

Digitalisierung der Universitätssammlungen der FAU Erlangen-Nürnberg

Auf dem Weg zum Semantic Web in Eigenregie

Martin Scholz und Udo Andraschke

(vorname.nachname@fau.de)

Die Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) besitzt über 20 Sammlungen aus den verschiedensten Fachbereichen. [1] Nicht minder verschieden gestalten sich Grad und Umfang ihrer Erfassung und Digitalisierung. Mit der Einrichtung einer Zentralkustodie im Jahr 2011, die die Aktivitäten und Ausrichtung der Sammlungen bündeln und sie als wichtige wissenschaftliche Infrastrukturen weiter ausbauen sollte, wurde auch das Ziel formuliert, die digitale Datenerfassung und Präsentation der Sammlungen voranzutreiben. Von zentraler Bedeutung sind dabei gemeinsame Erfassungsstandards und -formate sowie eine gemeinsame Software-Lösung und Webpräsenz.

Die Wahl der geeigneten Software-Infrastruktur fiel bewusst auf die Virtuelle Forschungsumgebung WissKI (wiss-ki.eu) [4], da sie

- a) unter einer Open Source-Lizenz verfügbar ist (GPL),
- b) konsequent auf offenen Standards und Formaten aufbaut und
- c) an der FAU mitentwickelt wird.

WissKI wird seit 2009 von der Arbeitsgruppe Digital Humanities des Departments für Informatik der FAU in Kooperation mit dem Germanischen Nationalmuseum in Nürnberg sowie dem Zoologischen Forschungsmuseum Alexander Koenig in Bonn als web-basiertes Content Management System für die Dokumentation von Kulturerbe im musealen und wissenschaftlichen Kontext entwickelt und befindet sich in den genannten Institutionen bereits im Einsatz. WissKI bietet den Nutzern gewohnte Eingabe- und Präsentationsschnittstellen, wie etwa feldbasierte Eingabemasken oder die Möglichkeit zum Freitext. Die Daten werden jedoch im Hintergrund nativ auf Basis von Semantic Web-Technologien (Ontologien, RDF [2]) erfasst. Dies ermöglicht auch technisch ungeschulten Nutzern das Einpflegen hoch vernetzter Datenbestände – sowohl lokal als auch global – und gleichzeitig das Erfassen der Bedeutung der Daten, um deren Interpretierbarkeit auf lange Zeit zu sichern. Dabei schreibt das System keine ontologischen Kategorien vor, sondern kann innerhalb des jeweiligen Anwendungsbereichs frei angepasst werden. WissKI ist nicht als zentraler Webdienst konzipiert, vielmehr kann die Software kostenlos heruntergeladen und auf einem Server als an die eigenen Bedürfnisse angepasste WissKI-Instanz eingesetzt werden.

Als fachübergreifende, verbindende Ontologie – eine sog. Referenzontologie – wurde der offene Standard CIDOC CRM [3] (ISO 21127) bzw. die OWL DL-Implementation Erlangen CRM (erlangen-crm.org) [5] gewählt, da das CIDOC CRM

- a) speziell auf die Dokumentation von Kulturerbe ausgerichtet ist und
- b) als international anerkannter Standard Sicherheit in Langzeitfragen gibt.

Die Referenzontologie garantiert zum einen ein Mindestmaß an fachübergreifender Interpretierbarkeit der Daten durch die Definition grundlegender Klassifikationsstrukturen und ermöglicht zum anderen die modulare Erweiterung um

fachspezifische Begrifflichkeiten.

Zur Umsetzung des angezeigten Vorhabens wurde das Pilotprojekt *WissKI@Sammlungen der FAU* [6] ins Leben gerufen. Partner sind neben der Zentralkustodie und der AG Digital Humanities drei ausgesuchte Universitätssammlungen: das Herbarium Erlangense, die Informatiksammlung Erlangen sowie die Schulgeschichtliche Sammlung. Die beteiligten Sammlungen spiegeln die oben genannte Heterogenität in hohem Maße wieder, so dass die unterschiedlichen Eigenarten und Bedürfnisse der Sammlungen der FAU weitgehend repräsentiert sind.

