

Wikidata: Nutzungsmöglichkeiten und Anwendungsbeispiele für den Bereich Digital Cultural Heritage

Müller-Birn, Claudia

clmb@inf.fu-berlin.de

Freie Universität Berlin, Deutschland

Schelbert, Georg

georg.schelbert@hu-berlin.de

Humboldt-Universität zu Berlin

Raspe, Martin

raspe@biblhertz.it

Max-Planck-Institut für Kunstgeschichte Rom

Wübbena, Thorsten

wuebbena@kunst.uni-frankfurt.de

Goethe-Universität Frankfurt

Einführung

Das im Jahr 2012 gegründete Projekt Wikidata ist ein Schwesterprojekt der Wikipedia, welches strukturierte Daten verwaltet, die in Wikipedia oder in anderen Wikimedia-Projekten verwendet werden können. Der Wikidata-Ansatz basiert auf einer Reihe von Designentscheidungen, wie der offenen Bearbeitung, der Pluralität (Daten aus unterschiedlichen Quellen mit unterschiedlicher Aussage sind erlaubt) und der Mehrsprachigkeit (Vrande[#] & Krötzsch, 2014).

Die strukturierten Daten sind nach konkreten Entitäten (Items) organisiert, die mit Aussagen (Statements) beschrieben werden. Jede Entität hat ein sprachspezifisches Label. Eine Entität, beispielsweise Q12418 besitzt das Label „Mona Lisa“. Sogenannte Eigenschaften (Properties) erlauben dann die Beschreibung der Entitäten mit Aussagen. Q12418 (Subjekt) ist eine Instanz „instance-of“ (P31 Property) der Entität „Painting“ (Q3305213 Objekt) und sagt aus, dass es sich hier um ein Gemälde handelt. Mit dieser Aussagestruktur bestehend aus Subjekt-Prädikat-Objekt bietet Wikidata maximale Flexibilität, da Wikidata erlaubt, neue Eigenschaften zu erstellen und Aussagen über sie zu machen. Durch die zusätzliche (eigentlich verpflichtende) Angabe von Referenzen kann Wikidata die Funktion einer sekundären, verifizierbaren Wissensbasis übernehmen. Durch die

Angabe von Referenzen werden in Wikidata nicht in erster Linie Tatsachen über die Welt gespeichert, sondern vielmehr Verlinkungen zu Quellen erstellt. Die Umstände, dass derzeit noch häufig de facto Referenzen fehlen und dass grundsätzlich Personen Falscheingaben machen werden können, sind kritisch zu berücksichtigen, jedoch nicht Gegenstand des Workshops.

Im September 2017 umfasste Wikidata Informationen von über 29 Millionen Entitäten, die mit Hilfe von über 175.000 registrierten Nutzern erstellt wurden. Ein wesentlicher Mehrwert von Wikidata liegt in der Nutzung der strukturierten Daten, denn diese können von Menschen und Maschinen gleichermaßen „verstanden“ werden (Müller-Birn et al., 2015). Das wurde dadurch ermöglicht, dass das bestehende Faktenwissen in terminologische Wissen überführt wurde (Erxleben et al., 2014). Somit können einerseits die Daten über ein User Interface von Menschen bearbeitet sowie abgefragt und andererseits (unter anderem) über den Wikidata SPARQL Query Service (<https://query.wikidata.org/>), von Softwareanwendungen verwendet werden. Darüber hinaus bieten Anwendungen wie SQID die Möglichkeit, das terminologische Wissen zu explorieren. Wikidata findet mittlerweile in vielfältigen Bereichen Anwendung: wie beispielsweise in interaktiven Abfrageansichten zu bestimmten Themen (z. B. das Oscar-Portal der Zeitung FAZ online), aber auch zur Unterstützung von Q&A-Systeme, wie beispielsweise Apples Siri Suchmaschine.

Wikidata im Bereich Digital Cultural Heritage

Solche oder ähnliche Anwendungen sind aber ebenfalls im Bereich der Geistes- und Kulturwissenschaften denkbar, wie die Beiträge zu der diesjährigen WikidataCon (https://www.wikidata.org/wiki/Wikidata:WikidataCon_2017) zeigen. Die wachsende Vernetzung zwischen Datenrepositorien und der zunehmende Einsatz entsprechender Datenmodelle (z.B. Graphen-Datenbanken, LOD) sowie Normdaten im Cultural Heritage-Bereich machen Wikidata zu einer interessanten Infrastruktur, die auch in diesem Gebiet eingesetzt werden könnte.

