

Datenbank

Vernetztes Arbeiten im Corpus-Projekt

Für das Vorhaben entsteht derzeit eine auf die Anforderungen der Kunstgeschichte zugeschnittene digitale Forschungsumgebung.

VON WERNER KÖHLER

Abb. 1: Linking Open Data Cloud Diagram, von Max Schmachtenberg, Christian Bizer, Anja Jentzsch und Richard Cyganiak, 2014 (<http://lod-cloud.net>). Das Bild zeigt die Verknüpfungen zwischen Datenbeständen, die als Linked Data in offenen Formaten publiziert sind.

Das ZIEL DES Akademie-Projekts *Corpus der barocken Deckenmalerei in Deutschland* ist die umfassende Darstellung der zwischen 1550 und 1800 entstandenen Werke der Wand- und Deckenmalerei auf dem Gebiet der Bundesrepublik Deutschland in einer Webanwendung. Das Projekt beinhaltet die Erforschung und Dokumentation des Denkmalbestands von etwa 5.000 bekannten Objekten durch eine Gruppe von Forscherinnen und Forschern der Ludwig-Maximilians-Universität München, wobei auch zerstörte, jedoch durch historisches Quellenmaterial rekonstruierbare Werkkomplexe miteinbezogen werden. Das Projekt hat eine Laufzeit von 25 Jahren, in der die Malereien

umfassend erforscht, dokumentiert und in ihren historischen und kunsthistorischen Kontext eingeordnet und interpretiert werden.

Das Deutsche Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg (DDK) hat innerhalb des Projektes im Wesentlichen zwei Aufgaben. Zum einen führt es die Fotokampagnen zur digitalen Reproduktion der Wand- und Deckengemälde durch, zum anderen ist es für die IT-Betreuung und die Entwicklung der digitalen Forschungsumgebung als webbasierter Arbeits- und Publikationsplattform zuständig. Letztere Aufgabe wird durch die IT-Abteilung des DDK geleistet.

Anforderungen an die digitale Forschungsumgebung

Für die zu entwickelnde digitale Forschungsumgebung des Projekts *Corpus der barocken Deckenmalerei in Deutschland* gibt es eine Reihe von Anforderungen:

Orientierung

Die Entwicklung einer tragfähigen Software-Lösung für ein so langfristig ausgerichtetes Unternehmen verlangt aus Perspektive der Softwaretechnik eine sorgfältige Planung. Da es eine konfektionierte Standardsoftware für die hochspezialisierten Aufgaben im Bereich kunsthistorischer Forschungsprojekte nicht gibt, müssen hier Lösungen erst konzipiert und gefunden werden. Dabei wird für die Planung zunächst ein Zeitraum von etwa zwei Jahren in den Blick genommen, denn auf die ferne Zukunft hin kann niemand heute vorhersehen, wie das Softwareprodukt am Ende des Projekts in 25 Jahren aussehen wird.

Aktuell gilt es daher, bei der Entwicklung der digitalen Forschungsumgebung als Arbeits- und Publikationsplattform innerhalb des Projekts eine zukunfts offene Lösung anzustreben und die aktuell wichtigen Entwicklungen des *World Wide Web* (WWW) mitzubetrachten und mitzugestalten. Zentral hierfür sind die Konzepte des *Semantic Web* und *Linked Data*: www.w3.org/standards/semanticweb sowie www.w3.org/standards/semanticweb/data.

Semantic Web und Linked Data

Sehr knapp gesagt meint *Semantic Web*, die Daten innerhalb des WWW semantisch „bedeutungsvoller“ in dem Sinne zu gestalten bzw. auszuzeichnen, dass sie maschinell einfacher und angemessener interpretiert werden können. Beispielsweise können Homographen (dieselbe Schreibweise bei unterschiedlicher Bedeutung, z. B. „Bank“ oder „Kohl“) ohne explizite Markierung der semantischen Differenz maschinell nicht leicht unterschieden werden.

Linked Data meint die Verlinkung und Vernetzung von Daten im WWW (Abb. 1). So sollte bei der Entwicklung einer Webanwendung für die kunsthistorische Forschung von vornherein auf die Einbettung in internationale Forschungskontexte und auf die informationstechnischen Interaktionen mit anderen Webanwendungen fokussiert werden. Forschung findet heutzutage nicht isoliert statt, sondern im WWW, und Forschungspublikationen dürfen keine Insel-Lösungen sein, sondern müssen von Anfang an in offenen Formaten erfolgen, die vernetzt genutzt werden können.

