

7 Gelehrten-geschichte im Biographischen Lexikon des Kaiserthums Oesterreich – Vom Transkript in Wikisource zur Struktur in Wikidata

Christian Erlinger^{1,2}

Abstract

Das *Biographische Lexikon des Kaiserthums Oesterreich* (BLKÖ) von Constant von Wurzbach 1856–1891 herausgegeben, ist mit über 24.000 Einträgen die umfassendste biographische Sammlung des Habsburger Reiches des 18. und frühen 19. Jahrhunderts. Das in 60 Bänden erschienene Werk wurde auf der Plattform der deutschsprachigen Wikisource,³ eines Schwesternprojektes von Wikipedia für die Transkription gemeinfreier Werke, im Volltext erschlossen und dort nach Einträgen geordnet veröffentlicht. Ergänzend dazu wurde das Lexikon in seiner Gesamtheit sowohl bibliographisch in Wikidata, dem Wiki-Schwesternprojekt in Form eines Knowledge-Graphen, erfasst als auch biographisch für die jeweiligen Einträge des Lexikons ein Matching bzw. eine Neuanlage von Wikidata-Einträgen durchgeführt. Für die Gelehrten-geschichte enthält der *Wurzbach* vielfältiges Material. Ausbildungsstationen, Berufungsgeschichten, Werklisten sind zusätzlich zu klassischen biographischen Informationen vielfach vorkommende Informationen. »Gelehrte« stellen im *Wurzbach* selbstredend eine oft biographierte Personengruppe dar, Karriereschritte, ihre Beziehungen zu Wirkungsorten, ihre Verflechtungen mit anderen Forschenden sind durchwegs im Lexikon vorkommende Informationen. Diese können im korrespondierenden biographischen Wikidata-Item strukturiert erfasst und durch Belegfunktion mit dem Wurzbach-Eintrag eindeutig verlinkt und belegt werden.

Der vorliegende Beitrag möchte aufzeigen, wie mit gemeinfreien und im Wiki-Kontext erschlossenen Quellen strukturierte Informationen zur Gelehrten-geschichte erschlossen und nutzbar gemacht werden. Dies betrifft die strukturierte Erfassung in Wikidata, eine semantische Annotation im Quelltext in Wikisource (mittels RDFa) und zeigt auch, wie ein Mapping zwischen dem Wikidata-Schema und der PCP-Ontologie aussehen kann und welche Vorteile insbesondere das Referenzmodell in Wikidata für die Aufbereitung historischer Quellen und Informationen bietet.

¹ ORCID: <https://orcid.org/0000-0001-7872-9617>

² Zentral- und Hochschulbibliothek Luzern

³ https://de.wikisource.org/wiki/Biographisches_Lexikon_des_Kaiserthums_Oesterreich (besucht am 20.01.2022)

7.1 Das Biographische Lexikon des Kaiserthums Oesterreich (BLKÖ)

Constant von Wurzbach hat fast in der gesamten zweiten Hälfte des 19. Jahrhunderts an dem Mammutprojekt einer umfassenden biographischen Enzyklopädie der damaligen österreichischen Monarchie mit all seinen Kronländern aus dem heutigen Ost- und Mitteleuropa gearbeitet. Mit 60 Bänden und 24.254 Einträgen stellt der *Wurzbach* unübertroffen das biographische Standard- und Referenzwerk für die biographisch-historische Forschung zu Personen des Habsburger Reiches im 18. und 19. Jahrhundert dar.⁴ Wurzbach wurde 1849 zum Leiter der administrativen Bibliothek des Innenministeriums ernannt und bekleidete nominell dieses Amt bis ins Jahr 1891.⁵ Allerdings wurde er für seine Arbeit am biographischen Lexikon 1874 freigestellt und widmete sich fortan ausschließlich der Fertigstellung seiner umfangreichen Sammlung zu den Persönlichkeiten des Kaiserreichs.⁶ Wengleich zeitgenössische Begutachter des Werkes Kritik an der Auswahl und dem subjektiv geprägten Stil übten, respektive der lexikalischen Biographik generell kritisch gegenüber standen, so bleibt unbestritten, dass der *Wurzbach* nicht nur retrospektiv seine Bedeutung als biographische Ressource errungen hat, sondern dass er durch die höchst umfassende Sammlung an zeitgenössischen Quellenmaterialien selbst eine Vielfalt an unterschiedlichen publizistischen und schriftstellerischen Werken eingebunden und referenziert hat, was insbesondere für die weiteren Möglichkeiten der Analyse des Lexikons von großem Wert ist. Diese Materialsammlung besteht nach wie vor in ihrer Einheit und ist in der Wienbibliothek im Rathaus aufbewahrt.⁷ Wurzbach hat sich in seiner Arbeit, womöglich als einer der ersten Enzyklopädisten, weniger auf archivarische, sondern mehr auf publizistische Quellen verlassen.⁸ Die oft nachträgliche Kritik am *Wurzbach* einerseits bezüglich der verwendeten Materialien, andererseits betreffend die Auswahl oder den Stil seiner Biographien sollen aber hier nicht weiter vertieft werden.

Ganz im Gegensatz zu den oft zynischen Kritikern Wurzbachs und im Anschluss an das grundlegende Lob der Arbeit, wie es auch Glossy in seiner Biographie Wurzbachs in der *Allgemeinen Deutschen Biographie* übt,⁹ soll in dieser Arbeit gezeigt werden, welche Möglichkeiten in den aktuell vorhandenen technologischen Werkzeugen bestehen, ganz konkret mit dem offenen Wissensgraph Wikidata, um sowohl biographische als auch bibliographische Zusammenhänge in der Tiefe strukturiert zu erschließen und auszuwerten. Auch abweichende, d. h. aus heutiger Sicht falsche biographische Angaben kön-

⁴ H. Reiterer. „Wurzbach von Tannenberg“. In: *Österreichisches Biographisches Lexikon* 16 (2021), S. 377–378. DOI: 10.1553/0x003d2c8b

⁵ H. Reiterer (2021) (s. Anm. 4)

⁶ ebd.

