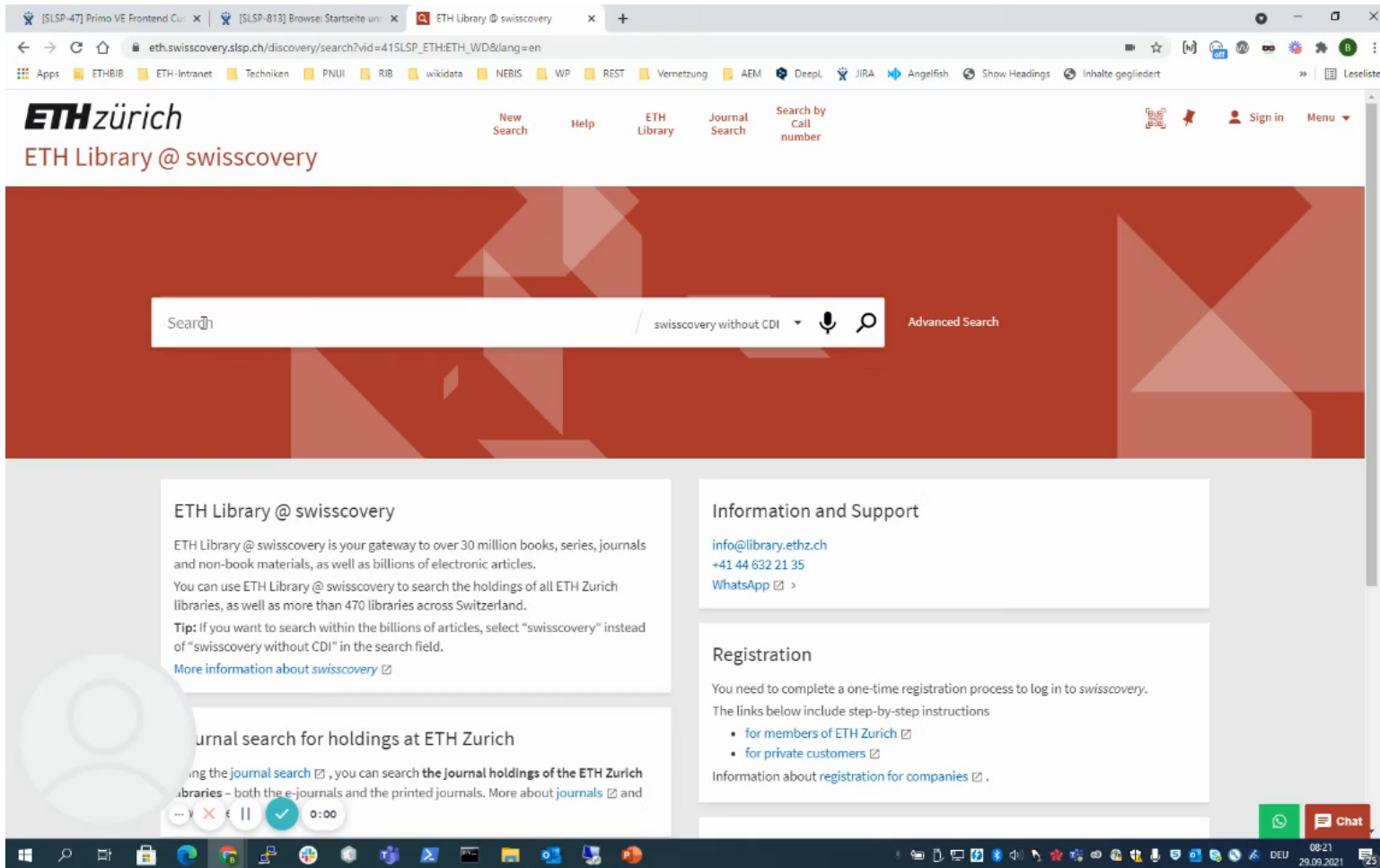
- ARCHIVAL MATERIAL**
Typoskript Kopie von: Eduard Rübel (1876-1960) an Hermann Staudinger (1881-1965)
12.11.1924
Check for available services >
- BOOK**
Der wissenschaftliche Nachlass von Hermann Staudinger(1881-1965) im Deutschen Museum : Katalog
1982
Available at ETH-Bibliothek HS Handschriften (C CONV 2817) >

Below the search results, a "Sorting by persons with image and by relevance in Wikidata" section is visible. This section lists three persons:

- Otto Staudinger**
German lepidopterist (1830-1900)
Birthday: 02.05.1830
Search for this person
- Hermann Staudinger**
German chemist, winner of the 1953 Nobel Prize in Chemistry
Birthday: 23.03.1881
Search for this person
- Ursula Staudinger**
German psychologist

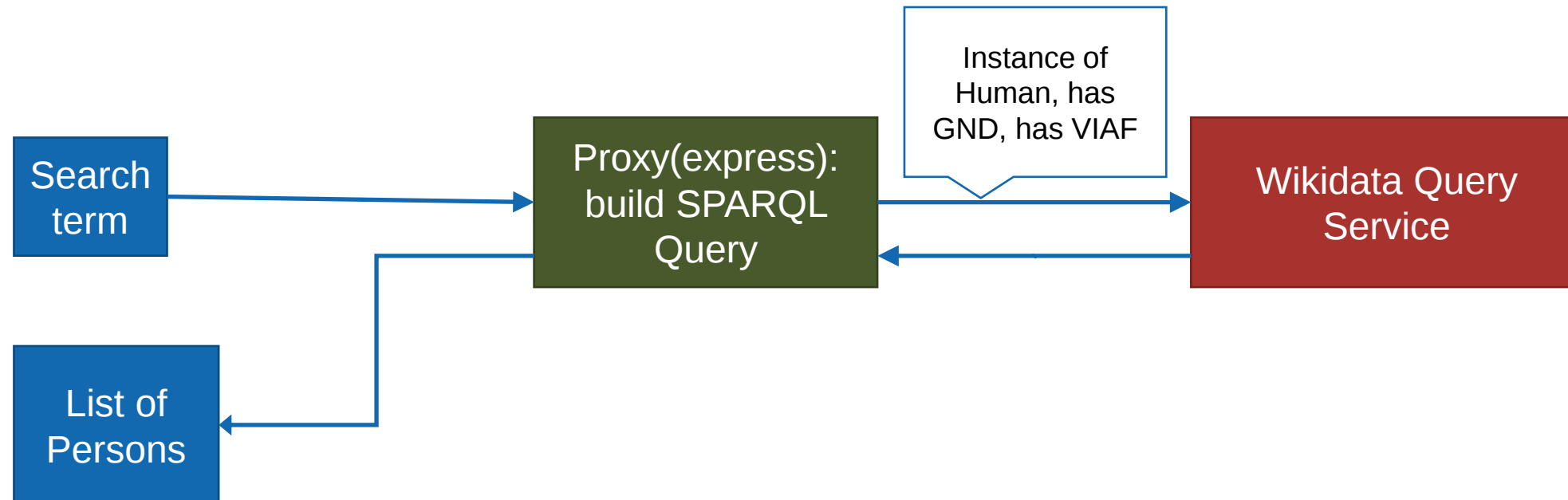


List of persons video



List of persons - How is it done?

- Primo VE Customizing in AngularJS
- The search term (string) is wrapped in a SPARQL query and sent to the Wikidata Query Service (<https://query.wikidata.org/sparql>) via a proxy.



Where to put the person list?

Only a part of our customers use this feature, so it has to be visible, but not distracting or annoying.

The image shows a screenshot of the ETH Library search interface. On the left, a search bar contains the text 'Mann'. Below it, a sidebar titled 'Haben Sie nach einer Person gesucht?' (Have you searched for a person?) is visible. A blue arrow points from this sidebar to a 'Linked Data Search' sidebar on the right. The 'Linked Data Search' sidebar contains two options: 'Do you search for a place (CH)?' and 'Do you search for a person?'. The main search results area on the right shows a list of results. The first result is a book titled 'Katalog der Thomas-Mann-Sammlung Band 4, Alphabetischer Katalog der Thomas-Mann-Sammlung Reprint 2018 - [2018]'. The second result is an image of a person, identified as 'Mann, Mittelholzer, Walter Tschadseeflug, 2.12.1930-23.1.1931'.

Mann

optimieren

Relevanz ▼

☐ 0 ausgewählt SEITE 1 187.258 Ergebnisse

1 BILD [Mann]

Linked Data Search PLUS

Do you search for a place (CH)?

Do you search for a person?

