

Praxislabor Digitale Geisteswissenschaften

Einführung in Wikidata

oder besser gesagt:
*Was ist Wikidata
und wie kann ich es für
meine Forschung nutzen?*



Julia Beck
j.beck@ub.uni-frankfurt.de

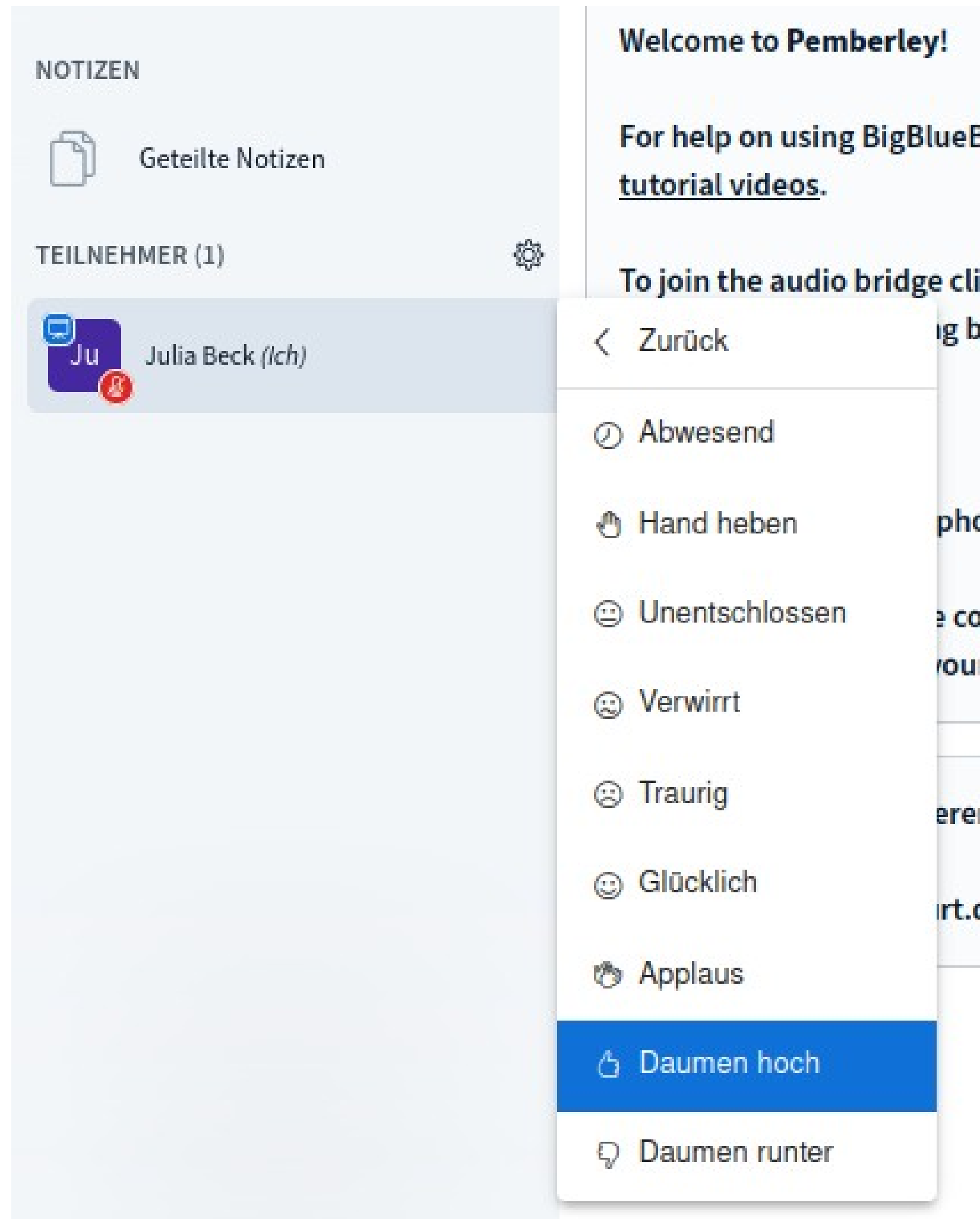
Kursinhalt „Einführung in Wikidata“

Heute:





- Bisherige Erfahrungen?
- Aufbau
- grundlegende Konzepte
- selbst Änderungen durchführen
- Forschen mit dem *Query Service*

Nächste Woche:

Übungen



Bisherige Erfahrungen mit Wikidata?