⁷ Wienbibliothek im Rathaus: Dokumentation. <https://www.wienbibliothek.at/bestaende-sammlungen/dokumentation> (besucht am 08.06.2022)

⁸ Karl Glossy. „Art. 'Wurzbach, Constant von'“. In: *Allgemeine Deutsche Biographie* 55 (1910). Hrsg. von Historischen Kommission bei der Bayerischen Akademie der Wissenschaften, S. 135–138. URL: https://de.wikisource.org/w/index.php?title=ADB:Wurzbach-Tannenberg,_Constant_Ritter_von (besucht am 08.06.2022)

⁹ ebd.

nen dergestalt sichtbar gemacht werden und ermöglichen dadurch eine möglichst transparente Nachvollziehbarkeit, wie und woher sich womöglich über einen längeren Zeitraum fest etablierte Meinungen zu biographischen Irrtümern tradieren konnten.

Da Wurzbach neben der Beschreibung adeliger Persönlichkeiten wie der »Staatsmänner, Militärs [...], verdiente Staatsbeamte«¹⁰ auch vielfach der Verzeichnung der Gelehrten in seinem Werk Platz einräumte und durch die vorhandene Erfassung und Erschließung des Lexikons in Wikisource und Wikidata eine interoperable und offene Grundlage zur Nachnutzung besteht, war es naheliegend, die Anwendbarkeit der Ontologie professoraler Karrieremuster (PCP)¹¹ auf Basis der biographischen Daten Wurzbachs zu überprüfen. Wie kann eine solche Ontologie mit dem Datenmodell von Wikidata in Einklang bzw. in einen offenen Austausch gebracht werden und bestehen ergänzend Möglichkeiten, semantische Annotationen in Bezug auf den inhaltlichen Fokus der PCP-Ontologie direkt im Quelltext des biographischen Lexikons einzusetzen?

7.2 Der Wurzbach im Wiki*Versum

Das BLKÖ steht auf der Plattform *austrian literature online* (ALO) in einer gescannten Ausgabe ohne Volltexterkennung zur Verfügung.¹² Die 60 Bände des Werkes können dort in einer seitenweise strukturierten Bildanzeige eingesehen und heruntergeladen werden. Basierend auf diesem Scan besteht in der deutschsprachigen Wikisource eine mittlerweile komplett transkribierte und zumindest einmal korrigierte Volltextversion des gesamten *Wurzbach* offen zur Verfügung.¹³ Parallel zur Textausgabe in Wikisource wurde das gesamte BLKÖ 2020 in Wikidata bibliographisch erschlossen und mit den biographischen Einheiten vernetzt, bzw. fehlende biographische Items mit zumindest rudimentären Angaben in Wikidata initial angelegt. Weitere Scans des *Wurzbach* stehen in Google Books¹⁴ und dem austria-forum.org¹⁵ zur Verfügung. Weiters besteht eine um OCR angereicherte Version des ALO-Scans im Internet Archive.¹⁶

¹⁰ Constant von Wurzbach. „Vorrede“. In: *Biographisches Lexikon des Kaiserthums Oesterreich*. Bd. 1. 1867, S. IV. URL: [https://de.wikisource.org/w/index.php?title=BLK%C3%96:Vorrede_\(Band_1\)](https://de.wikisource.org/w/index.php?title=BLK%C3%96:Vorrede_(Band_1)) (besucht am 08.06.2022)

¹¹ Jennifer Blanke u. a. *Professorale Karrieremuster der Frühen Neuzeit - Version 0.2 (pcp-on-web)*. 2022. URL: <https://pcp-on-web.de/ontology/0.2/index-de.html> (besucht am 08.06.2022)

¹² Biographisches Lexikon des Kaiserthums Österreich, in: *austrian literature online*. <https://www.literature.at/collection.alo?objid=11104&from=1&to=50> (besucht am 08.06.2022)

¹³ Biographisches Lexikon des Kaiserthums Oesterreich, in: Wikisource. https://de.wikisource.org/wiki/Biographisches_Lexikon_des_Kaiserthums_Oesterreich (besucht am 08.06.2022)

¹⁴ Biographisches Lexikon des Kaiserthums Oesterreich, in Google Book Search, verzeichnet im GenWiki: [https://wiki.genealogy.net/Biographisches_Lexikon_des_Kaiserthums_Oesterreich_\(von_Wurzbach\)](https://wiki.genealogy.net/Biographisches_Lexikon_des_Kaiserthums_Oesterreich_(von_Wurzbach)) (besucht am 08.06.2022)

¹⁵ Biographisches Lexikon des Kaiserthums Oesterreich, in: *austria-forum.org*. <https://austria-forum.org/web-books/kategorie/lexika/wurzbach-lexikon> (besucht am 08.06.2022)

¹⁶ Biographisches Lexikon des Kaiserthums Oesterreich, in internet archive. <https://archive.org/search.php?query=subject%3A%22Wurzbach%20OCR%22> (besucht am 08.06.2022)