Das Pilotprojekt hat experimentellen Charakter. Es soll WissKI für den Einsatz in den Sammlungen erproben und einen Migrationspfad für die gesamten Universitätssammlungen entwickeln. Ein wichtiger Teilaspekt ist dabei der Transfer der Daten aus den bestehenden Datenbanksystemen in zuvor eingerichtete WissKI-Instanzen. Dennoch versteht sich das Vorhaben nicht als rein technikgetrieben, sondern sieht die Software als ein Instrument zum Ausbau der Sammlungen zu Forschungsinfrastrukturen.

Das Projekt startete im November 2013 ohne größere finanzielle Ausstattung. Treibende Kraft war und ist das Eigeninteresse der beteiligten Partner. Seit Mitte 2014 wird das Projekt durch eine studentische Hilfskraft unterstützt. Aufgrund der räumlichen Nähe aller Projektbeteiligten haben sich Workshops in regelmäßigen Abständen als Arbeitsmodus bewährt. Von großer Bedeutung ist dabei der gegenseitige Austausch, sowohl zwischen den Sammlungen untereinander als auch zwischen den Sammlungen und der Informatik, repräsentiert durch die AG Digital Humanities. Parallel dazu wurde eine WissKI-Instanz als Sandbox zum Üben eingerichtet und mit einem Forum und Wiki ausgestattet, so dass bspw. auch Tutorien erstellt und gemeinsam bearbeitet werden können. In der ersten Phase des Projekts (von Ende 2013 bis Mitte 2014) wurden monatliche Arbeitstreffen mit allen Projektteilnehmern anberaumt, um Themen zu behandeln, die alle Sammlungen angehen und um eine gemeinsame Wissensbasis zu schaffen. Nach ca. 6 Monaten wurden ergänzend Treffen zwischen der AG Digital Humanities (IT) und je einer Sammlung (Anwender) eingeführt, um den Spezifika der einzelnen Sammlungen besser Rechnung zu tragen.

Das Projekt wurde aufgrund der inhaltlichen Aufgaben in zwei Phasen unterteilt: Die bereits abgeschlossene Phase 1 beinhaltet alle Maßnahmen bis zum Transfer der Daten nach WissKI, die sich in fünf Schritte gliedern lassen:

1. Vertrautmachen mit der (Semantic) Web-Technologie und der WissKI-Infrastruktur
2. Identifikation von Gemeinsamkeiten und Unterschieden der bereits vorhandenen Daten und Datenbankschemata
3. Erstellen bzw. Erweiterung der Domänenontologie auf Basis des CIDOC CRM
4. Definition der Eingabemasken und Datenfelder und entsprechende Konfiguration der WissKI-Software
5. Iteration der Punkte 2-4

Phase 2 widmet sich dem Datentransfer und dem Aufbau eines gemeinsamen Portals in folgenden Schritten:

6. Definition der Abbildungsvorschriften zwischen bestehenden Datenbanken und Domänenontologie
7. Import von Testdaten aus den bestehenden Datenbanken

8. Test und Korrektur des vorgenommenen Imports
9. Iteration der Punkte 6-8
10. Einbinden bzw. Erstellen von Normdaten
11. Einbinden des Datenbestandes in ein gemeinsames Präsentationsportal

Alle Schritte wurden und werden begleitend in Form von Tutorien dokumentiert. Sie spiegeln Erfahrungen, Diskussionen und Best Practices wieder und bilden einen wichtigen Eckpfeiler für die Migration weiterer Sammlungen der FAU. Sie stellen in ihrer Gesamtheit eine stetig wachsende Gebrauchsanweisung für den Einsatz von WissKI auf der einen und Leitfaden zur semantische Modellierung von Sammlungen auf der anderen Seite dar. Die Tutorien sind öffentlich zugänglich und können auch Dritten außerhalb der FAU als Leitfäden dienen. [7]

Zum jetzigen Zeitpunkt (November 2014) ist Phase 1 abgeschlossen, Phase 2 befindet sich noch in der Umsetzung.