-  Daumen runter vor dieser Veranstaltung zwar mal gehört, aber nie wirklich mit beschäftigt
-  Verwirrt bisschen was probiert / angesehen, aber nicht weit gekommen
-  Daumen hoch nutze ich ab und an
-  Abwesend lange her, ich will mich updaten

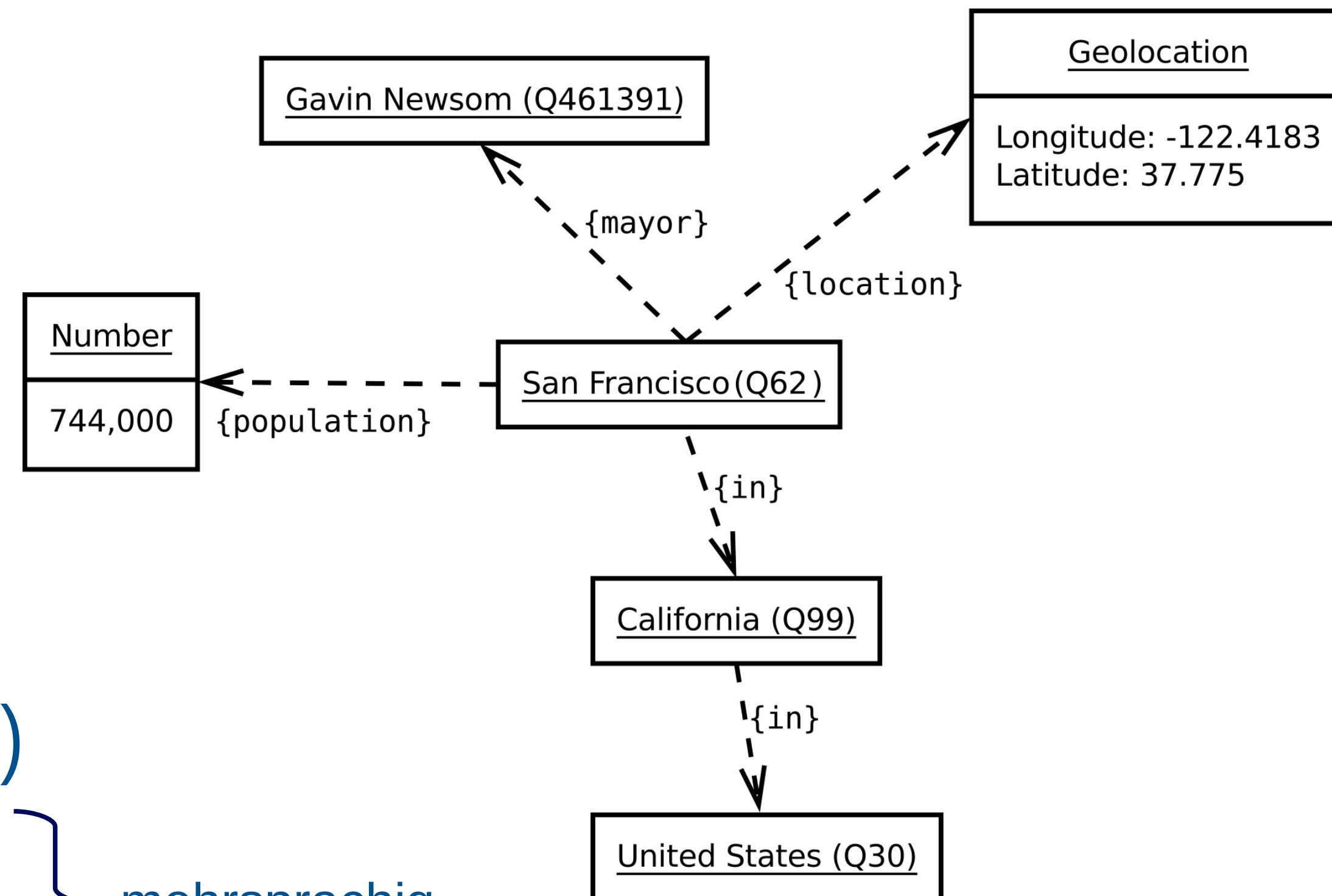
Was ist Wikidata eigentlich?

“Wikidata ist eine freie Wissensdatenbank, die von Menschen und Maschinen gelesen und bearbeitet werden kann.”