7.2.1 Wikisource

Wikisource ist ein 2003 gestartetes Projekt der Wikimedia Foundation mit dem Ziel, Volltexttranskripte von gemeinfreien Quellen in freier Lizenz zu erzeugen und zur Verfügung zu stellen.¹⁷ Wenngleich mit einer im Frühjahr 2022 erreichten Gesamtanzahl von knapp über 50.000 Werken im Vergleich zu großen Editions- oder Digitalisierungsplattform die Wikisource eher »klein« anmuten mag, so entfaltet sich die Bedeutung des hier eingestellten, bearbeiten und dargebotenen Quellenmaterials in seiner Qualität. Wikisource bietet sich einerseits als Nische an, um »unbekannte Quellenmaterialien ans Licht zu bringen und Einzeltexte oder Textkorpora zu präsentieren, die unter der Aufmerksamkeitsschwelle großer Editionsprojekte liegen.«¹⁸ Andererseits steht in Wikisource – im Vergleich zu den akademischen Editionsprojekten – die »gefällige Präsentation der Texte«¹⁹ im Vordergrund. Dies bedeutet aber keinesfalls einen besonderen Mangel an Qualität oder Sorgfältigkeit in der Transkription und Korrektur der Texte. Die Einfachheit der »Editions-Richtlinien«²⁰ der Wikisource soll einen leichten Einstieg als Mitarbeiter*in der Plattform erlauben; grundlegende editorische Leitlinie ist dabei, eine »möglichst originalgetreue inhaltliche Wiedergabe des Textes der Vorlage«²¹ zu erhalten. Das »Vier-Augen-Prinzip« versucht sicherzustellen, dass das Transkript der Quelle von zumindest zwei unterschiedlichen Personen gelesen und geprüft wurde.²² Die Quellenmaterialien sind dementsprechend in ihrem Bearbeitungsfortschritt (unkorrigiert – korrigiert – fertig) markiert. Die Textauswahl erfolgt – wie für ein Wiki-Projekt nicht anders vorstellbar – ausschließlich auf Basis der Community-Vorschläge. Texte mit weniger als 50 Seiten in der Vorlageform können sofort eingestellt und transkribiert werden, für größere Projekte gilt die Regel, dass diese Werke der Community vorgestellt werden und erst nach einer zehntägigen Stillhaltefrist, in der Einwände gegen das Projekt eingebracht werden können, eingestellt werden dürfen.

Neben vielen kleinen Quellen wurden und werden auch größere Projekte in der Wikisource editiert, die – sofern sich eine Institution dafür gefunden hätte – wohl auch in einem akademischen Editionsprojekt nicht schlecht aufgehoben wären: Dazu zählen neben der laufenden Erschließung des illustrierten Familienblattes *Die Gartenlaube*, die beiden biographischen Werke der *Allgemeinen Deutschen Biographie* und des hier diskutierten BLKÖ. Diese drei Großprojekte zeichnen sich unter anderem auch durch eine Gemeinsamkeit aus, dass sie nämlich allesamt bibliographisch auf Artikel- bzw. Lemma-Ebene in Wikidata strukturiert erfasst, das heißt »katalogisiert« sind

¹⁷ Susanne Haaf. „Rezension der Deutschsprachigen Wikisource“. In: *RIDE* (2018). DOI: 10.18716/ride.a.8.4

¹⁸ Jürgen Nemitz. „Wikisource – Beobachtungen zu einer kollaborativen digitalen Bibliothek“. In: *Mehr als Stadt, Land, Fluss: Festschrift für Ursula Braasch-Schwersmann*. Hrsg. von Lutz Vogel u. a. Neustadt an der Aisch, 2020, S. 338

¹⁹ Susanne Haaf (2018) (siehe Anm. ??)

²⁰ Wikisource Editionsrichtlinien, <https://de.wikisource.org/wiki/Wikisource:Editionsrichtlinien> (besucht am 08.06.2022)

²¹ ebd.

²² Wikisource: Hilfe:Korrekturlesen, <https://de.wikisource.org/wiki/Hilfe:Korrekturlesen> (besucht am 08.06.2022)

und somit einer analytischen Auswertung auf Basis von verlinkten Metadaten offenstehen.²³

Das BLKÖ in Wikisource unterscheidet sich entgegen den meisten anderen sowohl kleinen als auch großen Quellentexten in Wikisource dadurch, dass die Vorlagenscans nicht in Wikimedia Commons gespeichert sind, sondern unmittelbar aus einer externen Quelle eingebunden sind. Im konkreten Fall wurde für jedes Lemma des BLKÖ eine eigenständige Wikisource-Seite angelegt und mit dem Digitalisat der Plattform *austrian literature online* verlinkt. Im Bearbeitungsmodus eines BLKÖ-Eintrags erscheint somit neben der Wikisyntax das Digitalisat dieser Plattform. Zusätzlich zu den Volltexten je Lemma sind in Wikisource Inhaltsverzeichnisse je Band, verlinkte Register sowie Metadaten zum jeweiligen Lemma in einer Infobox aufgelistet. Diese Metadaten umfassen den Lemma-Titel, die genaue Fundstelle aus Band- und Seitenzahl, Vorgänger- und Nachfolger-Lemma sowie eine Vernetzung zu einem GND-Item, dem Wikidata-Item oder ggf. der Wikipedia-Seite der im jeweiligen Text beschriebenen Person. Darüber hinaus verfügt jedes BLKÖ-Lemma über ein eigenständiges bibliographisches Wikidata-Item welches über die Links in der Seitennavigation erreichbar ist.

7.2.2 Wikidata

Wikidata ist ein 2013 gegründetes Projekt der Wikimedia Foundation, um das globale Wissen in strukturierter Art und Weise in Form eines Knowledge Graphs zu erfassen.²⁴ Für Wikidata wurde mit Wikibase eine Erweiterung der allen Wiki-Projekten zugrunde liegenden Software MediaWiki entwickelt, die es ermöglicht, Wissen in Form von RDF-Tripeln zu speichern. Wikidata zeichnet sich durch seine Offenheit sowohl hinsichtlich der Daten als auch des Datenmodells, seine Multilingualität, die Vernetztheit und Zentralität der Daten aus. Wikidata ist einerseits offen um für faktisch jedwedes Konzept oder reales Objekt ein Item mit strukturierten und verlinkten Informationen aufnehmen zu können, andererseits bietet es für sämtliche Seiten in einem beliebigen Wiki-Schwesternprojekt die Möglichkeit (sei es Wikipedia, Wikisource etc.) ein beschreibendes Wikidata-Item mit Sitelink zur jeweiligen Projektseite zu erstellen. Für Wikisource-Projekte sind dies dann bibliographische Items in Wikidata, die einen solchen Sitelink zum jeweiligen Werk beinhalten. Für das BLKÖ existiert somit neben bibliographischen Items zum Lexikon selbst bzw. zu jedem einzelnen der 60 Bände, auch zu jedem Lemma (und somit jeder Wikisource-Seite des BLKÖ) ein eigenständiges bibliographisches Wikidata-Item.