Sort by Relevanz ▼

Show only ▲

Available in the libraries

Online

☐ 0 selected PAGE 1 227,741 Results

1 **BOOK**
Katalog der Thomas-Mann-Sammlung Band 4, Alphabetischer Katalog der Thomas-Mann-Sammlung Reprint 2018 - [2018]
Not available at ETH Zurich
Online access only at other sw

2 **IMAGE**
Mann
Mittelholzer, Walter
Tschadseeflug, 2.12.1930-23.1.1931

Choose person page

- If the user selects a person, then a person page is rendered.
- A person page represents the person (“Thing”), is linkable in both directions, and collects resources belonging to that person.
- An URL represents this page.

Hermann Staudinger

More precise or more results

More information about the person

0 selected 164 Results

1

ARCHIVAL MATERIAL

Hermann Staudinger (1881-1965) an Eduard Rübel (1876-1960)

Staudinger, Hermann

11.12.1924

Check for available services >

2

ARCHIVAL MATERIAL

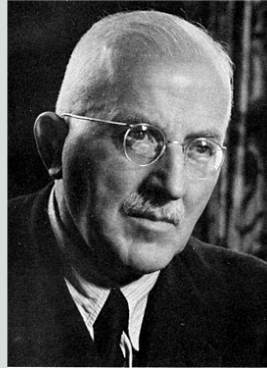
Hermann Staudinger (1881-1965) an Eduard Rübel (1876-1960)

Staudinger, Hermann

10.11.1924

Check for available services >

Hermann Staudinger: More information about the person



Information from Wikidata and the GND

Chemiker

German chemist, winner of the 1953 Nobel Prize in Chemistry

Born: 23. März 1881, Worms

Died: 9. September 1965, Freiburg im Breisgau

1912-1925 Prof. an der Eidgenössischen Technischen Hochschule Zürich, 1926-1951 Prof. für Organische Chemie an der Universität Freiburg; 1953 Nobelpreisträger für Chemie ; Chemiker, Deutschland, Schweiz

Information regarding the license status of embedded media files [Wikimedia Commons](#) .

[Search for this Person](#)

Links in Archive

[ETH Zurich University Archives \(Inventarnummer: CH-001807-7:hs 348\)](#) >

Links from Wikidata

[Wikidata](#) >

[Wikimedia Commons](#) >

[Historisches Lexikon der Schweiz](#) >

[GND \(Gemeinsame Normdatei der Deutschen Nationalbibliothek\)](#) >

[Library of Congress](#) >

Links from Metagrid

Links powered by Metagrid

[Historisches Lexikon der Schweiz](#) >



Choose person page video

The screenshot shows a web browser window displaying the ETH Library search results for the query 'any.contains.staudinger'. The page is titled 'Tweak your results' and shows 1,396 results. The search results are sorted by 'Relevance' and show only 'Available in the libraries' and 'Online' resources. The search results are filtered by 'Resource Type' (Books, Dissertations, Ebooks) and 'Year' (1668 to 2021). The search results are displayed in a list format, showing the name, title, and a small portrait of the person. The first result is 'Otto Staudinger', a German lepidopterist (1830-1900), with a birthday of 02.05.1830. The second result is 'Hermann Staudinger', a German chemist, winner of the 1953 Nobel Prize in Chemistry, with a birthday of 23.03.1881. The third result is 'Ursula Staudinger', a German psychologist, with a birthday of 03.04.1959. Each result has a 'Search for this person' button. The page also includes a 'Linked Data Search' section and a 'Show More' link. The bottom of the page shows a video player interface with a play button and a '0:00 for this person' label.

eth.swisscovery.slsp.ch/discovery/search?query=any.contains.staudinger&tab=discovery_network&search_scope=DiscoveryNetwork&vid=41SLSP_ETH:ETH_WD&lang=en&offset=0#eth-person...

0 selected PAGE 1 1,396 Results Save query

Sorting by persons with image and by relevance in Wikidata

Otto Staudinger
German lepidopterist (1830-1900)
Birthday: 02.05.1830
Search for this person

Hermann Staudinger
German chemist, winner of the 1953 Nobel Prize in Chemistry
Birthday: 23.03.1881
Search for this person

Ursula Staudinger
German psychologist
Birthday: 03.04.1959
0:00 for this person

Linked Data Search PLUS
Do you search for a person?