Kulturhistorische Datenbanken können beispielsweise die in ihnen enthaltenen Entitäten (historische Personen, Objekte, Orte etc.) auf Wikidata referenzieren, neue Zusammenhänge in Form von Aussagen an Wikidata zurückspielen und bestehende Aussagen aus Wikidata verwenden. Im Fall von Personen-Datensätzen enthält Wikidata beispielsweise Bezeichner, wie z.B. VIAF, GND oder der ULAN ID (Union List of Artist Names: Identifier von der Getty Union List von Künstlernamen). Wikidata stellt hier bereits ein internationales Bindeglied dar, da die überwiegend national (z.B. durch nationale Bibliotheken wie die DNB, BNF oder Library of Congress) oder disziplinär (z.B. ULAN für die Kunstgeschichte)

organisierten Normdatenrepositorien nicht alle Bezeichner abdecken bzw. auf diese verweisen. Im Bereich der Werke (Bauwerke, Kunstwerke und Artefakte aller Art) könnte diese übergreifende Rolle von Wikidata noch bedeutender werden. Anders als bei Personen existieren hier noch immer wenige Normdatenrepositorien. Zudem ist die Zahl der Artefakte wesentlich höher und ihre Definition viel weniger eindeutig (es kann von einer Zahl von mehreren -zig Millionen „relevanter“ Kunstwerke weltweit ausgegangen werden; die Zahl der für die verschiedenen kulturhistorischen Disziplinen relevanten Artefakte ist um ein Vielfaches höher). Es ist unwahrscheinlich, dass die betreffenden GLAM-Institutionen, d.h. kulturellen Institutionen und Gedächtnisorganisationen, hier in absehbarer Zeit ein ausreichendes und vor allem standardisiertes Datenmaterial bereitstellen können. Auch dort wo sich einschlägige Institutionen der Aufgabe angenommen haben, bleibt das entweder auf eine nationale Ebene beschränkt (z.B. mit den Datenbanken Merimee oder Joconde in Frankreich, oder dem RKD in den Niederlanden), oder droht unausgewogen und fragmentarisch zu bleiben (z.B. CONA, das Cultural Objects Name Authority des Getty Research Institute). Das Deutsche Dokumentationszentrum für Kunstgeschichte, Foto Marburg, hat zwar vielfach die Bedeutung von Werknormdaten unterstrichen (Locher/Warke 2013), jedoch bislang keinen Vorschlag für deren Bereitstellung gemacht. Nach heutigem Ermessen kann wohl auch nicht davon ausgegangen werden, dass es möglich oder sinnvoll ist, ein vollständiges Referenzrepositorium aller Bau- und Kunstwerke anzustreben. Selbst die Europeana (mit der Europeana ID) als zusammenfassender Aggregator ist hier noch lückenhaft und bietet vor allem keine Möglichkeit, weiteres Material direkt hinzuzufügen.

Wikidata kann hier eine bestehende Lücke schließen und die Funktion eines Drehkreuzes (Hub) erhalten, von dem aus auf weitere digitale Repräsentationen der Werke wie auch auf zugehörige Wissensbestände verwiesen wird. Freilich muss hierzu der Bestand in Wikidata – parallel zu den genannten Repositorien – ständig weiter ausgebaut werden. Projekte wie beispielsweise die „Sum of all paintings“ treiben diese Aufgabe systematisch voran (Poulter, 2017).

Für die praktische Arbeit im Bereich des Kulturerbes und der Kunstgeschichte bietet Wikidata nicht zuletzt den Vorteil der freien Zugänglichkeit und der verwendeten offenen technischen Standards. Selbst kleinste und sehr spezialisierte Projekte können so fehlende Gegenstände in Wikidata selbst ergänzen und sich damit zugleich in den globalen Wissensbestand einschreiben. Wohin sich Wikidata entwickeln wird, ist nicht absehbar. Es werden sich in jedem Fall viele weiteren Anwendungsgebiete herausbilden und ein Anwendungsfall könnte eine offene Metadatenbank für das Weltkulturerbe sein. Die bereits erwähnte Einschränkung der Manipulierbarkeit ist hierbei einerseits kritisch zu sehen, jedoch steht zu erwarten, dass das Zusammenspiel zahlreicher Statements einschließlich zugehöriger Referenzquellen sowie die Referenzierung auf

andere Normdaten einen sich selbst bestätigenden oder für automatisierte Überprüfungen immer besser geeigneten Datenbestand erzeugen wird.