Wie in Abbildung 7.1²⁵ gezeigt, werden für die Erzeugung eines Wikidata-Items unterschiedliche existierende Werte aus den Wik-

²³ Ein Desiderat, welches in der Rezension der Wikisource von Haaf (siehe Anm. 17) noch als Manko angeführt wurde.

²⁴ Denny Vandrecic. „Wikidata: A Free Collaborative Knowledgebase“. In: *Communication of the ACM* 10 (2014), S. 78–85. doi: 10.1145/2629489

²⁵ Beispiellemma sowohl für Abbildung 7.1 als auch für Abbildung 7.2: Constant Wurzbach. „Gaßner, Ferdinand Simon“. In: *Biographisches Lexikon des Kaiserthums Oesterreich*. Bd. 5. 1859, S. 98. URL: https://de.wikisource.org/wiki/BLK%C3%96:Ga%C3%9Fner,_Ferdinand_Simon (besucht am 08.06.2022)

bibliographische Verweise für Aufnahmen in anderen externen Katalogen.

Die besondere Wirkmacht von Wikidata kommt allerdings dadurch zustande, dass neben den bibliographischen Informationen auch die biographischen Informationen zu den beschriebenen Personen erfasst werden können. Wie gezeigt verlinkt jedes Lemma-Katalogisat über die Schlagwort-Eigenschaft ein Personen-Item in Wikidata. Aus den Wikisource-Texten des BLKÖ wiederum lassen sich diese Items mit grundlegenden biographischen Informationen strukturiert erzeugen bzw. anreichern.

Biographisch

The screenshot shows the Wikidata page for the BLKÖ entry 'BLKÖ-Gaßner, Ferdinand Simon'. The main subject statement is 'described by source'. Other statements include 'student of', 'date of birth', and 'stated in'. The page also displays a snippet of the original text from the BLKÖ source, which is a biographical entry for Ferdinand Simon Gaßner, detailing his life, education, and musical career.

Abbildung 7.2: Graphische Darstellung von biographischen Wikidata-Eigenschaften auf Basis der textuellen Vorlage in Wikisource

Die biographischen Wikidata-Items besitzen, wie in Abbildung 7.2 gezeigt, reziprok zum Main-Subject-Statement des bibliographischen BLKÖ-Items das Statement »described by source« wo die Person mit der beschreibenden Stelle, d. h. mit dem Lemma im BLKÖ vernetzt ist. Die einzelnen Lexikoneinträge des *Wurzbach* bieten vielfältige biographische Informationen, die strukturiert im Personen-Item in Wikidata erfasst werden können. Dazu zählen neben Geburts- und Sterbedaten, ausführlichen Namensangaben und Verwandtschaftsverhältnissen auch Informationen zu Ausbildung, Beruf oder anderen persönlichen Netzwerken.

Das Datenmodell in Wikidata erlaubt es, jeder Aussage auch eine detaillierte Referenz einzuschreiben. Das heißt, dass jedem Geburtsdatum aus dem BLKÖ im Wikidata-Item immer die exakte Fundstelle der Information im *Wurzbach* eingeschrieben werden kann. Stellt sich aufgrund neuerer oder bereits bekannter Forschung heraus, dass das im *Wurzbach* erfasste Geburtsdatum nicht korrekt ist, so wird in Wikidata der »Rank« des Statements mit dem Wert aus dem BLKÖ als »deprecated« entsprechend gekennzeichnet. Es braucht daher keine Berichtigung, aber auch keine Beschönigung des *Wurzbach*, denn der zeigt eben jene Werte, die erfasst wurden. Wikidata erlaubt aber durch die Granularität des Datenmodells, insbesondere durch die Möglichkeit, jede Aussage einzeln und getrennt zu belegen,

Abfragen nach »ungültigen« Aussagen einer historischen Quelle zu stellen.²⁶

7.3 »Academic Patterns« in Wikidata

Akademische Karriereverläufe sind in ihren Grundzügen auch im BLKÖ durchaus Bestandteil der Biographien. So werden Universitäten genannt, an denen die biographierten Personen studiert oder gelehrt haben, Schüler und Lehrer werden angeführt, berufliche Tätigkeiten und Stationen an denen diese ausgeübt werden und wurden finden sich ebenso. Darüber hinaus kann es auch zu detaillierten zeitlichen Zuordnungen (Studienabschluss oder Ausübung einer bestimmten beruflichen Position in einem definierten Zeitfenster) kommen. All diese in der Quelle auffindbaren Informationen lassen sich in strukturierter Form in den biographischen Wikidata-Items erfassen. Einige dieser spezifischen Eigenschaften sind in der folgenden Tabelle 7.1 aufgelistet und kurz erläutert.