- seit 2012 Teil des Wikiversums (Wikipedia, Wiktionary, Wikisource, ...)
- sehr große Datenbank (100.690.804 Items, Stand 07.12.2022), die ständig wächst
- enthält strukturierte Daten
- jede*r kann mitmachen (5.753.528 registrierte User, Stand 07.12.2022); die Daten sind frei verfügbar (CC0 1.0 Public Domain) und können modifiziert, kopiert und weitergegeben werden
- so umfangreich, dass wir nur an der Oberfläche kratzen können (Hilfe zur Selbsthilfe)

Wikidata Items

- bilden *Dinge* menschlichen Wissens ab:
Begriffe / Konzepte / Objekte / Personen
/ Ereignisse / Lexeme (seit 2018) / ...
 - sind *relevant* (~ Link ins übrige Wikiversum)
 - sind *eindeutig identifizierbar*
 - sind miteinander *verknüpft*
-
- haben einen Identifier (Q-Number / L-Number)
 - haben mind. eine Bezeichnung (*Label*)
 - haben mind. eine Beschreibung (*Description*)
 - haben eine beliebige Anzahl Aussagen über das Item (*Statements*)
 - können weitere Identifier haben



Wikidata Item Beispiel



- Person: Jane Austen
- Identifier **Q36322**
- Bezeichnung: „Jane Austen“@en
- Beschreibung: „English novelist (1775–1817)“@en
- Aussagen (Subjekt – Prädikat – Objekt):
 - Jane Austen ist ein Mensch.
 - Jane Austen ist weiblich.
 - Jane Austens Geburtsdatum ist am 16.12.1775.
 - ...
- Weitere ID: **pnd118505173**

Das Konzept hinter Aussagen

Wikidata arbeitet mit RDF (Resource Description Framework) mit einer Graphdatenbank

	Relationale Datenbank	Graphdatenbank
Kategorie	Mengen (Tabellen mit Spalten und Zeilen)	Pfade (Graphen mit Knoten und Kanten), NoSQL
Besonderheit	Relationen von Tabellenspalten	Speichern von Relationen als Daten
Abfragesprache	SQL	SPARQL
Performance	Schnell bei großen Mengen	Schnell bei komplexen Anfragen

Das Konzept hinter Aussagen

RDF (Resource Description Framework) ist ein konzeptuelles Datenmodell.

Subjekt $\xrightarrow{\text{Prädikat}}$ Objekt.