Ausrichtung des Workshops

Das Ziel des Workshops liegt darin, eine fundierte Einführung in Wikidata zu geben (Weitere Informationen zum Ablauf und Materialien unter https://blogs.fu-berlin.de/dhd2018_wikidata-workshop/ verfügbar). Im Rahmen dieser Einführung wird den Teilnehmer_innen der Semantic Web Technologie-Stack (Triple, RDF, OWL, etc.) vermittelt und Wikidata in bestehende andere Initiativen (z.B. DBpedia) eingeordnet. In einem praxisorientierten Tutorial wird eine Einführung in Wikidatas Abfragesprache SPARQL gegeben. Die Teilnehmer_innen sollten nach dem Workshop ein grundlegendes Verständnis zu den Möglichkeiten des Einsatzes von Wikidata in ihrer Forschungspraxis erlangt und erste Möglichkeiten zum selbstständigen Einsatz kennengelernt haben. Dies wird sichergestellt, indem anhand mehrerer Fallbeispiele aus dem Bereich der kunsthistorischen Bilddatenbanken – unter anderem der Historischen Glasdias-Sammlung des Instituts für Kunst- und Bildgeschichte der HU Berlin – praktische Ansätze zur Nutzung von Wikidata in der geisteswissenschaftlichen Forschung aufgezeigt werden. Dabei werden Probleme diskutiert und exemplarische Lösungen durchgespielt werden. Hierzu gehören Fragen, wie vorhandene Datensätze zu Kulturgütern (automatisiert) aufgefunden, wie neue Entitäten möglichst effektiv ohne die Erzeugung von Dubletten angelegt und bestehende Aussagen abgefragt werden können. Zu den exemplarischen Lösungen gehören unter anderem ein Web-Service für die Ermittlung von Wikidata-IDs unter Verwendung der Google-Bildersuche und ein Skript zur Extraktion von Daten unter Nutzung der GND-Nummern. Nachfolgend werden die bisher geplanten Beispielanwendungen kurz eingeführt.

Beispielanwendung 1: Die Datenbank der digitalisierten Glasdias des Instituts für Kunst- und Bildgeschichte der Humboldt-Universität in Berlin. (Georg Schelbert, Humboldt-Universität zu Berlin; Martin Raspe, MPI f. Kunstgeschichte in Rom)

Fotografien oder Dias, die Kunstwerke zeigen, werden normalerweise erfasst, indem – neben einigen Metadaten zum Bild – das abgebildete Kunstwerk beschrieben wird. Dies erfolgt oft nach bestimmten Standards, die dann Pflichtfelder und Terminologien vorgeben (Harpring 2010). Allerdings bedeutet die Beschreibung der auf Dias befindlichen Kunstwerke einen im Grunde überflüssigen Aufwand, da es sich zumeist um bekannte oder sogar berühmte Kunstwerke handelt. Bei diesen Kunstwerken

ist anzunehmen, dass diese bereits beschrieben wurden. Daher wurde entschieden, die Fotografien oder Dias mit Normdaten-Identifikatoren aus Wikidata zu versehen. So muss z.B. zumindest ein Teil der Beschriftung entziffert oder das Bildmotiv erkannt werden, um dann in Wikidata gesucht werden zu können.

Um diesen Prozess zu beschleunigen bzw. zuverlässiger zu machen (indem zu eingegebenen Namen der Objekte zugleich Bilder zur Überprüfung der Übereinstimmungen angezeigt werden) wurde prototypisch ein Webservice entwickelt, der (mit einer entsprechend konfigurierten Google „custom search engine“) Bilder sucht. Er sucht dabei mit Vorrang bei Wikipedia, und ermittelt dazu die passenden Wikidata-IDs. Dieser Service soll im Rahmen des Workshops als Ausgangspunkt für praktische Arbeiten und weitere Diskussionen dienen.