Wikidata-Property (ID & Label)	Beschreibung
P69 – educated at	Verlinkt eine Person mit einer Bildungsinstitution
P39 – position held	Verlinkt eine Person mit einer beruflichen Position (Professur, Leitungsaufgabe etc)
P1066 – student of	Verlinkt eine Person mit einer anderen Person, die als Lehrer fungiert
P1416 – affiliation	Verlinkt eine Person (oder Organisation) mit einer anderen Organisation, muss dabei kein »arbeitsrechtliches« Verhältnis bestanden haben
P802 – student	Reziprok zu P1066 – verlinkt eine Person mit ihren jeweiligen »Studierenden«
P812 – academic major	Kann zur näheren Beschreibung des Statements »educated at« als Darstellung des gewählten Faches dienen
P1026 – acad. thesis	Vernetzt eine Person, insbesondere zur detaillierten Erfassung des »educated at«-Statements mit einem bibliographischen Wikidata-Item einer akademischen Abschlussarbeit
P4101 – dissertation submitted to	Verlinkt – semantisch um eine Spur deutlicher als »student of« die Vernetzung zwischen Dissertant*in und der betreuenden Person
P9161 – dissertation committee member	Ermöglicht die Verlinkung zwischen Dissertant*in und weiteren nicht primär betreuenden Personen einer Doktoratsarbeit

Tabelle 7.1: Ausgewählte Wikidata-Eigenschaften zur Beschreibung akademischer Karriereverläufe

Die folgende Abbildung 7.3 zeigt strukturiert erfasste Informationen zu Bildungsstationen des von Wurzbach biographierten Musikers und Komponisten Ferdinand Simon Gaßner. Erfasst und verlinkt wurde, dass Gaßner in Gießen einen Dokortitel der Philosophie im Jahr 1819 erhalten hat, dass er bei Gottfried Weber studiert hat (wiewohl diese Beziehung nicht nachweislich mit dem Studium in Gießen steht, und somit auch bei erstem Statement diese Beziehung nicht zur näheren Spezifizierung eingebaut wurde).

In Abbildung 7.4 werden für Ferdinand Simon Gaßner Informationen zu seiner beruflichen Orientierung und seinen Positionen dargestellt. Das Amt des »Musikdirektors«, welches er im zweiten Jahrzehnt des 19. Jahrhunderts an der Universität Gießen innehatte, wurde mit den entsprechenden Werten detailliert dargestellt.

Allen Aussagen rund um den Karriereverlauf Gaßners gemein ist, dass es sich um Informationen handelt, die im BLKÖ aufzufinden sind. Dementsprechend ist allen Aussagen die Belegstelle mit der

²⁶ Die Wikidata-Abfrage <https://w.wiki/4J8q> illustriert den beschriebenen Sachverhalt bei Erfassung abweichender Geburtsdaten in unterschiedlichen Quellen: Das BLKÖ und die GND weisen für die gleiche Person unterschiedliche Geburtsdaten aus (besucht am 08.06.2022).

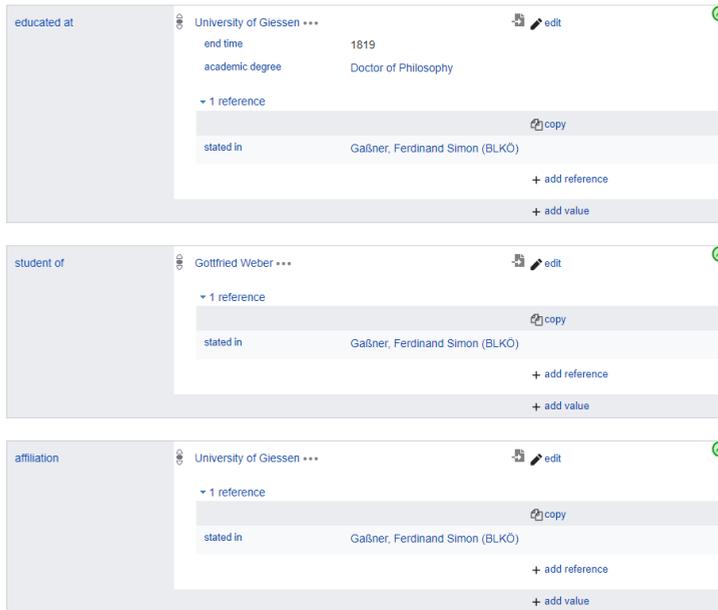


Abbildung 7.3: Wikidata-Statements zu akademischen Karriereverläufen am Beispiel von Ferdinand Simon Gaßner²⁷

Aussage »stated in« und dem Link auf das bibliographische Item des Lexikoneintrags angefügt.

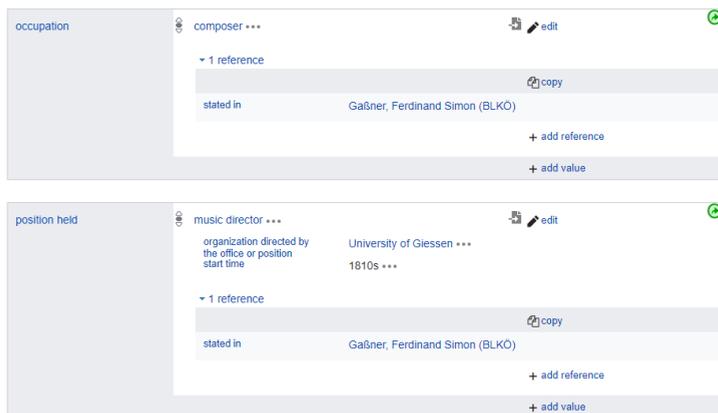


Abbildung 7.4: Wikidata-Statements zu beruflichen Aussagen am Beispiel von Ferdinand Simon Gaßner²⁸

7.3.1 Ontology-Mapping PCP und Wikidata

Die PCP-Ontologie kann sowohl in ihren Klassen, als auch ihren Properties mit äquivalenten Entitäten in Wikidata gematcht werden. Im Rahmen dieser Arbeit wurden alle PCP-Entitäten (Eigenschaften und Klassen) überprüft und jeweils übereinstimmende Treffer mit den Eigenschaften in Tabelle 7.2 zugeordnet. Insgesamt konnten 99

Entitäten sinnvoll mit korrespondierenden Elementen in Wikidata gematcht werden.²⁹ Methodisch wurde das Matching für die Entitäten mit OpenRefine durchgeführt; die Überprüfung der Übereinstimmung externer Properties musste manuell durchgeführt werden, da hier – zumindest bekannterweise noch – keine Möglichkeit existiert, anstelle der Objekt-Entitäten in Wikidata Properties als Gegenstand der Reconciliation auszuwählen.

