

Digitale Sammlungserschließung mit WissKI und CIDOC CRM

Scholz, Martin

martin.scholz@fau.de
Friedrich-Alexander-Universität Erlangen-Nürnberg,
Deutschland

Wagner, Sarah

s.wagner@gnm.de
Germanisches Nationalmuseum, Deutschland

Die systematische Erfassung und wissenschaftliche Erschließung einer Sammlung sind grundlegende Voraussetzungen, um ihr wissenschaftliches Potential sichtbar zu machen. Häufig aber fehlen Software-Lösungen und Know-How für eine flächendeckende Digitalisierung und Online-Präsenz.

Dieser Workshop führt anhand praktischer Beispiele in die digitale Sammlungsarbeit mit WissKI und in die Modellierung mit dem CIDOC Conceptual Reference Model (CRM) ein. Durch die praktische Arbeit lernen die Teilnehmer die im Projekt „Objekte im Netz“ bereitgestellte Modellierung sowie die Konfiguration der Virtuellen Forschungsumgebung (VFU) für universitäre Sammlungen kennen.

Erschließung und Digitalisierung von Sammlungen

Neben Museen beherbergen auch Universitäten einen großen Schatz an Sammlungen, die der Wissenschaftsrat 2011 „als wertvolle Infrastruktur für [...] Forschung“ mit „beachtliche[m] wissenschaftliche[n] Potential“ identifiziert hat.¹ Allein in Deutschland existieren rund 1000 Sammlungen an über 80 Universitäten.² Zwar sind darunter auch renommierte Sammlungen, doch leidet das Gros an unzureichender Erschließung, Sichtbarkeit, Betreuung, Pflege oder Unterbringung.³ Auch bei der Digitalisierung gibt es enormen Aufholbedarf: Lediglich ein Drittel der Sammlungen sind digital zugänglich. Grund dafür sind u.a. auch das Fehlen von Software-Lösungen und Know-how für eine flächendeckende Digitalisierung und Online-Präsenz.

Seit einigen Jahren gibt es vermehrt Anstrengungen, universitäre Sammlungen aus ihrem Dornröschenschlaf zu wecken und sie zu einer wichtigen Ergänzung objektgebundener Forschung und Lehre weiter zu entwickeln. Dies drückt sich unter anderem in deutschlandweiten Förderprogrammen aus, wie etwa

der „Allianz für universitäre Sammlungen“ des Bundesministeriums für Bildung und Forschung. Das darin geförderte Projekt „Objekte im Netz“⁴ konzentriert sich auf die Digitalisierung universitärer Sammlungen und entwickelt in einer Kooperation zwischen der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) und dem Germanischen Nationalmuseum Nürnberg (GNM) eine gemeinsame Erschließungs- und Digitalisierungsstrategie für die Sammlungen der FAU, um die wissenschaftliche Nutzbarkeit der reichhaltigen Bestände zu verbessern.

Im Fokus stehen jedoch nicht einzelne Sammlungen oder Fachbereiche, sondern die Bereitstellung von Software-Werkzeugen und Lösungswegen, um die digitale Erschließung und Verfügbarkeit an breiter Front voranzutreiben. Die über 20 Sammlungen der FAU bilden dabei eine äußerst heterogene Entwicklungs- und Testlandschaft, um Lösungen zu erarbeiten, die über die FAU hinaus anwendbar sind. Die nötige Generizität der Ansätze und die Nachhaltigkeit sind daher zentrale Herausforderungen, wobei bei letzterem die langfristige Interpretierbarkeit der Daten im Blickpunkt des Projekts steht. Daneben müssen die meist knappen personellen und finanziellen Mittel berücksichtigt werden.

Als besonders geeignet zur Umsetzung der Ziele erscheinen auf technischer Seite Lösungen, die unter freien Lizenzen (Open Source) zur Verfügung stehen und die Ideen des Semantic Web implementieren: Flexible Wissensnetze mit klar definierter Semantik, die weltweit – und damit auch sammlungsübergreifend – verknüpft werden können. Das Projekt erweitert daher die virtuelle Forschungs- und Dokumentationsumgebung WissKI⁵ zu einem Werkzeug für die digitale Sammlungserschließung und stellt auf verschiedene Sammlungsbereiche abgestimmte Konfigurationen der Software sowie Leitfäden zur Verfügung. Für die standardisierte semantische Auszeichnung der Daten kommt das CIDOC Conceptual Reference Model (CRM) zum Einsatz.

