

Darstellung heterogenen und dynamischen Wissens mit CIDOC CRM und WissKI

Scholz, Martin

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

Goerz, Guenther

guenther.goerz@fau.de
Friedrich-Alexander Universität Erlangen-Nürnberg,
Deutschland

Wagner, Sarah

s.wagner@gnm.de
Germanisches Nationalmuseum Nürnberg

Fichtner, Mark

m.fichtner@wiss-ki.eu
Germanisches Nationalmuseum Nürnberg

Dieser Vortrag stellt die praktische Arbeit mit WissKI und die ontologische Modellierung mit dem CIDOC-CRM anhand zweier Use Cases vor. Dabei steht der dynamische Umgang mit heterogenen Daten im Fokus.

Die WissKI-Software

Die virtuelle Forschungsumgebung "WissKI" (Wissenschaftliche Kommunikations-Infrastruktur, <http://wiss-ki.eu/>) entstand aus Anforderungen an die kooperative Forschung im Bereich des Kulturerbes und seiner Dokumentation im digitalen Medium. Im Rahmen des DFG-geförderten, gleichnamigen Projekts WissKI wurde die Softwareplattform auf der Basis des Open-Source Content Management Systems Drupal (<http://drupal.org>) in Zusammenarbeit zwischen dem Germanischen Nationalmuseum Nürnberg (GNM), dem Zoologischen Forschungsmuseum Alexander Koenig in Bonn (ZFMK) und der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) entwickelt. Fokus des Systems sind neben der einfachen Bereitstellung und offenen Verfügbarkeit von Quellmaterialien – strukturierten Texten, Grafiken, Bildern, Video, Audio – und Metadaten in digitaler Form, auch das interaktive und vernetzte Arbeiten auf der Basis semantischer Tiefenerschließung. Indem die typischen Eigenschaften von gängigen Content Management Systemen unberührt bleiben, verfügt das System über eine detaillierte Nutzersteuerung mit Rechteverwaltung und ist in der Lage Web-Inhalte wie

Webseiten, Foren und Wikis zu verwalten und online zu präsentieren.

Die Software ist als Open-Source kostenfrei unter <http://github.com/WissKI> (Görz et al. 2009-*) veröffentlicht und kann dementsprechend nachgenutzt und erweitert werden.

Semantische Modellierung mit dem CIDOC-CRM

Bei der semantischen Modellierung kommt dem Conceptual Reference Model (CIDOC-CRM, ISO 21127) von ICOM-CIDOC als formaler Referenzontologie eine Schlüsselrolle zu. Unsere Implementation in der Web Ontology Language bildet in Verbindung mit verschiedenen Werkzeugen des Semantic Web die Grundlage des WissKI-Systems. Auch diese Softwarekomponente ist als Open-Source kostenfrei verfügbar und kann unter <http://erlangen-crm.org> heruntergeladen werden.

Die Datenakquisition im WissKI-System erfolgt primär über karteikartenähnliche Formulare oder Freitextfelder, beides tradierte Formen der Datenerfassung in den vorliegenden Anwendungsbereichen. Die Daten werden vom System im Hintergrund jedoch durch das CIDOC-CRM semantisch erschlossen, d. h. neben den Daten wird auch ihre ontologiebezogene Bedeutung für Mensch und Maschine lesbar gespeichert. Hierzu bedient sich WissKI konsequent der Techniken des Semantic Web wie RDF und OWL-DL und zielt auf Veröffentlichung der Daten als Linked Open Data. Die eingegebenen Daten bilden einen Wissensgraphen, der auf einfache Weise weltweit und (potentiell) transdisziplinär vernetzt werden kann. Die formale Ontologie, die über eine hierarchische Struktur von Konzepten und Eigenschaften und einer standardisierten logischen Sprache ein System von Fachbegriffen bildet, dient der Erfassung und semantischen Erschließung der Daten. Je nach Fachgebiet und Sammlungsschwerpunkt kann das CIDOC-CRM als grundlegende Ontologie um fachspezifische Begriffe mit Anwendungsontologien erweitert werden. Diese flexible Art der Wissensspeicherung lässt schnell und unproblematisch Änderungen des Datenmodells am laufenden System zu. Durch die Einbindung einer standardisierten logischen Sprache und die Nutzung von Open-Source-Software ermöglicht WissKI eine hohe Anschlussfähigkeit, leichte Zugänglichkeit und Langzeit-Interpretierbarkeit der Daten.