Wikidata-Property (ID & Label)	Beschreibung
P1628 – equivalent property	beschreibt ein exaktes Matching der Wikidata-Eigenschaft mit einer externen Eigenschaft
P2235 – equivalent superproperty	beschreibt ein Matching einer Wikidata-Eigenschaft mit einer externen Eigenschaft von der erstere einen Teil repräsentiert
P2236 – equivalent subproperty	beschreibt ein Matching einer Wikidata-Eigenschaft mit einer externen Eigenschaft von letztere einen Teil repräsentiert.
P1709 – equivalent class	beschreibt ein Matching eines Wikidata-Objektes mit einer externen Klasse.
P4900 – broader Concept	beschreibt das Matcing eines Wikidata-Objektes mit einer externen Klasse, die ein thematisch übergeordnetes, breiteres Konzept beschreibt (entspricht exakt skos:broaderConcept)

Tabelle 7.2: Wikidata-Properties für das Matching externer Ontologien

Das Matching einer externen Ontologie mit Wikidata ist bei einem derart dynamischen und sich permanent wandelnden Knowledge-Graphen natürlich immer auch ein Abbild des jeweiligen Zeitpunktes. Ziel eines Mappingprozesses ist es, einen möglichst hohen Anteil an genau übereinstimmenden Entitäten zu finden, da dies die Interoperabilität der mit den jeweiligen Entitäten beschriebenen Daten erhöht. In Tabelle 7.3 ist ersichtlich, dass von den 70 Klassen-Entitäten 67 einer exakten Übereinstimmung zugeführt werden konnten. Bei den Object-Properties konnten hingegen nur drei, bei den Datatype-Properties wiederum neun mit einer exakt gleichbedeutenden Wikidata-Eigenschaft zusammengeführt werden.

Das generelle Matching-Ergebnis zeigt, dass bei den Entitäten eine sehr hohe Trefferquote erzielt wurde mit nur 14 Entitäten, die keinem Wikidata-Item, weder exakt noch in einem über- oder untergeordneten Sinnzusammenhang, zugewiesen werden konnten. Wesentlich schlechter ist die Trefferquote allerdings bei den Properties. 31 der Object-Properties und zwölf der Datatype-Properties konnten nicht gematcht werden. Dieser Umstand liegt weniger in der Gesamtzahl an verfügbaren Eigenschaften in Wikidata,³⁰ sondern insbesondere in grundlegend unterschiedlichen Herangehensweisen an die Modellierung der vorliegenden Ontologie. So besitzt Wikidata bspw. keine eigenen oder auch ähnlichen Properties für »latinisierte Namen« (pcp:latinForename oder pcp:latinSurname); für diesen Zweck ist die multilinguale Konzeption von Wikidata völlig hinreichend, bzw. können in den jeweiligen Namensfeldern Übersetzungen oder Transliterationen erfasst und durch einen Qualifier näher definiert werden. Auch Properties wie abstrakt formuliert »has work«, existieren zwar in gewissen Formen in Wikidata (zB »notable works«) generell wird

²⁹ Das Mapping wurde als RDF-Turtle-File veröffentlicht: ERLINGER: Wikidata mapping of the PCP-Ontology [Data set]. Zenodo. (13.07.2022) DOI:10.5281/zenodo.6829140

³⁰ Per 4. Juli 2022 existierten in Wikidata 10.100 Properties (vgl. WikidataMeter. Twitter (04.07.2022), <https://twitter.com/WikidataMeter/status/1543814030908788737>). Allgemein: Abfrage aller in Wikidata verfügbaren Properties: <https://w.wiki/5SmG>. (besucht am 14.07.2022)

bei Werkbezügen der umgekehrte Modellierungsweg gegangen, d. h. auch das Werk wird erfasst und durch die Verlinkung mit den oder der Urheber*in vernetzt.

Letztlich ist aber auch die prinzipielle Offenheit des Datenmodells in Wikidata ein Umstand, der womöglich bei einigen der im Zuge dieser Arbeit nicht gematchten Eigenschaften durchaus noch Übereinstimmungen bringen könnte, je nach Betrachtungsweise des Durchführenden. Hier darf angemerkt werden, dass der Autor im Matching eine eher defensive Strategie durchgeführt hat, d. h. Matchings nur dann gesetzt hat, wenn diese wirklich sehr augenscheinlich waren.

Ein weiterer Grund für die Unmöglichkeit des Mappings zwischen der PCP-Ontologie und Wikidata liegt dann vor, wenn die Logik zwischen Properties und Objects genau umgekehrt ist, wie im Fall der Geschlechtsauszeichnung: »male«, »female«, »divers« sind Eigenschaften und »Gender« wird als Klasse definiert, in Wikidata liegt hingegen die Eigenschaft »sex or gender« vor.

PCP-Type	P1709	P4900	P1628	P2235	P2236	None	Summe
owl:Class	67	3				14	84
owl:ObjectProperty			17	7	1	31	56
owl:DatatypeProperty			10	4		12	26
Summe	67	3	27	11	1	57	166

Tabelle 7.3: Verteilung der Matching-Properties im Ontologie-Matching

Durch das Eintragen der Mappings mit externen Ontologien in die jeweiligen Wikidata-Items bzw. auch Properties entsteht somit die Möglichkeit, Datenbestände in Wikidata als RDF-Tripel mit Eigenschaften externer Beschreibungssysteme auszugeben. Beispielhaft wird die Wikidata SPARQL-Abfrage für den zuvor bereits genannten Ferdinand Simon Gaßner gezeigt. Die Abfrage prüft alle vom Item Gaßners ausgehenden sowie alle nach Gaßner eingehenden Statements und gibt jene Tripel aus, bei denen in der Wikidata-Property ein äquivalentes Matching in die PCP-Ontologie vorhanden ist.