Im Vortrag soll daher auf Phase 2 nicht näher eingegangen werden. Vielmehr sollen Phase 1 analysiert und anhand von Beispielen einige der Hindernisse und Chancen des Vorgehens hervorgehoben werden:

1. Das Projektformat mit regelmäßigen Workshops und die aktive Einbindung von Sammlungsmitarbeitern setzt deren Bereitschaft voraus, sich eingehend mit aus Sammlungssicht meist fachfremden Methoden und Techniken auseinanderzusetzen. Im Bereich des Semantic Web handelt es sich zudem um ein relativ neues und dynamisches Gebiet der Informatik, das nicht mit jahrzehntelanger Erfahrung und entsprechend ausgereiften Werkzeugen aufwarten kann wie etwa relationale Datenbanksysteme. Naturgemäß bildet auch der Zeitaufwand (und damit indirekt die personellen Kapazitäten einer Sammlung) eine Hürde.
2. Im Gegenzug festigt das Format bei den Sammlungsmitarbeitern das Verständnis und die Akzeptanz für die eingesetzten Technologien und Methoden. Das Eigeninteresse der Sammlungen wird klar erkennbar, was in den Augen der Autoren die erfolgreiche Umsetzung des Vorhabens trotz begrenzter Mittel entscheidend begünstigt.
3. Nicht zuletzt bauen die Sammlungen über die beteiligten Mitarbeiter Kompetenz im Bereich IT und semantischer Modellierung auf. Die Sammlungen können sich im Idealfall untereinander austauschen und mithin gegenseitig helfen und unterstützen. Realistischerweise ist dies im Pilotprojekt bei einfacheren Fragestellungen der Bedienung und Modellierung gegeben.
4. Die ontologische Modellierung und der fächerübergreifende Charakter der Workshops führen zu einer vertieften Reflexion über die eigene Sammlung und den sammlungs- bzw. fächerübergreifenden Kontext. So wurde bspw. das Bewusstsein für fachspezifische Termini bei gleichzeitiger Notwendigkeit für gemeinsame Terminologien gestärkt. Das Auftreten teils unerwarteter Überschneidungen in den einzelnen Disziplinen ermöglichte u.a. auch eine genauere Definition sonst kaum weiter hinterfragter Begrifflichkeiten.
5. Das CIDOC CRM als Referenzontologie bietet hier einen guten Ausgangspunkt, um gemeinsame Strukturen und Prozesse trotz unterschiedlicher Fachbegriffe herauszuarbeiten und deren Bedeutung klar zu formulieren. Andererseits können die in der Referenzontologie vorgegebenen Strukturen zu zunächst eigenwilligen Ergebnissen führen. Die starke Betonung von Ereignissen im CIDOC CRM steht zum Beispiel im scheinbaren Widerspruch zur objektzentrierten Dokumentation vieler Sammlungen und erfordert teilweise ein Überdenken tradierter Muster und eine Korrektur institutionalisierter Denkgewohnheiten. Dies kann wiederum die oben genannte Reflexion anregen.

6. Durch die Analyse der Datenstrukturen werden darüber hinausgehende Gemeinsamkeiten sichtbar. Grundlegende Herausforderungen wie die Einbindung und Verwendung gemeinsamer Normdaten und die Sicherung der Datenqualität können benannt, diskutiert und einheitlich angegangen werden.

Nach Abschluss der ersten Projektphase kann somit resümiert werden, dass die Umsetzung des vorliegenden Vorhabens – der einheitlichen Digitalisierungen der Sammlungen der FAU in Eigenregie – zwar einen deutlichen Einsatz von den Sammlungen selbst einfordert, sich aber Mehrwerte ergeben haben, die für sie mittel- und langfristig von Vorteil sind.

Literatur & Internetseiten:

- [1] U. Andraschke und Marion Ruisinger, Die Sammlungen der Universität Erlangen-Nürnberg, 2007.
- [2] G. Antoniou, P. Groth und F. van Harmelen, A Semantic Web Primer, MIT Press, 2012.
- [3] Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, Matthew Stiff (Hrsg.), Definition of the CIDOC Conceptual Reference Model, 2011.
- [4] M. Scholz und G. Goerz, WissKI: A Virtual Research Environment for Cultural Heritage. In Proceedings of ECAI. 2012, 1017-1018.
- [5] <http://erlangen-crm.org> (aufgerufen am 09.11.2014)
- [6] <http://wisski.cs.fau.de/sammlungen> (aufgerufen am 09.11.2014)
- [7] <http://wisski.cs.fau.de/sammlungen/tutorials> (aufgerufen am 09.11.2014)