Exemplarische Anwendungsszenarien

EDEN

Die Epigraphische Datenbank Erlangen-Nürnberg (EDEN, Dreyer et al. 2014-*) ist eine Online-Datenbank für antike Inschriften aus Kleinasien. Momentan beschränkt sie sich auf die drei antiken griechischen Siedlungen Metropolis in Ionien, Magnesia am Mäander und Apollonia am Rhyndakos. Inhaltlich wird die Datenbank seit 2012 von Boris Dreyer (Professur für Alte Geschichte, FAU) in enger Zusammenarbeit mit Archäologen der FAU und dem Lehrstuhl für Graphische Datenverarbeitung sowie archäologischen Kollegen in der Türkei gepflegt und weiterentwickelt. Ziel ist die Schaffung eines effektiven Werkzeugs zur interdisziplinären und internationalen Forschung, das Forscher, Studenten und die breite Öffentlichkeit im Spannungsbereich zwischen Alter Geschichte und Archäologie nutzen können. Während sich Althistoriker primär mit den immateriellen Eigenschaften der Inschrift, also Textinhalt und -form, beschäftigen, ist für Archäologen der materielle Träger der Inschrift von größerem Interesse. Die Verknüpfung und gemeinsame Präsentation von Daten für beide Disziplinen bildet ein Novum in diesem Bereich. Daher waren von Anfang an web-basiertes, kollaboratives Arbeiten sowie flexible Datenhaltung wichtige Voraussetzungen der technischen Infrastruktur.

Der Fokus der Datenbank ist (noch) nicht das Bereitstellen einer großen Anzahl an Inschriften, sondern die Angabe detaillierter Metadaten aus verschiedenen Disziplinen: zu den edierten Inschriften kommen zahlreiche Metadaten in textueller und tabellarischer Form, u. a. Funddaten, wissenschaftliche Kommentare, Übersetzungen und hochauflösende Bilder. Diese geben wertvolles Wissen für Historiker und Archäologen wieder. Entsprechend den Entwicklungen durch Kooperationen und neue Forschungsschwerpunkte rücken frühere Randinformationen schnell in den Fokus der Datenbank und werden sukzessive verfeinert und ergänzt. Ebenso werden weiterführende Informationen zu wiederkehrenden Themen wie bestimmten Herrschern, Genres oder Orten laufend hinzugefügt. Die Einbindung von 3D-Modellen soll die Datenbank mittelfristig zu virtuellen Ausgrabungsstätten erweitern.

Sammlung Rück

Seit 2015 widmen sich Forscher am GNM einem einzigartigen Bestand an Musikinstrumenten. Die Sammlung Rück war die größte deutsche Sammlung historischer Musikinstrumente in Privatbesitz und wurde 1880 vom Pianisten und Lehrer Wilhelm Rück gegründet. Nach seinem Tod setzten sich seine Söhne den systematischen Ausbau der Sammlung und die Dokumentation der Entwicklung abendländischer Musikinstrumente zum Ziel. Eine seit 1929 akribisch geführte Dokumentation aller Sammlungsaktivitäten ist Ulrich Rück, dem Sohn des Sammlungsgründers, zu verdanken. Diese Korrespondenz mit über 1000 Partnern umfasst 17.200 Briefe und Postkarten, die zusammen mit

1.500 Musikinstrumenten und weiterem Material 1962 vom GNM erworben wurden. Alle Dokumente werden seither im Historischen Archiv, die Objekte selbst im Sammlungsreferat für Musikinstrumente aufbewahrt.

Die Korrespondenz zum Aufbau der Sammlung Rück bietet heute eine einzigartige Quelle, Einblicke in das Handeln mit Musikinstrumenten und in die Entstehung einer Sammlung von Objekten des Kulturlebens in der Zeit zwischen Weltwirtschaftskrise und Wirtschaftswunder gewinnen zu können. Besondere Bedeutung kommt dem Schriftverkehr mit Gutachtern bezüglich Qualität und Marktwert der Instrumente zu, da hieraus ermöglicht wird, einen Preisspiegel für den Handel mit historischen Musikinstrumenten der damaligen Zeit zu erstellen. Dieser soll zum Abschluss des Projekts gemeinsam mit Recherche-Ergebnissen zu den Erwerbsumständen und unmittelbaren Vorbesitzern online zugänglich gemacht werden. Weitere wichtige Aspekte der Untersuchung sind Sammlungs-, Kommunikations- und Marketingstrategien, die zu einer europaweiten Vernetzung der Sammlung führten, und die in einer Monographie dargestellt werden.