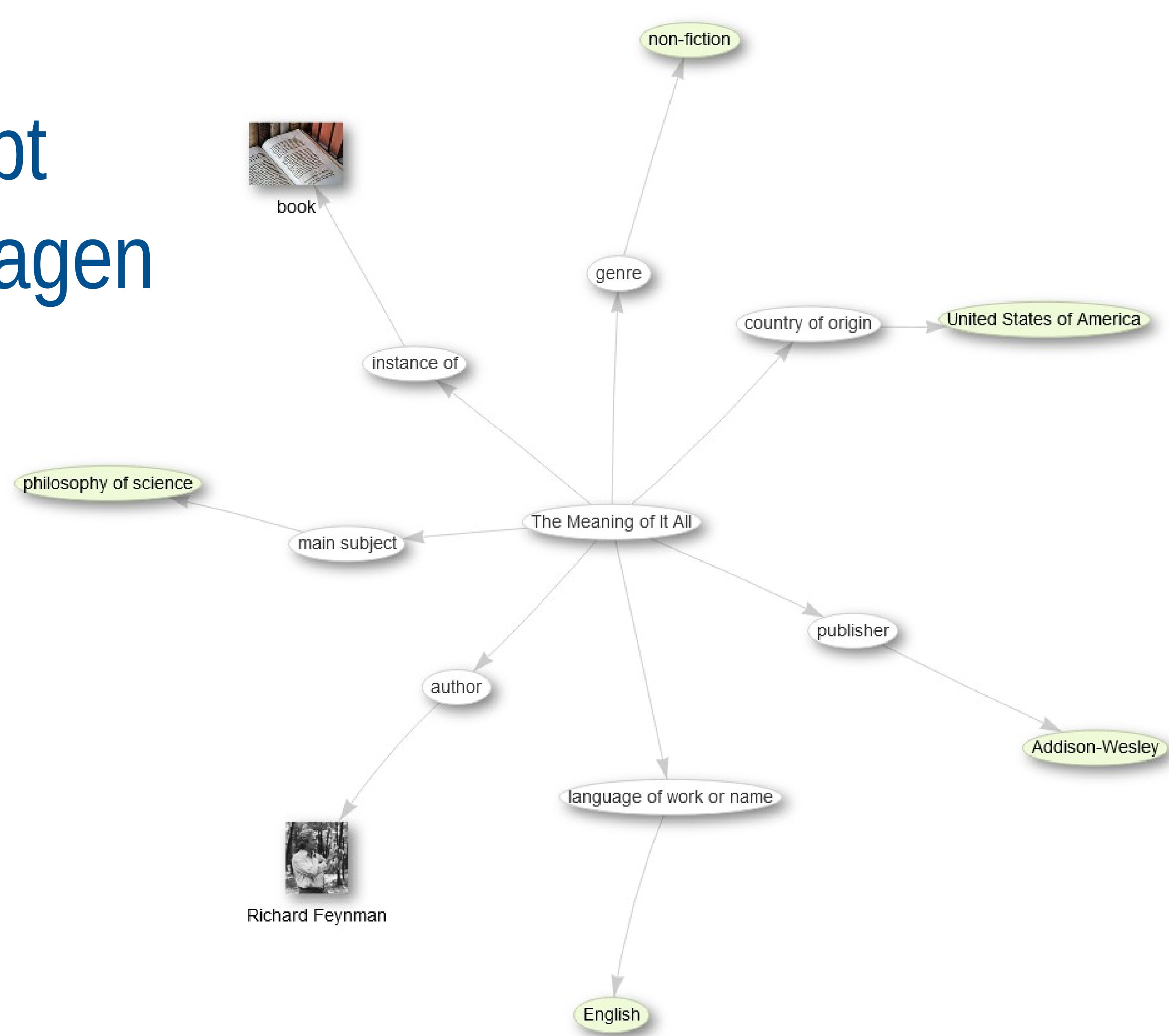
Mit diesem sogenannten *Triple* kann man logische Aussagen formulieren.

Subjekte sind dabei immer Ressourcen, die per URI (Uniform Resource Identifier), identifizierbar sind. Objekte können ebenfalls Ressourcen sein oder Literale (in Wikidata auch „quantitative Werte“ genannt).

Aussagen können qualifiziert werden, indem man Aussagen über eine Relation macht.

Subjekt	Prädikat	Objekt	Qualifizierer
Jane Austen	ging an die	Reading Abbey Girls' School.	<i>Und zwar von 1785 bis 1786.</i>

Das Konzept hinter Aussagen



label — **Douglas Adams** (Q42) — **item identifier**

description — English writer and humorist
Douglas Noël Adams | Douglas Noel Adams — **aliases**
▶ In more languages

Statements

property — **educated at** — **value**

rank — **St John's College** — **qualifiers**

end time	1974
academic major	English literature
academic degree	Bachelor of Arts
start time	1971

statement group — **opened references**

▼ 2 references

stated in	Encyclopædia Britannica Online
reference URL	http://www.nndb.com/people/731/000023662/
original language of work	English
retrieved	7 December 2013
publisher	NNDB
title	Douglas Adams (English)

+ add reference

collapsed reference

▲ ● ▼ **Brentwood School**

end time	1970
start time	1959

▶ 0 references

+ add (statement)

Wikidata Item Beispiel

Wikidata Item anlegen

“Oh, *XYZ* existiert ja noch gar nicht.“

“Ich habe einen Fehler gefunden.“

- Vorherige Prüfung, ob Item nicht bereits existiert.
- Es gibt Richtlinien, wie **Bezeichner** und **Beschreibungen** auszusehen haben.
- Beim ersten Mal am besten nur einzelne Statements bearbeiten oder neu anlegen, am Beispiel bereits bestehender Einträge.
- Referenzen suchen, die neu angelegt Aussage bestätigen.

Wikidata Masseneinträge

“Ich habe viele tolle Forschungsergebnisse, die ich mit der Welt teilen möchte und alle noch nicht in Wikidata sind. Muss ich die wirklich alle per Hand eingeben?“

- Für Masseneinträge kann das Tool **QuickStatements** genutzt werden.