Aufgrund einer erfolgreichen Pilotstudie kann das Projekt bereits auf erste Ergebnisse verweisen. Im Rahmen des Workshops wird die bereits publizierte generische Konfiguration vorgestellt und von den Teilnehmern angewandt.

Die Werkzeuge - WissKI und CIDOC CRM

Semantische Technologien - im Speziellen Semantic Web und Linked Open Data - erfreuen sich zunehmender Beliebtheit in den Digital Humanities. Für objektbasierte Forschung bieten die flexible, netzwerkartige Grundstruktur des Resource Description Framework (RDF) und darauf aufbauende Formate ein adäquates Mittel zur Repräsentation, Verwaltung und Publikation von (Meta-)Daten. Zahlreiche VFUs unterstützen das Erstellen von komplexen Wissensnetzen

und deren Export in Tripelformaten. Wichtige Normdateien und Thesauri stehen als Linked Open Data zur Verfügung. Ontologien wie das CIDOC CRM bilden das semantische Rückgrat dieses Ansatzes und garantieren ein Mindestmaß an Interoperabilität und Datenaustausch, das über das klassische Verlinken von Web-Dokumenten hinausgeht.

Wenngleich die Nutzung semantischer Technologien zunimmt, stellt der praktische Einsatz unerfahrene oder wenig technikaffine Nutzer meist vor große Herausforderungen. Dies gilt weniger für die Beherrschung bestimmter Formate und Werkzeuge als vielmehr für die semantische Modellierung der Daten, d.h. die Erstellung von und den richtigen Umgang mit Ontologien. Da hierbei die Bedeutung der Daten formalisiert niedergelegt wird, ist mitunter ein gehöriges Maß an Wissen über einen Anwendungs-/Fachbereich erforderlich, um Modellierungsfehler zu vermeiden und so eine spätere korrekte Interpretation zu gewährleisten. Insbesondere das CIDOC CRM⁶, das eine Top-Level-Ontologie für die Dokumentation kulturellen Erbes darstellt, steht immer wieder in der Kritik, für Einsteiger zu komplex zu sein.

Die virtuelle Forschungs- und Dokumentationsumgebung WissKI nimmt sich dieser Herausforderung an. Die browserbasierte Software ist das Produkt aus zwei DFG-geförderten Projekten und entstand aus Anforderungen an die kooperative Forschung in Museen bzw. im Bereich des Kulturerbes und seiner Dokumentation im digitalen Medium. Zentraler Fokus von WissKI ist das vernetzte Arbeiten auf Basis semantischer Tiefenerschließung von Forschungsdaten. Eine Schlüsselrolle kommt hierbei dem CIDOC CRM zu, das um projektspezifische Anwendungsontologien erweitert werden kann.

Aus Nutzersicht ist das System an die tradierten Formen der Datenakquise und -präsentation angelehnt. Die Daten werden jedoch semantisch aufbereitet und nativ als RDF mitsamt Ontologie-Konstrukten gespeichert. Dem Nutzer werden so die Vorteile von Linked Open Data und Semantic Web zugänglich, ohne dass dieser sich mit technischen und ontologischen Details auseinandersetzen muss. Kern dieses Ansatzes ist eine Abbildung zwischen den tradierten, meist datensatz-basierten, tabellarischen Darstellungen und der graphbasierten Wissensrepräsentation, die die ontologiegestützte, formale Semantik der verwendeten Datenfelder beinhaltet. Diese Abbildung wird von einem inhaltlichen Administrator festgelegt und ist für die Nutzer standardmäßig nicht sichtbar. Die formale Semantik muss also nicht verstanden werden, um das System effektiv zu nutzen. Abbildungen oder Teile davon können zwischen verschiedenen Systemen wiederverwendet und erweitert werden, so dass sich Best-Practice-Modellierungen herausbilden.

