## Wikidata-Datenmodell

Wie bereits im Kapitel 4 gesagt wurde, ist Wikidata eine multilinguale Graph-Datenbank, in der sowohl der Inhalt als auch Struktur des Graphen bearbeitet werden können. Es wird einem streng semantischen Prinzip durch Properties gefolgt. Klassen bzw. Instanzen können dabei frei erzeugt werden. Da die Daten in Wikidata als RDF-Tripeln gespeichert werden, lässt sich das Datenmodell von Wikidata auf einfacher Ebene gut mit RDF abbilden: mit Item (rdf:subject), Property (rdf:predicate) und Wert (rdf:object) (vgl. [RDF und Wikidata](https://www.wikidata.org/wiki/Wikidata:Relation_between_properties_in_RDF_and_in_Wikidata)). Im Rahmen dieser Arbeit wurde zwei Typen von Items verwendet: Person mit biographischer Datensätze und Weg zu Missionsorten.

Da sich die Informationen über die Personen nicht als homogen zeigten, wurde sämtliches verfügbares Material im Datenmodell aufgenommen. Die Tabelle 5.2 zeig die Modellierung aller denkbaren Relationen zu einer Person. Wikidata-Elementen wurden mithilfe der Wikidata-Eigenschaft P31 (Instanz bzw. Klasse) ein Typ Q5 (ein Mensch) zugewiesen. Es gibt die typische Information für eine Person wie Geschlecht, Verwandtschaften, Ehrpartner:in, die Ausbildung, wichtige Datenangaben. Ein ergänzendes Element des Datenmodell ist die Verwendung von Qualifikatoren. Sie ermöglichen es, Relationen mit zusätzlichen Kontextinformationen anzureichern: mit Datumsangaben (pq:580, pq:582 und pq:585) oder mit Quellenangaben zu Primärdigitalisat (pq:584). Um die in dieser Arbeit erzeugte Daten von anderen abzugrenzen, wurde ein Q-ID *WikiProjekt Digitalisat2Graph* erstellt und die Daten diesem zugeordnet.

|  |  |  |
| --- | --- | --- |
| Wikidata-Property | Beschreibung | Beispiel |
| rdfs:label (@de) | *Label* des Items (Deutsch, Literal) | Johann Christian Breutel (wd:Q21415674) |
| wdt:P31 | *Instanz* von (verlinkt) | Mensch (wd:Q5) |
| wdt:P19 | *Geburtsort* (verlinkt) | Weißenburg in Bayern (wd:Q44500) |
| wdt:P20 | *Sterbeort* (verlinkt) | Berthelsdorf (wd:Q502553) |
| wdt:P21 | *Geschlecht* (verlinkt) | männlich (wd:Q6581097) |
| wdt:P22 | *Vater* (verlinkt) | Johann Christian Breutel (wd:Q21415674) |
| wdt:P25 | *Mutter* (verlinkt) | Sophie Breutel (wd:Q125843987) |
| wdt:P26 | *Ehrpartner(in)* (verlinkt) | Sophie Breutel (wd:Q125843987) |
| wdt:P40 | *Kind(er)* (verlinkt) | Johann Christian Breutel (wd:Q21415674) |
| wdt:P69 | *besuchte Bildungseinrichtung* | Universität Leipzig (wd:Q154804) |
| wdt:P106 | *Tätigkeit* (verlinkt) | Botaniker (wd:Q2374149) |
| wdt:P463 | *Mitglied von* (verlinkt) | Herrnhuter Brüdergemeine (wd:Q159318) |
| wdt:P551 | *Wohnsitz* (verlinkt) | Berthelsdorf (wd:Q502553) |
| wdt:P569 | *Geburtsdatum* (Literal), ISO: YYYY-MM-DD/ YYYY-MM/ YYYY | 1788-01-21 |
| wdt:P570 | *Sterbedatum* (Literal), ISO: YYYY-MM-DD/ YYYY-MM/ YYYY | 1875-02-18 |
| wdt:P793 | *Schlüsselereignis* (verlinkt) | Weg zum Missionsort (Südafrika) (wd:Q135874728)/ Konfirmation (wd:Q214802) |
| wdt:P937 | *Wirkungsort* (verlinkt) | Südafrika (wd:Q258) |
| wdt:P1066 | *Schüler von* (verlinkt) | Johann Jürgen Stein (wd:Q123456933) |
| wdt:P1971 | *Anzahl der Kind(er),* (Literal) | 7 |
| wdt:P3373 | *Geschwister* (verlinkt) | Johann Christian Breutel (wd:Q21415674) |
| wdt:1344 | *Teilnahme an* (verlinkt) | Weg zum Missonsort (Südafrika): Johann G. und Johanna D. Bonatz (wd:Q135874728) |
| wdt:P5008 | *auf der Arbeitsliste des Wikimedia-Projektes* (verlinkt) | Digitalisat2Graph (wd:Q136079887) |
| pq:P580 | *Startzeitpunkt (Literal), ISO: YYYY-MM-DD/YYYY-MM/YYYY* | 1822-06-08 |
| pq:P582 | *Endzeitpunkt (Literal), ISO: YYYY-MM-DD/YYYY-MM/YYYY* | 1822-06-08 |
| pq:P585 | *Zeitpunkt/Stand (Literal), ISO: YYYY-MM-DD/YYYY-MM/YYYY* | 1822-06-08 |
| pq:P854 | *URL der Fundstelle (URL)* | URL |