Paste a tab/newline-delimited command sequence from the original QuickStatements here.
 You can pass such commands as URL parameters by appending "#v1=COMMANDS" to the URL.
 For convenience, you can replace tabs by "|" and newlines by "¶".
 Note: MERGE command does NOT work yet.
 You can remove specific statements by prefixing a line with "-"
 Quantities with error can be entered as 1.2±0.3 (for 1.2±0.3)

```
CREATE
LASTP31 Q5 S143 Q24731821
LASTLen "Giovanni Francesco Rossi"
LASTP21 Q6581097 S143 Q24731821 S813
+2017-10-04T00:00:00Z/11
LASTP1472 "Giovanni Francesco Rossi" S143 Q24731821
S813 +2017-10-04T00:00:00Z/11
LASTP19 Q220 S143 Q24731821 S813
+2017-10-04T00:00:00Z/11
LASTP20 Q220 S143 Q24731821 S813
+2017-10-04T00:00:00Z/11
LASTP373 "Giovanni Francesco Rossi" S143 Q24731821
S813 +2017-10-04T00:00:00Z/11
```

Close Import

Wikidata Versionierung / Logs

Diff selection: Mark the radio buttons of the revisions to compare and hit enter or the button at the bottom.

Legend: **(cur)** = difference with latest revision, **(prev)** = difference with preceding revision, **m** = minor edit.

(latest | earliest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Read

View history

Search Wikidata

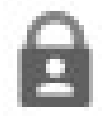


Compare selected revisions

- (cur | prev) 17:04, 19 November 2022 Magnus Manske (talk | contribs) . . (289,514 bytes) **(+365)** . . (Created claim: *Levadia Library ID (P11185): 59de6bd221c0c8f86d488fe9, batch #105389*)
- (cur | prev) 00:39, 18 November 2022 Qesotiotyo (talk | contribs) . . (289,149 bytes) **(+359)** . . (Created claim: *Famous Birthdays ID (P11194): people/jane-austen, batch #105291*) (Tag: quickstatements [2.0])
- (cur | prev) 11:28, 16 November 2022 VojtěchDostálBot (talk | contribs) . . (288,790 bytes) **(+1,503)** . . (Created claim: *field of work (P101): literature (Q8242), field of work from NKC, see details*)
- (cur | prev) 10:02, 16 November 2022 Magnus Manske (talk | contribs) . . (287,287 bytes) **(+345)** . . (Created claim: *MMB ID (P11156): 17133, batch #105198*) (Tag: quickstatements [2.0])
- (cur | prev) 09:44, 16 November 2022 Magnus Manske (talk | contribs) . . (286,942 bytes) **(+345)** . . (Created claim: *Municipal Library of Trikala ID (P11155): 77671, batch #105197*) (Tag: quickstatements [2.0])
- (cur | prev) 10:33, 3 November 2022 Theklan (talk | contribs) . . (286,597 bytes) **(+451)** . . (Added qualifier: *point in time (P585): 31 October 2022, #quickstatements; #temporary_batch_166; #temporary_batch_167*)
- (cur | prev) 10:33, 3 November 2022 Theklan (talk | contribs) . . (286,146 bytes) **(+413)** . . (Created claim: *on focus list of Wikimedia project (P5008): Wikipedia:Vital articles/Level/4 (Q6173448, quickstatements [2.0])*)
- (cur | prev) 02:07, 24 October 2022 MatSuBot (talk | contribs) . . (285,733 bytes) **(+70)** . . (Changed label, description and/or aliases in kw: *add missing labels*)
- (cur | prev) 20:45, 19 October 2022 Aqchampion (talk | contribs) . . (285,663 bytes) **(+93)** . . (Added link to [kwwiki]: *Jane Austen*) (Tag: Sitelink Change from Connected Wiki)
- (cur | prev) 17:36, 10 October 2022 JhsBot (talk | contribs) . . (285,570 bytes) **(+105)** . . (Added link to [igwikiquote]: *Jane Austen*)
- (cur | prev) 17:56, 8 October 2022 Henry Merrivale (talk | contribs) . . (285,465 bytes) **(+344)** . . (Created claim: *Literary Encyclopedia ID (P9001): 1-7064, #quickstatements; #temporary_batch_166*)
- (cur | prev) 21:01, 5 October 2022 VojtěchDostálBot (talk | contribs) . . (285,121 bytes) **(+371)** . . (Added qualifier: *reason for deprecated rank (P2241): source known to be unreliable (Q22979 (details))*) (Tag: WikibaseJS-cli)
- (cur | prev) 14:19, 5 October 2022 Spinster300 (talk | contribs) . . (284,750 bytes) **(+1,423)** . . (Created claim: *National Library of Ireland ID (P10227): vtls000071750, Added with moreIdentifiers*)
- (cur | prev) 02:13, 5 October 2022 VojtěchDostálBot (talk | contribs) . . (283,327 bytes) **(+5)** . . (Updated Item: *Deprecating ethnicity statements sourced only from Wikimedia projects, see Wikidata:Bot_requests#request_to_depreciated_ethnic_group_only_sourced_with_P143_(2021-10-23) (details)*) (Tag: WikibaseJS-cli)
- (cur | prev) 15:29, 16 September 2022 Matlin (talk | contribs) . . (283,322 bytes) **(+343)** . . (Created claim: *Fandango person ID (P8125): 29301, batch #98384*) (Tag: quickstatements [2.0])
- (cur | prev) 14:41, 13 September 2022 RoastPig (talk | contribs) . . (282,979 bytes) **(+113)** . . (Added link to [srwikiquote]: *Џејн Остун*) (Tag: Wikidata User Interface)
- (cur | prev) 13:53, 11 September 2022 Սահակ (talk | contribs) . . (282,866 bytes) **(-1,140)** . . (Changed claim: *date of birth (P569): 16 December 1775*) (Tag: Wikidata User Interface)
- (cur | prev) 13:49, 11 September 2022 Սահակ (talk | contribs) . . (284,006 bytes) **(0)** . . (Changed Armenian description: *անգլիացի վիպասան*) (Tag: Wikidata User Interface)