Sort by Relevance
Show only
Available in the libraries
Online
Online - Open Access
Resource Type
Books (306)
Dissertations (206)
Ebooks (75)
Show More
Year
From 1668 To 2021 Refine
ETH Zürich
ETH Zürich (180)
ETH Zürich (132)



Person page links

Links from Metagrid

Links powered by [Metagrid](#) 

[Historisches Lexikon der Schweiz](#)  >

[Helveticat](#)  >

[Elites suisses au XXe siècle](#)  >

Links from beacon.findbuch

Links from the SeeAlso-Service [pnd-aks](#) , which is based on [BEACON](#) 

[Personenseite bei der Deutschen Digitalen Bibliothek \(156\)](#)  >

[Deutsche Biographie](#)  >

Students or doctoral students (from Wikidata)



[Leopold Ružička \(Croatian scientist \(1887-1976\), Born: 1887-09-13\)](#)

Teachers (from Wikidata)



[Daniel Vorländer \(German chemist, Born: 1867-06-11\)](#)

Links in Archive

[ETH Zurich University Archives \(Inventarnummer: CH-001807-7:Hs 348\)](#)  >

Researcher profiles via IDs from Wikidata

[ORCID record](#)  >

[Publons profile](#)  >

[Google Scholar profile](#)  >

[ResearchGate profile](#)  >

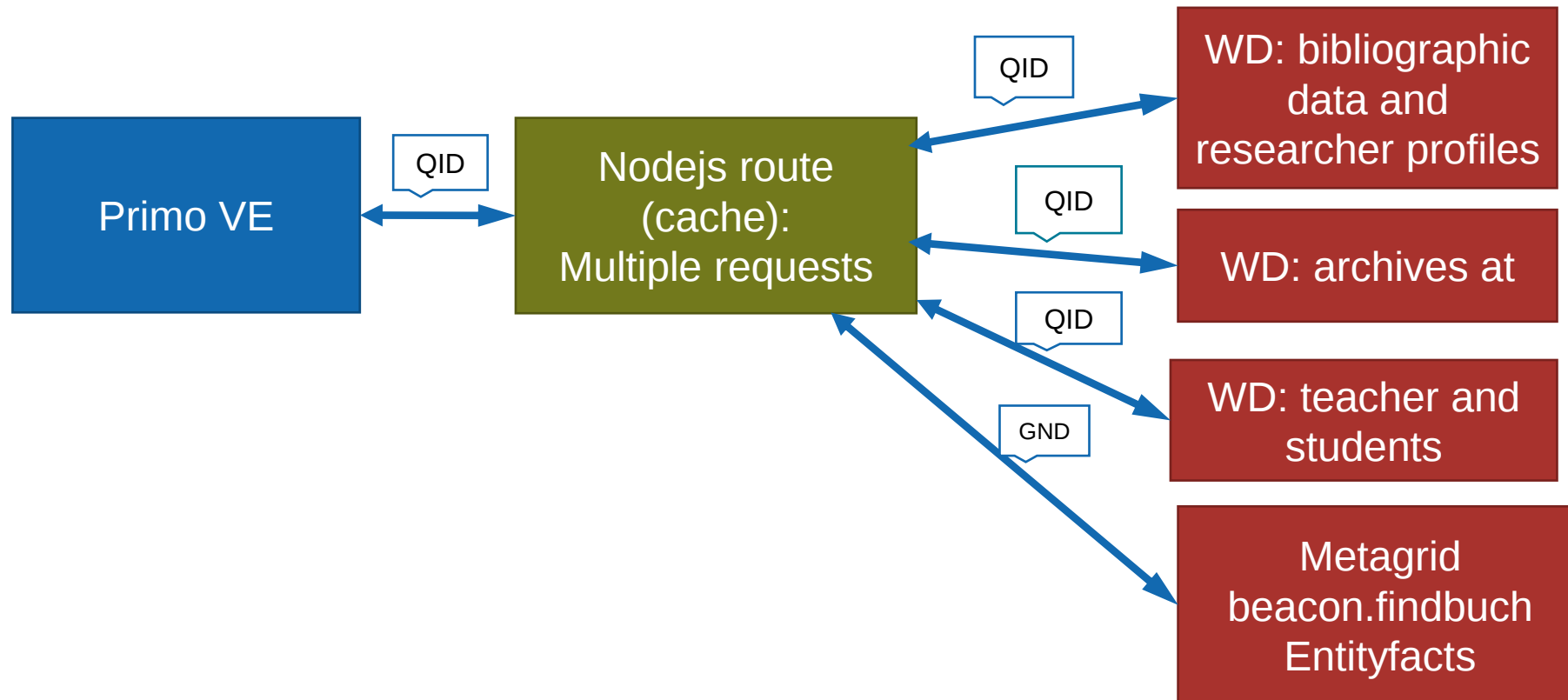
[Mendeley profile](#)  >

[Dimensions profile](#)  >



Person page - How is it done?