Tab. 5.2: Datenmodell für biographische Datensätzen in Wikidata

Der zweite Typ des Datenmodells umfasst die Wege zu Missionsorten. In den untersuchten Lebensläufen wurden neben biographischen Angaben auch die Reisen von Missionaren beschrieben: sowohl die Anreise zu einer Missionsstelle als auch Reisen zwischen verschiedenen Missionsorten (Elim, Silo usw.). Letztere waren jedoch nicht konsistent dokumentiert und die Beschreibungen dieser Reisen waren häufig unterbrochen. Dies machte die chronologische Rekonstruktion nicht möglich. Daher wurde entschieden, sich ausschließlich auf die Wege bis zur jeweiligen Missionsstelle zu konzentrieren. Dabei wurden drei solche Zielrichtungen identifiziert: Westindien, Labrador und Südafrika.

In den Lebensbeschreibungen wurden jeweils die Route sowie die Zusammensetzung der Mitreisenden festgehalten. Diese Informationen wurden zuerst in Tripeln-Form modelliert. Anschließend wurden die einzelnen Reisen in Wikidata je Reise nach Rote und Mitreisenden gruppiert und als eingeständige Einträge in Wikidata angelegt.

Das Modell erlaubt einige Eigenschaften zu beschreiben (Tabelle 5.3). Über wdt:P31 (Instanz von) wurde die Klasse *Route* definiert. Für jede Route wurden der geographische Start- (wdt:P1427) und Endpunkt (wdt:1444) erfasst und bei Bedarf lassen sich auch Zwischenstationen (wdt:P2825) optional angeben. Besonders wichtig ist hier der Einsatz von Qualifikatoren. Sie präzisieren die Aussagen zu Route, durch die Modellierung eines Datums (pq:P585). Die Quellennachweise wurden als Referenzen hinterlegt und über P854 verknüpft.

|  |  |  |
| --- | --- | --- |
| Wikidata-Property | Beschreibung | Beispiel |
| rdfs:label (@de) | *Label* des Items (Deutsch, Literal) | Weg zum Missonsort (Südafrika) (Q135874728) |
| wdt:P31 | *Instanz* von (verlinkt) | Route (Q1322323) |
| wdt:1427 | *geografischer Startpunkt* (verlinkt) | Herrnhut (wd:Q165140) |
| wdt:P1444 | *geografischer Endpunkt* (verlinkt) | Genadendal (wd:Q1501295) |
| wdt:P2825 | *Zwischenpunkt einer Reise* (verlinkt) | Herrnhut (wd:Q165140) |
| pq:P585 | *Zeitpunkt/Stand* (Literal), ISO: YYYY-MM-DD/YYYY-MM/YYYY | 1822-06-08 |
| pq:P854 | *URL der Fundstelle* (URL) | URL |
| wdt:P5008 | *auf der Arbeitsliste des Wikimedia-Projektes* (verlinkt) | Digitalisat2Graph (wd:Q136079887) |