Die Open-Source-Lizenzierung aller in diesem Workshop verwendeten Werkzeuge und Standards ist ein wichtiger Aspekt. Die kostenfreie Nutzung trägt zum einen der häufig angespannten finanziellen Situation universitärer Sammlungen Rechnung und ist zum anderen Bestandteil des partizipativen Konzepts: Anwender können die

Materialien nutzen, sie an ihre Bedürfnisse anpassen und wiederum der Community zur Verfügung stellen.

Zielgruppe sowie Inhalt und Ziele des Workshops

Der Workshop richtet sich an alle, die mit Sammlungsobjekten oder mit Objekten des kulturellen Erbes im Allgemeinen arbeiten und diese digital dokumentieren oder erschließen. Auch spricht der Workshop interessierte Wissenschaftler an, die Objekte standardisiert dokumentieren und ihre Metadaten semantisch aufbereiten möchten. Es werden von den Teilnehmern keine Vorkenntnisse für die VFU WissKI oder das CIDOC CRM vorausgesetzt.

Der Workshop zeigt anhand praktischer Beispiele, wie Erfassungsschemata und -modi aus der universitären Sammlungslandschaft mithilfe der Referenzontologie CIDOC CRM und der VFU WissKI auf Objekte universitärer Sammlungen bzw. des kulturellen Erbes im Allgemeinen umgesetzt werden können.

Während des Workshops arbeiten die Teilnehmer mit ihrem eigenen WissKI-System, wahlweise einzeln oder in Kleingruppen. Dabei stehen weniger die informationstechnischen Details der Werkzeuge im Vordergrund. Vielmehr werden die nötigen Schritte bis zum effektiv einsetzbaren System vermittelt und durchgeführt. Angefangen bei der Installation und einigen grundlegenden Funktionalitäten, binden die Teilnehmer die vom Projekt „Objekte im Netz“ angebotene Konfiguration zur Sammlungserschließung in WissKI ein und erhalten somit ein einsetzbares System mit standardisierten Eingabe- und Anzeigemöglichkeiten. Darauf aufbauend werden Möglichkeiten der einfachen Anpassung der semantischen Modellierung aufgezeigt und selbständig geübt. Das Erfassen von (selbst mitgebrachten) Datensätzen rundet die praktische Einführung ab.

Neben einer allgemeinen Einführung in das Arbeiten mit WissKI und der semantischen Dokumentation von Daten sind die Teilnehmer nach dem Workshop in der Lage, einfache Erfassungsmasken zu modellieren, Daten mit WissKI zu erfassen und zu recherchieren.

Kurzbiographien

Martin Scholz ist einer der Hauptentwickler der Virtuellen Forschungsumgebung WissKI. Er studierte Informatik und Sinologie an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Nach seinem Diplom 2008 arbeitete er für die Arbeitsgemeinschaft Digital Humanities der FAU für das DFG-geförderte Projekt „Wissenschaftliche Kommunikationsinfrastruktur“ (WissKI). Seit 2017 engagiert er sich für die Digitalisierung der Sammlungen der FAU im Rahmen des BMBF-geförderten Projekts

„Objekte im Netz“. Seine Forschungsinteressen liegen in den Digital Humanities, insbesondere in den Bereichen Wissensrepräsentation, Semantic Web und Verarbeitung natürlicher Sprache.

Martin Scholz

Friedrich-Alexander-Universität

Erlangen-Nürnberg

Referat H2 – Zentralkustodie

Hugenottenplatz 1a, 91054 Erlangen

martin.scholz@fau.de

Sarah Wagner ist Kunsthistorikerin und arbeitet seit 2012 in der Abteilung für Kulturinformatik am Germanischen Nationalmuseum Nürnberg. Sie studierte Kunstgeschichte und Museumsarbeit in Bamberg, Erlangen und Leiden (NL) und betreut seit 2014 verschiedene Forschungsprojekte, die mit WissKI arbeiten. Aktuell ist sie für das BMBF-geförderte Kooperationsprojekt „Objekte im Netz“ tätig und vertritt dort die Seite des Museums. Ihre Forschungsschwerpunkte liegen in der frühneuzeitlichen Sammlungspraxis und der semantischen Wissensmodellierung.