Subject	PCP-Property	Wikidata äquivalente Property	Objekt
Ferdinand Simon Gaßner (wd:Q1405801)	pcp:dateOfBirth	wd:P569	1798-01-06T00:00:00Z
Ferdinand Simon Gaßner (wd:Q1405801)	pcp:dateOfBirth	wd:P569	1798-01-16T00:00:00Z
Ferdinand Simon Gaßner (wd:Q1405801)	pcp:gnd	wd:P227	116450088
Ferdinand Simon Gaßner (wd:Q1405801)	pcp:dateOfDeath	wd:P570	1851-02-25T00:00:00Z
Ferdinand Simon Gaßner (wd:Q1405801)	pcp:forename	wd:P735	Ferdinand (wd:Q3276175)
Gottfried Weber (wd:Q215604)	pcp:student	wd:P802	Ferdinand Simon Gaßner (wd:Q1405801)
Gaßner, Ferdinand Simon (ADB) (wd:Q21233880)	pcp:hasKeyword	wd:P921	Ferdinand Simon Gaßner (wd:Q1405801)
Gaßner, Ferdinand Simon (ADB) (wd:Q21233880)	pcp:hasSubject	wd:P921	Ferdinand Simon Gaßner (wd:Q1405801)
Artikel in Riemanns Musiklexikon (wd:Q27773928)	pcp:hasKeyword	wd:P921	Ferdinand Simon Gaßner (wd:Q1405801)
Artikel in Riemanns Musiklexikon (wd:Q27773928)	pcp:hasSubject	wd:P921	Ferdinand Simon Gaßner (wd:Q1405801)
Gaßner, Ferdinand Simon (BLKÖ) (wd:Q88685147)	pcp:hasKeyword	wd:P921	Ferdinand Simon Gaßner (wd:Q1405801)
Gaßner, Ferdinand Simon (BLKÖ) (wd:Q88685147)	pcp:hasSubject	wd:P921	Ferdinand Simon Gaßner (wd:Q1405801)

Tabelle 7.4: Mapping der Wikidata-Statements in die PCP-Ontologie ausgehend von und eingehend nach Ferdinand Simon Gaßner. (SPARQL-Abfrage: <https://w.wiki/5T28>)

In gleicher Weise können neben den Properties nun auch Klassen, die als Subjekt- oder Objekt fungieren und in die PCP-Ontologie äquivalent gemappt sind, abgefragt werden.³¹

Dass sich wie dargestellt Ontologien ad hoc wohl in vielen Fällen nicht zur Gänze mappen lassen, ist aber auch kein großer Nachteil. Fehlende Entitäten können ohne weiteres leicht nachgetragen werden. Bei fehlenden Properties hingegen müsste zuerst wohl überlegt werden, ob eine neue Property im Gesamtsystem der existierenden Eigenschaften sinnvoll ist. Das Anlegen neuer Eigenschaften in Wikidata setzt ohnehin einen Diskussionsprozess in der Community voraus,³² wo dieser Umstand letztlich auch noch einmal zu Tage treten kann. Aber auch hier ist prinzipiell nicht ausgeschlossen, den Abdeckungsgrad zwischen externer Ontologie und Wikidata zu erhöhen, mit Ausnahme jener Fälle, wo eine Vertauschung von Eigenschafts- und Objektwert (wie bei der Darstellung von Geschlechtern in PCP) zwischen den beiden Systemen vorliegt.

7.3.2 Semantische Annotation in Wikisource mit RDFa

Nicht alle im Quelltext auftauchenden Entitäten sind immer sehr eindeutig oder letztlich auch »gehaltvoll« genug, um in Wikidata in Form eines RDF-Statements strukturiert erfasst werden zu können. Die Annotation von Fakten unmittelbar im Quelltext stellt dabei durchaus eine Möglichkeit dar, den Quelltext selbst maschinenlesbar auszuzeichnen und durch Verlinkungen mit externen Datenbeständen anzureichern.

In Form von RDFa steht eine Syntax zu Verfügung, die es erlaubt, direkt in den HTML-Code einer Webresource semantische Strukturen einzubauen.³³ Durch Umklammerung von Stellen mit Tags, denen in den Attributen sowohl die Verlinkung zur beschreibenden Entität als auch durch Angabe von Properties das semantische Verhältnis von getagter Entität und dem vorliegenden Gesamtkonzept eingeschrieben ist, werden Aussagen modelliert. In Wikisource steht eine Vorlage »Annotate QID«³⁴ zur Verfügung, die wie in Abbildung 7.5 erlaubt, durch Umspannung einer Quelltextstelle und Angabe der Q-ID eines Wikidata-Items RDFa-Tags in der Ausgabe zu erzeugen. Das um eine beschreibende Komponente ergänzte »title«-Attribut im RDFa-Tag wird aus einer Wikidata-Abfrage aus den Feldern »rdf:label« und »schema:description« unmittelbar beim Rendering der Seite erzeugt und bietet somit eine sehr dezente aber grundsätzliche vorhandene Information über die in der Quellenstelle angegebene Entität. Der Mehrwert steigt natürlich noch wesentlich, wenn die in der Quelle angegebene Namensform wesentlich von der üblichen Form abweicht.

Mit einer einfachen Erweiterung der Vorlage kann wie in Abbildung 7.6 angeführt zusätzlich zur Objekt-Entität auch noch eine Property übergeben werden, die dann den semantischen Gehalt des RDFa-Tags noch zusätzlich erhöht.