Darüber hinaus stellt sich die Frage, wie Wissen, das aus dem Umgang mit den Bildern – etwa auf der Basis der Beschriftungen – entsteht, jedoch noch nicht Wikidata enthalten ist, in Wikidata eingespielt werden kann. Hierfür werden Möglichkeiten (potentielle Workflows und Ansätze) im Rahmen des Workshops diskutiert.

Beispielanwendung 2: Die ConedaKOR-Datenbank im Kunsthistorischen Institut in Frankfurt (Thorsten Wübbena, DFK Paris / Kunstgeschichtliches Institut der Univ. Frankfurt)

Aufgrund ihres Kenntnisstands sind Studierende in einer frühen Phase des Kunstgeschichtsstudiums häufig nicht in der Lage, über die textuelle Suche in einem Bilddatenbanksystem zu einem befriedigenden Ergebnis zu kommen. Sie wissen schlichtweg nicht, wonach sie suchen müssen bzw. welche Suchbegriffe sie verwenden müssen, um das entsprechende Resultat zu erhalten. Beim Einsatz von ConedaKOR hilft hier sicher die Option, die Daten visuell zu explorieren. So reicht es zum Beispiel, die aufbewahrende Institution zu kennen, um zu einem bestimmten Bildwerk zu gelangen. Mittels Relationen werden die übergeordneten Entitäten durch die diversen Räume bis hin zu den Werken sichtbar. Diese Möglichkeit stellt verständlich keine spezifische Besonderheit dar, macht aber deutlich, dass nur eine hohe Datendichte dieses Vorgehen wirklich sinnvoll macht. Um nun die eingeschränkten personellen Ressourcen im Frankfurter Kunstgeschichtlichen Institut zu unterstützen, wurde eine semi-automatisierte Lösung im Zusammenspiel mit dem Datenbestand aus Wikidata konzipiert. Da in der Frankfurter ConedaKOR-Installation die GND-ID für Personen und Orte abgelegt wird und genau diese Identifier auch in Wikidata vorhanden sind, werden die Relationen zwischen den – mit einer GND-ID versehenen – Entitäten in Wikidata mit denen in der Frankfurter KOR-Installation verglichen. Jede dort nicht vorhandene Relation wird eingetragen und mit

einem entsprechenden Label („Aussage aus Wikidata übernommen“) versehen. Neben diesem Verfahren wird mit einer speziell entwickelten Web-Extension noch eine weitere Möglichkeit der systemübergreifenden Nutzung von Wikidata vorgestellt. Hierbei wird zum einen auf jeder Website, die eine Wikidata-ID enthält im eigenen Datenbanksystem nachgesehen, ob Bildmaterial zum Objekt vorhanden ist, welches sich die Nutzer_innen dann direkt anzeigen lassen können. Ebenso kann auf diesem Weg neues Bildmaterial in das eigene System hochgeladen werden. Damit ist die Nutzung der Datenbasis in Wikidata möglich, fehlende Objekte können direkt dort ergänzt werden und die rechtlich häufig schwierige Abbildungssituation wird weiterhin im System der Nutzer_innen geklärt. Auch die Übernahme von Wikidata-Inhalten in die eigene Datenbank ist auf diesem Weg möglich.

Beitragende

Claudia Müller-Birn: Prof. Dr. rer. nat. Claudia Müller-Birn erforscht Fragestellungen im Bereich der Computer-Supported Cooperative Work und Social Computing. Ihr Ziel ist es, basierend auf einem besseren Verständnis der bestehenden Wissensprozesse neuartige Interaktionskonzepte zu entwickeln. Daher verbindet sie die Bereiche der Datenanalyse von Online Communities (z.B. Wikidata) mit dem Design von Kollaborationssoftware (z.B. Annotationssoftware neonion) eng miteinander. Ein zentraler Anwendungsbereich ihrer Forschung ist der Bereich der Scientific Collaboration, in welchen sie unter anderem durch Einsatz von Linked Data für eine nachhaltigere Nutzung von Forschungsdaten eintritt.