- The collection of all information is offloaded to our express/nodejs system.



Precision vs Recall

- Here, we are in the area of tension between "important things are missing" and "there is too much wrong". This problem is also described by the term pair "Precision vs Recall".
- While "Precision" refers to the percentage of the actual result that is relevant, "Recall" refers to the percentage of the total relevant results that appear in the results list.

Wenjun Wu: More precise or more results?

You see **8 results** of the search for a unique identification number (GND) of the person.

Attention: You may not see all results because the identification number (GND) of the person is not assigned to all resources.

Select one of the search options:

9 results

- Search for the name , restricted by the identification number (GND) of the person OR the year of birth.
- You may not see all results for this person because the year of birth or identification number (GND) is not assigned to all resources.

24 results

- Search by name only
- You may also see wrong results if several people have the same name.

31 results

- Search for the name and name variants (from Wikidata)
- You may also see wrong results if several people have one of the name variants.



Precision vs Recall: Options

- Search for GND ID
- Search by name (from Wikidata, first name and last name) AND (GND ID OR date of birth)
- Search by name (from Wikidata, first name and last name)
- Search for name variants (from Wikidata)



Precision vs Recall - user survey

- This is not without complexity. Should we burden the users with this?
- This feature was welcomed because it makes transparent a situation that is there anyway.
- Which of the search alternatives is useful depends on the data and is different for different persons.



Person cards and links to person page video

eth.swisscovery.slsp.ch/discovery/fulldisplay?docid=alma991141182249705501&context=L&vid=41SLSP_ETH.ETH_WD&lang=en&search_scope=DiscoveryNetwork&adaptor=Local%20sea...

3 versions found. See all versions >

Top
Loan and request options
Details
Links
Send to

Loan and request options

OTHER LOAN AND REQUEST OPTIONS

Region Zentralschweiz
Available in institution

Details

Title: Zur Interpretation des modernen Dramas : Brecht, Dürrenmatt, Frisch
Attribution: hrsg. v. Rolf Geissler unter Mitarbeit v. Therese Poser u. Wilhelm Ziskoven
Publication: Frankfurt am Main : M. Diesterweg
Edition/Date: 10. Aufl. - 1981
Format: 144 S. ; 23 cm
Language: German
Contributor: Geissler, Rolf >
Poser, Therese >
Ziskoven, Wilhelm >

Subject: Dürrenmatt, Friedrich 1921-1990 >
Brecht, Bertolt 1898-1956 >
Dürrenmatt, Friedrich >
Frisch, Max 1911-1991 >
Drama >

Display in the GND

Brecht, Bertolt 1898-1956 Search for this Person
Frisch, Max 1911-1991 Search for this Person
Dürrenmatt, Friedrich 1921-1990 Search for this Person
Drama

Links



Questions answered?

- Q1: The person page represents the person and is linkable in both directions (QID is part of URL).
- Q2: The user can choose between persons with the same name.
- Q3: With the person's metadata, controlled searches for the person's resources can be made in several ways.
- Q4: We test the search term to see if it could mean a person or a place and provide results for it. This could also be a possibility for other types, e.g. "buildings".
- Q4: We use wikidata graphs to create relationships between teacher and student in our discovery system.



Links

- https://www.wikidata.org/wiki/User:Beat_Estermann/WD_Hackdays_2019
- Gitlab Repo 2019: <https://gitlab.com/ethlibrary/wd-hackdays-2019/graph-based-query-expansion/>
- Person page ETH Library @ swisscovery:
https://eth.swisscovery.slsp.ch/discovery/search?search_scope=DiscoveryNetwork&tab=discovery_network&query=any,contains,%5Bwd%2Fperson%5DQ926125&vid=41SLSP_ETH:ETH&lang=en
- Person list in our test view:
https://eth.swisscovery.slsp.ch/discovery/search?query=any,contains,staudinger&tab=discovery_network&search_scope=DiscoveryNetwork&vid=41SLSP_ETH:ETH_WD&lang=en&offset=0

Suggestions of people after the search input will come to the productive view in the first Q1 of 2022.



Thank you!
Questions?

