

# Wikidata Quality

Wikidata ist der Name einer frei bearbeitbaren Datenbank, die 2012 gegründet wurde und unter anderem das Ziel hat, Wikipedia zu unterstützen. Das Projekt wurde von Wikimedia Deutschland gestartet und stellt als gemeinsame Quelle bestimmte Datentypen für Wikimedia-Projekte bereit, zum Beispiel Geburtsdaten oder sonstige allgemeingültige Daten, die in den Artikeln aller Wikipedien verwendet werden können. Bis Mai 2014 wurden 17 Millionen Datenelemente eingetragen. Diese werden sowohl in Wikipedia und ihren Schwesterprojekten als auch außerhalb von Wikimedia genutzt. Die Aufgabe des Projektteams ist es, in Kooperation mit dem **Projektpartner Wikimedia Deutschland**, Werkzeuge und Methoden zu entwerfen und entwickeln, welche die Qualität der Daten in Wikidata überprüfen und verbessern.



## Projektbeschreibung

Wikidata kann wie Wikipedia frei editiert werden, erfordert jedoch Kenntnis über mögliche Aussagen (Eigenschaft-Wert-Paare), die über Datenelemente getroffen werden können und Expertenwissen über das Element selbst. Diese Freiheit kann dazu führen, dass fehlerhafte Daten eingetragen werden. Beispielsweise können Wertebereiche einer bestimmten Eigenschaft verletzt werden oder veraltete Daten gelangen in die Datenbank. Die Wikidata-Community hat bereits ein rudimentäres System, einen Bot, zur regelbasierten Konsistenzprüfung einzelner Eigenschaften erstellt. Dieser prüft beispielsweise die konkreten Werte einer Eigenschaft anhand der erlaubten Werte. Im Rahmen des Projektes soll dieses System in zwei Stoßrichtungen erweitert und verbessert werden.

**1. Transparente Fehlererkennung:** Inkonsistenzen in den Daten der Wikidata werden bisher teilweise durch Bots in Berichten gesammelt. Diese Sammlung soll für den Wikidata-Nutzer transparenter gestaltet werden, z.B. schon beim Bearbeiten eines Datenelementes. Zudem soll es der Community ermöglicht werden, Inkonsistenzen leicht aufzufinden und zu korrigieren.

**2. Validierung mit externen Daten:** Externe Quellen, wie die Internet Movie Database (IMDB), sollen in einem zweiten Schritt als Prüfinstanzen in die Konsistenzprüfung einbezogen werden. Z.B. können damit die Schauspieler eines Films mit den entsprechenden Daten in der IMDB abgeglichen werden und Wikidata-Nutzer über Inkonsistenzen beim Bearbeiten informiert werden.

**3. Anreicherung durch Quellinformationen:** Weiterhin soll die Benutzerfreundlichkeit beim Hinzufügen von Quellinformationen verbessert werden. Quellinformationen belegen Aussagen in Wikidata und sind damit entscheidend für die Qualität der eingetragenen Daten.

## Projektpartner

Wikidata Quality ist ein gemeinsam mit der Organisation Wikimedia Deutschland (Berlin) durchgeführtes Projekt. Die Wikimedia Deutschland – Gesellschaft zur Förderung Freien Wissens e. V. (WMDE) ist die deutsche Sektion der Wikimedia Foundation, die verschiedene Wikipedien und andere Projekte betreibt.



Wikimedia Deutschland wurde als erste nationale Wikimedia-Organisation von der Wikimedia Foundation anerkannt. Satzungsgemäßer Zweck des Vereins ist es, „die Erstellung, Sammlung und Verbreitung freier Inhalte (engl. Open Content) in selbstloser Tätigkeit zu fördern, um die Chancengleichheit beim Zugang zu Wissen und die Bildung zu fördern.“

Der Verein beschäftigt 68 Angestellte, davon 15 MitarbeiterInnen in der Softwareentwicklung. Wikidata ist das größte Projekt der Wikimedia Deutschland Softwareentwicklungs-Abteilung. Die Zusammenarbeit im Rahmen des Bachelorprojektes erfolgt in der Zusammenarbeit mit dem Wikidata-Team in Berlin.



Das Projekt für 6-8 Studenten beginnt im Oktober 2014 und wird durch Prof. Dr. Felix Naumann und Anja Jentzsch betreut. Fragen können gerne an [anja.jentzsch@hpi.uni-potsdam.de](mailto:anja.jentzsch@hpi.uni-potsdam.de) gerichtet werden.