Georg Schelbert: Dr. phil. Georg Schelbert interessiert sich als Architektur- und Kunsthistoriker insbesondere für die Kunstgeschichte der Stadt Rom in der frühen Neuzeit, sowie für ihre Dokumentation in Karten und Veduten (Urban iconography). Ferner beschäftigt er sich als Leiter der Mediathek des Instituts für Kunst- und Bildgeschichte mit der Geschichte kunsthistorischer Bildmedien (Fotografien, Diapositive). Innerhalb der Digital Humanities beschäftigt er sich schwerpunktmäßig mit kulturhistorischen Bild- und Forschungsdatenbanken (Datenmodelle, linked data, Wissensrepräsentation, open heritage, Chronotopographie).

Martin Raspe: Dr. phil. Martin Raspe arbeitet generell über italienische Kunst und Architektur sowie niederländische Kunst der frühen Neuzeit. Außerdem beschäftigt er sich mit der Kunsttopographie der Stadt Rom von der Antike bis zur Gegenwart. In den Digital Humanities interessieren ihn Bild- und Forschungsdatenbanken (Datenmodelle, Graphdatenbanken, Wissensrepräsentation), Verfahren und Tools.

Thorsten Wübbena: Thorsten Wübbena M.A. arbeitet über Themen der bildenden Kunst und deren mediale

Rezeption (insbesondere in Musikvideos). Er arbeitet an Projekten zum Einsatz von Informationstechnologie in der kunstgeschichtlichen Forschung. Innerhalb der Digitalen Kunstgeschichte interessieren ihn insbesondere kulturhistorische Bild- und Forschungsdatenbanken (Datenmodelle, Wissensrepräsentation).

(urn:nbn:de:kobv:11-100209272 [letzter Zugriff 25. September 2017]

Bibliographie

Erxleben, Fredo / Günther, Michael / Krötzsch, Markus / Mendez, Julian / Vrandečić, Denny (2014): „Introducing Wikidata to the Linked Data Web“, in: *Proceedings of the 13th International Semantic Web Conference*. New York: Springer-Verlag 50-65.

Harpring, Patricia (2010): *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*, Los Angeles: Getty Publications.

Kohle, Hubertus (2013): *Digitale Bildwissenschaft*, Glückstadt: Hülbusch.

Locher, Hubert / Warnke, Martin (2014): Ergebnisse des Round Table „Kritische Massen – Zur Anschlussfähigkeit digitaler Bildbestände an die aktuelle Kunsthistorische Forschung“, in: *Kunstgeschichte. Open Peer Reviewed Journal* 07 Okt 2014 <http://www.kunstgeschichte-ejournal.net/412/> [letzter Zugriff 25. September 2017]

Ohlig, Jens / Schelbert, Georg (2017): „Datenpartnerschaften mit Wikidata – Projekt Durchblick“, *Wikimedia DE-Blog*, 21. Aug. 2017 <https://blog.wikimedia.de/2017/08/21/datenpartnerschaften-mit-wikidata-projekt-durchblick/> [letzter Zugriff 25. September 2017]

Müller-Birn, Claudia / Karran, Benjamin / Lehmann, Janette / Luczak-Rösch, Markus (2015): „Peer-production system or collaborative ontology development effort: what is Wikidata?“ in *Proceedings of OpenSym 2015 - Conference on Open Collaboration*. San Francisco: ACM.

Patton, Glenn E. (2010): *Funktionale Anforderungen an Normdaten: Ein konzeptionelles Modell* (IFLA Working Group on Functional Requirements and Numbering of Authority Records – FRANAR) Berlin, New York: De Gruyter.

Poulter, Martin (2017): „Wikidata – the new hub for cultural heritage“, *Wikimedia UK-Blog*, 20. Jan. 2017 <https://blog.wikimedia.org.uk/2017/01/wikidata-the-new-hub-for-cultural-heritage/> [letzter Zugriff 25. September 2017]

Vrandečić, Denny / Krötzsch, Markus (2014): „Wikidata: a free collaborative knowledge base“, in: *Commun. ACM* 57(10), 78–85.

Woitas, Kathi: „Bibliografische Daten, Normdaten und Metadaten im Semantic Web – Konzepte der Bibliografischen Kontrolle im Wandel“, in: *Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft*, 10.05.2013