Tab. 5.3 Datenmodell für Routen in Wikidata

Die Verknüpfung zwischen Personen und Routen wurde auf Ebene der Personeninstanz modelliert. Die Zuordnung erfolgt mittels Property P1344 (*Teilnahme an*). Als P-Wert wurde die Q-ID des jeweiligen Roteobjekts hinterlegt. Die Modellierung der Missionswege wurde als besonders relevant erachtet, da sie ermöglichen, den Verlauf der Reise zu nachvollziehen und dabei die zentrale Umsteigpunkte zu identifizieren. Dadurch lassen sich die Knotenpunkte rekonstruieren, die für die Missionsarbeit eine wichtige Rolle spielen.

Mit dem Tool [QuickStatements](https://quickstatements.toolforge.org/) wurden die tabellarisch aufbereiten Daten in Wikidata hochgeladen. Jede Zeile kodiert eine Aussage zu einem Wikidata-Item und enthält mindestens Subjekt (Q-ID), Prädikat (P-ID) und Objekt (Wert). Optional können Qualifikatoren und Referenzen mitgegeben werden (Abbildung 5.8).

Ein Bild, das Text, Screenshot enthält.

KI-generierte Inhalte können fehlerhaft sein.

Abb. 5.8 Screenshot: QuickStatements

Nach dem Upload liegen die Informationen als Wikibase-Statement in Wikidata vor. Die RDF-Repräsintation wird von Wikidata bereitgestellt und kann u.a. im Turtle-Format aufgerufen werden. Eine Turtle-Datei beinhaltet [Präfix-Deklarationen](https://www.wikidata.org/wiki/EntitySchema:E49), Metadaten zum Datensatz und die direkten Aussagen eines Wikidata-Objekts mit Qualifikatoren und Referenzen. Das letzteres stellt die Abbildung 5.9 dar.

Ein Bild, das Text, Screenshot, Schrift enthält.

KI-generierte Inhalte können fehlerhaft sein.

Abb. 5.9 Ausschnitt aus der Turtle-Datei zum Datensatz Marianne Bonatz (Q135317408)

Die dargestellten Tripel folgen dem RDF-Datenmodell und spiegeln die semantische Struktur von Wikidata wider. Jede einzelne Aussage über ein Objekt wird durch eine eindeutige URI mit dem wds:-Präfix (Abbildung 5.8, rot markiert) gekennzeichnet. Dadurch wird die Aussage selbst als einstündige Ressource behandelt, sodass zusätzliche Informationen wie Qualifikatoren und Referenzen (Abbildung 5.8, lila markiert) direkt an die Aussage angehängt werden können.

Eine große Herausforderung bei der Datenmodellierung bestand in der Zuordnung von Relationen zwischen Hilfstabelle und Wikidata. Besonders anspruchsvoll war die Abgrenzung von Ortsaufenthalte (Wohnsitz, Wirkungsort) und Reisetätigkeit. Aufenthalte von Personen in Städten, die etwa dem Erlenen einer Fremdsprache oder dem langzeitlichen Besuch dienten, wurden als Wohnsitz modelliert. Als Wirkungsorte wurden alle Städte markiert, in denen die jeweilige Person länger- oder kurztätig war bspw. zu der Teilnahme an einer Konferenz oder zum Predigen. Allerdingst wurden kein projektspezifisches Kriterium zur Einstufung eines Wirkungsort und Wohnsitzes definiert (z.B. Mindestdauer oder Mindestzahl von Ereignissen). Als weitere soll der Kontext bzw. das Ziel von Reise, Besuch und Aufenthalt (Fremdsprache lernen, eine Kirche aufbauen, Kinder aus Missionsstellen nach Europa mitnehmen usw.) für die jeweiligen Personen präzisiert werden. Diese Beispiele verdeutlichen nur einen Teil der Herausforderungen bei der Datenmodellierung.