Sarah Wagner

Germanisches Nationalmuseum Nürnberg

Kornmarkt 1, 90402 Nürnberg

s.wagner@gnm.de

Fußnoten

1. Vgl. „Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen“, Verfügbar unter: <https://www.wissenschaftsrat.de/download/archiv/10464-11.pdf>) [letzter Zugriff: 10. Januar 2018]
2. Kennzahlen zu den folgenden Aussagen sind verfügbar unter: <https://portal.wissenschaftliche-sammlungen.de/kennzahlen/> [letzter Zugriff: 10. Januar 2018]
3. siehe Fußnote 1
4. Das Projekt wird vom Bundesministerium für Bildung und Forschung von 2017 bis 2020 im Rahmen der Förderlinie „Vernetzen - Erschließen - Forschen. Allianz für universitäre Sammlungen“ gefördert. Mehr Informationen unter URL: <http://objekte-im-netz.fau.de/> [letzter Zugriff: 10. Januar 2018]
5. WissKI (= „Wissenschaftliche Kommunikationsinfrastruktur“, URL: [letzter Zugriff: 10. Januar 2018]) basiert auf dem Open-Source Content Management System Drupal (URL: [letzter Zugriff: 10. Januar 2018]) und wurde in Zusammenarbeit zwischen dem Germanischen Nationalmuseum, Nürnberg, dem Zoologischen Forschungsmuseum Alexander Koenig, Bonn und der Friedrich-Alexander-Universität Erlangen-Nürnberg entwickelt.
6. Das CIDOC CRM wurde vom International Committee for Documentation als Teil des International Council of Museums (ICOM) als formale Referenzontologie erarbeitet und ist seit 2006 als ISO Norm (ISO 21127) anerkannt. In der „Erlangen CRM“ (URL: <http://erlangen-crm.org/> [letzter Zugriff: 10. Januar 2018]) auf

Basis der Web Ontology Language (OWL) liegt eine maschinenlesbare Version vor. Weitere Informationen zum CIDOC CRM unter URL: <http://cidoc-crm.org/> [letzter Zugriff: 10. Januar 2018].

Bibliographie

Definition of the CIDOC Conceptual Reference Model: Version 5.0.4., autor. durch die CIDOC CMR Special Interest Group (SIG), 2011. http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf) [letzter Zugriff 25.09.2017].

Görz, Günther: „WissKI: Semantische Annotation, Wissensverarbeitung und Wissenschaftskommunikation in einer virtuellen Forschungsumgebung“ in: Kunstgeschichte, Open Peer Reviewed Journal, urn:nbn:de:bvb:355-kuge-167-7 [letzter Zugriff 10.01.2018].

Hohmann, Georg (2011): „Die Anwendung von Ontologien zur Wissensrepräsentation und -kommunikation im Bereich des Kulturellen Erbes“ in: Schomburg, Silke u.a. (eds.): Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland. Köln: Hochschulbibliothekszentrum NRW 33-39.

Hohmann, Georg / Schieman, Bernhard (2013): „An Ontology-Based Communication System for Cultural Heritage. Approach and Progress of the WissKI Project“ in: Hans Bock u.a. (eds.): Scientific Computing and Cultural Heritage. Berlin: Springer 127-135.

Hohmann, Georg/Fichtner, Mark. Chancen und Herausforderungen in der praktischen Anwendung von Ontologien für das Kulturerbe. In: Digitales Kulturerbe. Bewahrung und Zugänglichkeit in der wissenschaftlichen Praxis. Vol. Kulturelle Überlieferung – digital. Karlsruhe 2015. S. 115-128.

Wissenschaftsrat (2011): *Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen*. Berlin <https://www.wissenschaftsrat.de/download/archiv/10464-11.pdf>) [letzter Zugriff 25.09.2017].