³¹ Wikidata SPARL-Query: <https://w.wiki/5TBp> (besucht am 15.07.2022)

³² Wikidata Property Proposal: https://www.wikidata.org/wiki/Wikidata:Property_proposal (besucht am 15.07.2022)

³³ RDFa-Info Website. <https://rdfa.info> (besucht am 15.07.2022)

³⁴ Vorlage:Annotate QID, https://de.wikisource.org/wiki/Vorlage:Annotate_QID (besucht am 15.07.2022)

Gaßner, Ferdinand Simon (*Compositeur und Musikschriftsteller*, geb. zu Wien 6. Jänner 1798, gest. zu Karlsruhe 25. Febr. 1851). Sohn des Malers *Simon G.* (s. d. S. 100). Zeigte früh großes Talent für Musik, erhielt in Karlsruhe Unterricht in der Violine und besuchte daselbst das Gymnasium. Später wählte er die Musik zu seinem Lebensberuf, bekam zuerst eine Stelle in der Karlsruher Hofcapelle, und als in Mainz 1816 ein neues Theater erbaut wurde, eine Anstellung daselbst als Violinist. Hier trat er auch in näherem Verkehr mit dem berühmten **Gottfried Weber**, unter dessen Leitung G. seine Kunstbildung vollendete. Nach einem von G. veranstalteten

`{{aqid|Q215604|Gottfried Weber}}`

`Gottfried Weber`

`{{aqid|Q215604|Gottfried Weber|Prop=P802}}`

`Gottfried Weber`

Abbildung 7.5: Anwendung und Darstellung des Ergebnisses bei Verwendung der Vorlage:Annotate QID in Wikisource

Abbildung 7.6: Erweiterung der semantischen Annotation um Property

7.4 Conclusio

Historische biographische Quellenmaterialien wie das in diesem Beitrag besprochene *Biographische Lexikon des Kaiserthums Österreich* bieten viele Detailinformationen, die einer strukturierten Beschreibung und Analyse zugänglich gemacht werden können. Die Darstellung akademischer Lebensläufe ist dabei sehr beachtenswert. Mit dem PCP-Ontologieschema liegt eine RDF-Beschreibung speziell für akademische Karrieremuster vor. Im Zuge der vorliegenden Arbeit wurde untersucht, ob diese Ontologie auch mit dem riesigen und flexiblen Datengraph von Wikidata in Übereinstimmung gebracht werden kann, bzw. ob dadurch grosse Teile der Beschreibungselemente der PCP-Ontologie durch Wikidata-Entitäten ersetzt werden könnten. Mehr als 50 % der PCP-Entitäten, sowohl Klassen wie auch Properties, lassen sich mit Wikidata-Einträgen in Übereinstimmung bringen. Eine vollständige Übereinstimmung ist aber nicht erzielbar, da dies den grundlegenden Ideen zur Modellgestaltung in den jeweiligen Systemen widersprechen würde.

Unzweifelhaft ist es durch die gematchten Entitäten möglich, gegenseitige Datenanreicherungen zu erzielen. Mit relativ überschaubarem Aufwand und unter Zuhilfenahme von niederschweligen und frei verfügbaren Technologien wie OpenRefine gelingt es sehr schnell und nachvollziehbar, ein Mapping zwischen Wikidata und einer externen Ontologie vorzubereiten und auch unmittelbar in Wikidata einzuspielen. Ist ein solches System dann in Wikidata verfügbar, so steht der gesamte Datenbestand des freien Datengraphs offen, um abgefragt zu werden. Dabei können die Werte übereinstimmend mit der fremden Ontologie ausgegeben und in andere Datenbanken oder Anwendungen überführt und nachgenutzt werden.

Literatur

Blanke, Jennifer u. a. *Professorale Karrieremuster der Frühen Neuzeit - Version 0.2 (pcp-on-web)*. 2022. URL: <https://pcp-on-web.de/ontology/0.2/index-de.html> (besucht am 08.06.2022).

- Glossy, Karl. „Art. 'Wurzbach, Constant von'“. In: *Allgemeine Deutsche Biographie* 55 (1910). Hrsg. von Historischen Kommission bei der Bayerischen Akademie der Wissenschaften, S. 135–138. URL: https://de.wikisource.org/w/index.php?title=ADB:Wurzbach-Tannenberg,_Constant-Ritter_von (besucht am 08.06.2022).
- Haaf, Susanne. „Rezension der Deutschsprachigen Wikisource“. In: *RIDE* (2018). DOI: 10.18716/ride.a.8.4.
- Nemitz, Jürgen. „Wikisource – Beobachtungen zu einer kollaborativen digitalen Bibliothek“. In: *Mehr als Stadt, Land, Fluss: Festschrift für Ursula Braasch-Schwersmann*. Hrsg. von Lutz Vogel u. a. Neustadt an der Aisch, 2020, S. 338.
- Reiterer, H. „Wurzbach von Tannenberg“. In: *Österreichisches Biographisches Lexikon* 16 (2021), S. 377–378. DOI: 10.1553/0x003d2c8b.
- Vandrecic, Denny. „Wikidata: A Free Collaborative Knowledgebase“. In: *Communication of the ACM* 10 (2014), S. 78–85. DOI: 10.1145/2629489.
- Wurzbach, Constant. „Gaßner, Ferdinand Simon“. In: *Biographisches Lexikon des Kaiserthums Oesterreich*. Bd. 5. 1859, S. 98. URL: https://de.wikisource.org/wiki/BLK%C3%96:Ga%C3%9Fner,_Ferdinand_Simon (besucht am 08.06.2022).
- Wurzbach, Constant von. „Vorrede“. In: *Biographisches Lexikon des Kaiserthums Oesterreich*. Bd. 1. 1867, S. IV. URL: [https://de.wikisource.org/w/index.php?title=BLK%C3%96:Vorrede_\(Band_1\)](https://de.wikisource.org/w/index.php?title=BLK%C3%96:Vorrede_(Band_1)) (besucht am 08.06.2022).