Wikidata Quality

“Das ist ja schön und gut, aber wenn jeder jederzeit bearbeiten kann, ist meine Forschung dann überhaupt valide/aussagekräftig/...?”

- Vorsicht bei globalen Aussagen: Wenn man in Wikidata sucht, findet man nicht alle Autorinnen auf der Welt, sondern alle Autorinnen, die in Wikidata verzeichnet sind.
- Es gibt auch „kein Wert“ oder „unbekannter Wert“.
- Items können vor Vandalismus geschützt sein. (siehe rechte obere Ecke )

Es gibt ganze Studien zur Qualität von Wikidata, z.B.:

<https://doi.org/10.1016/j.websem.2021.100679>

- Die meisten kommen zu dem Schluss, dass die kollektive Intelligenz die Qualität hochhält und stetig verbessert.

Wikidata Projekte

- Arbeit an Vervollständigung von Wikidata (z.B. für Gebäude/Orte oder Themengebiete)
- Einigung auf gemeinsame Standards und Systematiken

Beispiele:

Wikidata Projekt Performing Arts

https://www.wikidata.org/wiki/Wikidata:WikiProject_Performing_arts

Liste aller Projekte:

<https://www.wikidata.org/wiki/Wikidata:WikiProjects>

Wikidata Query Service

Abfragen werden in Wikidata mittels der Abfragesprache **SPARQL** gemacht und sehen erstmal etwas kryptisch aus, wenn man nicht bereits andere Abfragesprachen wie z.B. SQL kennt.

```
SELECT ?item ?itemLabel
WHERE
{
  ?item wdt:P31 wd:Q5.
  ?item wdt:P21 wd:Q6581072.
  ?item wdt:P106 wd:Q36180.
  SERVICE wikibase:label { bd:serviceParam
    wikibase:language "[AUTO_LANGUAGE],en". }
}
```


Wikidata Query Service

<https://query.wikidata.org/>

Wikidata bietet tolle Tutorials zum Lernen von SPARQL:

https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial

https://www.wikidata.org/wiki/File:Querying_Wikidata_with_SPARQL_for_Absolute_Beginners.webm

Aber auch der **QueryBuilder** bietet eine schöne Möglichkeit Abfragen zu bauen, ohne Erfahrung in SPARQL.

Beispiele:

https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/Building_a_query

Wikidata Tools

- Visualisierungstools: https://www.wikidata.org/wiki/Wikidata:Tools/Visualize_data
z.B. Crotos
- Lexikografische Daten:
https://www.wikidata.org/wiki/Wikidata:Tools/Lexicographical_data
- Allgemeine Übersicht: <https://www.wikidata.org/wiki/Wikidata:Tools>

Wikidata One Pager als kleine Übersicht:

<https://upload.wikimedia.org/wikipedia/commons/8/8d/Wikidata-in-brief-1.0.pdf>