Zentrale Herausforderung des Nachlasses Rück ist die Heterogenität der Materialien. So sind neben den Archivgütern, die inhaltlich tiefer erschlossen werden sollen, auch Objekt-, Personen- und Literaturdaten mit entsprechenden Kreuzverweisen zu erfassen. Auf der Basis dieser Verweise soll die Abfolge der Kommunikation mit den verschiedenen Kommunikationspartnern, die geschäftlichen Reisen der Familie Rück, Erwerbungen in ihren zeitlichen und räumlichen Beziehungen und der generelle Kontext zur damaligen Zeit dargestellt werden.

Der Großteil der Archivmaterialien aus dem Nachlass Rück ist mit Schreibmaschine geschrieben und deshalb gut lesbar. Eine händische Transkription der Materialien ist nicht notwendig. Die Erschließung erfolgt auf Basis von hochqualitativen Digitalisaten der Briefe, die anschließend die Wissenschaftler inhaltlich zusammenfassen. In diesen Zusammenfassungen werden entsprechende Referenzierungen zu den anderen Materialien des Projektes vorgenommen. Hierfür ist eine Verlinkung und idealerweise eine Auszeichnung aus dem Fließtext heraus notwendig. Gleichzeitig wird ein System benötigt, das die verknüpften Materialien geeignet präsentieren kann.

Praktische Umsetzung

EDEN

Durch die Einbindung von Drupal bringt WissKI einerseits die nötigen Voraussetzungen für web-basiertes, kollaboratives Arbeiten mit. Der generische Aufbau ermöglicht andererseits die Anpassung an die Erfordernisse der beteiligten Disziplinen. Weiterhin erleichtert der Einsatz von Ontologien zur Datenspeicherung einen oben angesprochenen Fokuswechsel, da Daten nicht neu erfasst

werden müssen, sondern lediglich die bereits bestehenden Randinformationen angereichert werden und somit der Detailgrad auf neue Anforderungen angehoben werden kann. In EDEN wurden beispielsweise die rudimentären, zunächst aufgrund von Nennungen im Text erfassten Personendaten zu Datensätzen ersten Ranges mit eigenem wissenschaftlichen Kommentar ausgebaut. Ebenso könnten die geographischen Informationen durch eine Kooperation mit Geographen ausgeweitet werden und EDEN zu einer für (kultur-)geographische Forschungen nutzbaren Quelle machen. Diese Änderungen des Datenmodells könnten mit WissKI problemlos parallel zum Einpflegen der Daten erfolgen. Somit kann die Datenbank nicht nur hinsichtlich der Datenmenge, sondern auch hinsichtlich der Fülle der Metadaten bei Bedarf stetig wachsen.

Sammlung Rück

Das im Archivbereich des GNM bisher benutzte Dokumentationssystem "Faust" (<http://www.land-software.de/>) ist für die Erfassung, Erschließung und Abbildung der Projektdaten in vielerlei Hinsicht unzureichend. Zum einen würde eine Anpassung des Archivsystems auf die Bedürfnisse des Projekts Rück dazu führen, dass in allen anderen Anwendungsbereichen entsprechende Felder auftauchen. Eine inhaltliche Tiefenerschließung in der Qualität des Projekts Rück ist für die sonstigen Archivmaterialien eher untypisch, da für die Erfassung letzterer insbesondere die Quantität im Vordergrund steht.

Faust hat bislang keine Komponente zur Text-Annotierung in Fließtexten, was für die semantische Tiefenerschließung der Archivmaterialien essenziell ist. Auch ist Faust nur bedingt in der Lage, die Verknüpfungsdichte an Daten darzustellen. Unter diesem Aspekt kann eines der Kernziele, das Netzwerk agierender Personen im zeitlichen und räumlichen Kontext abzubilden, kaum realisiert werden.

WissKI erfüllt die Bedürfnisse des Projekts, da es über einen Text-Annotierer für Fließtext verfügt und auf der Basis von Semantic Web-Techniken Verknüpfungen flexibel darstellt und auswertet. Auch konnte mit WissKI eine direkte Schnittstelle am Objektdatensatz zu MIMO (2009-*) (musical instrument museums) eingerichtet werden.

In WissKI können die Forschungsprimärdaten inklusive der Digitalisate in voller TIFF-Qualität und allen Annotierungen der Fachöffentlichkeit zur Verfügung gestellt und eine Präsentationsoberfläche für die breite Öffentlichkeit geschaffen werden.

Fazit und Ausblick

Obleich in ihren Disziplinen und Objektgattungen verschieden, haben beide Anwendungsfälle gemeinsame Herausforderungen: Zum einen sind sie mit heterogenen,

stark vernetzten Daten und anwendungsspezifischen Anforderungen konfrontiert. Zum anderen entwickeln sich die Anforderungen an die Software hinsichtlich Umfang und Vernetzung der Metadaten im Projektverlauf. WissKI ist jedoch aufgrund seiner Systemarchitektur darauf bestens vorbereitet.

Im derzeit noch laufenden Projekt WissKI² wird die Entwicklung der WissKI-Software konsequent weitergeführt. Neben den bereits bewährten Formularfeldern und Bildern sollen nun verstärkt auch interaktive Landkarten, Zeitstrahlen, 3D-Animationen und weitere Medientypen eingebunden und dargestellt werden. Damit ergeben sich weitere Möglichkeiten der Datenpräsentation und der Wissensvernetzung für die vorgestellten Anwendungsfälle.

Bibliographie

Dreyer, Boris / Holdenried, Marvin / Scholz, Martin (2014-*): *EDEN — Epigraphische Datenbank Erlangen-Nürnberg*. WissKi: Friedrich-Alexander-University, Erlangen-Nuremberg (FAU) <http://wisski.cs.fau.de/eden/> [letzter Zugriff 08. Januar 2016].

Görz, Günther (2011): "WissKI: Semantische Annotation, Wissensverarbeitung und Wissenschaftskommunikation in einer virtuellen Forschungsumgebung", in: *Kunstgeschichte. Open Peer Reviewed Journal* <http://www.kunstgeschichte-journal.net/167/> [letzter Zugriff 08. Januar 2016].

Görz, Günther / Scholz, Martin (2012): "WissKI: A Virtual Research Environment for Cultural Heritage", in: De Raedt, Luc, Luc De Raedt, Bessiere, Christian / Dubois, Didier / Doherty, Patrick / Frascioni, Paolo / Heintz, Fredrik / Lucas, Peter (eds.): *20th European Conference on Artificial Intelligence, ECAI 2012, Proceedings*. IOS Press <http://www2.lirmm.fr/ecai2012/> [letzter Zugriff 08. Januar 2016].

Görz, Günther / Scholz, Martin / Merz, Dorian / Krause, Siegfried / Fichtner, Mark / Reinfandt, Kerstin / Grobe, Peter / Pfeifer, Maria Anna (2009-*): *WissKi: Scientific Communication Infrastructure*. Friedrich-Alexander-University, Erlangen-Nuremberg (FAU); Digital Humanities Research Group, Department of Computer Science, Germanisches Nationalmuseum (GNM) Nuremberg; Biodiversity Informatics Group, Department of Museum Informatics, Zoologisches Forschungsmuseum Alexander Koenig (ZFMK) Bonn <http://wiss-ki.eu/> [letzter Zugriff: 08. Januar 2016].

Hohmann, Georg / Fichtner, Mark (2015): "Chancen und Herausforderungen in der praktischen Anwendung von Ontologien für das Kulturerbe", in: Robertson-von Trotha, Caroline Y. / Schneider, Ralf M. (eds.): *Digitales Kulturerbe. Bewahrung und Zugänglichkeit in der wissenschaftlichen Praxis (= Kulturelle Überlieferung – digital 2)*. Karlsruhe: Karlsruhe Scientific Publishing 115-128.

MIMO (2009-*): *Musical Instrument Museums Online*. Philharmonie de Paris: Cité de la musique, The University of Edinburgh, Germanisches Nationalmuseum, Muziekinstrumentenmuseum, Association "Amici del Museo degli Strumenti Musicali" <http://www.mimo-international.com/MIMO/accueil-ermes.aspx> [letzter Zugriff 08. Januar 2016].

Scholz, Martin / Holdenried, Marvin / Dreyer, Boris / Meyer-Wegener, Klaus / Görz, Günther (2014): "Und Semantik wuchs in EDEN - Eine Vorstellung und ein Erfahrungsbericht", in: *Magazin für digitale Editionswissenschaften* 1, 1: 22-30 https://www.mde.fau.de/files/2015/03/MdE-2015-01_3_Scholz_et_al.pdf [letzter Zugriff 08. Januar 2016].