Technische Formate

Dies betrifft sowohl technische Formate für Medien- und Textdateien als auch standardisierte Schnittstellen und Sprachen für die Referenzierung und semantische Auszeichnung von Inhalten. Hinsichtlich der Unterstützung technischer Formate gilt es z. B., für die Bereit-

ABB: HTTP://I00D-CLOUDNET; HTTP://WISS-KI.EU/SITES/WISS-KI.EU/FILES/IMAGES/WISSKI-DIAGRAMM.PNG

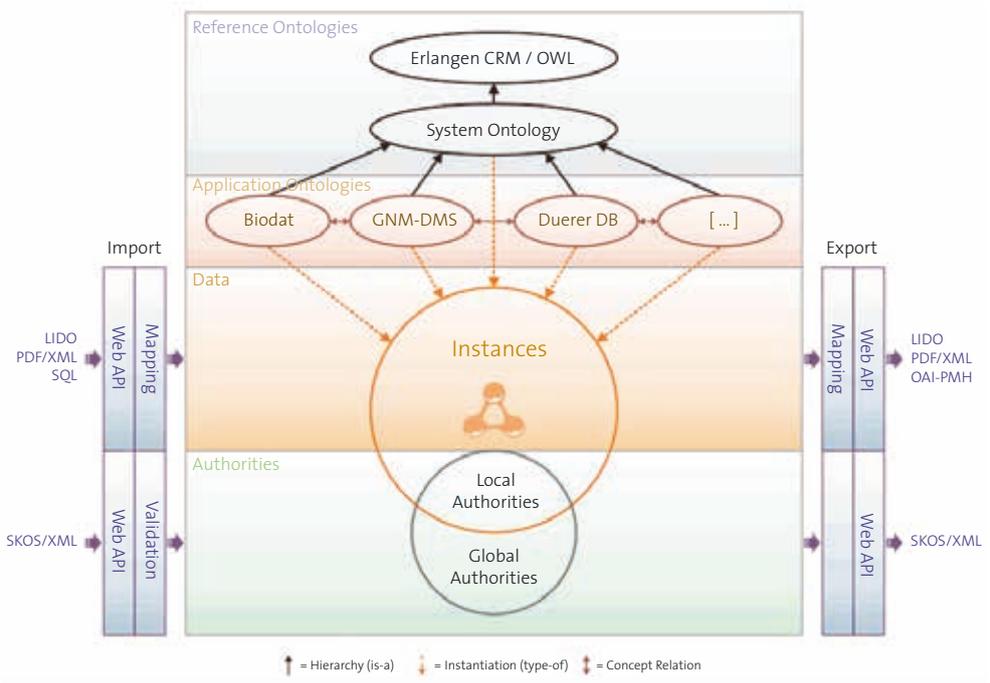


Abb. 2: WissKI Ontology Layercake. Das Bild zeigt die Schichtenarchitektur der digitalen Forschungsumgebung WissKI, die im Projekt *Corpus der barocken Deckenmalerei* verwendet wird.

stellung von 3D-Modellen und VR-Umgebungen deren Nachnutzbarkeit und Nachhaltigkeit für die Langzeitarchivierung sowie die Darstellung in unterschiedlichen Viewern zu beachten, wobei auch rechtliche Fragen eine nicht unbedeutende Rolle spielen können.

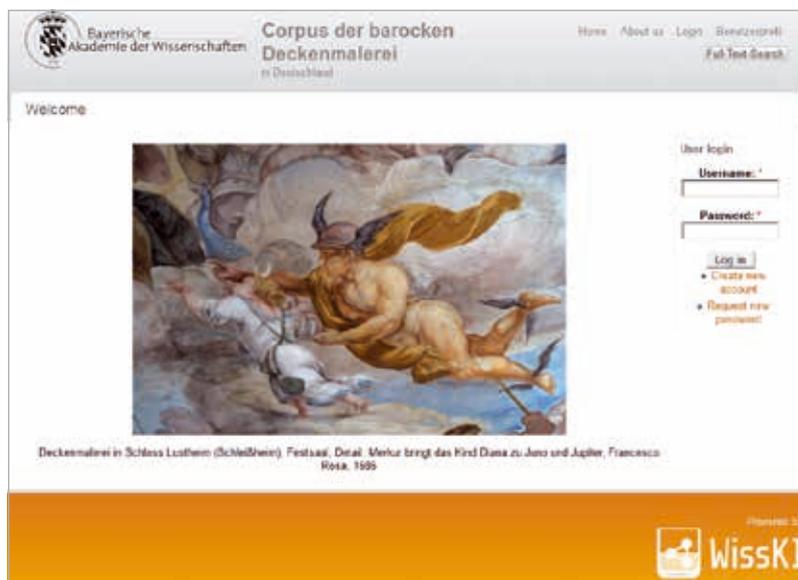
Normdateien, Thesauri, Vokabulare

Hinsichtlich standardisierter Schnittstellen bilden die international unterschiedliche Benennung von Fachbegriffen oder differierende Schreibweisen von Namen für Personen oder Geographika ein inhaltlich bedeutsames Problemfeld, z. B. im Kontext von Suchanfragen. Im internationalen Forschungszusammenhang sollten Suchanfragen, egal in welcher Sprache sie abgefasst sind, relevante Treffer aus verschiedenen Webanwendungen unabhängig davon zurückerhalten, in welcher Sprache die Informationen ursprünglich vorliegen. Eine Lösung hierfür stellen internationale Normdateien (*Authority Files*) dar, die nichtsemantische Identifikatoren (*IDs*) für Entitäten – wie z. B. Personen, Sachbegriffe, Geographika – bereitstellen und darüber hinaus häufig auch Konkordanzen zu unterschiedlichen Schreibweisen sammeln. Beispiele hierfür sind die *Gemeinsame Normdatei (GND)* der Deutschen Nationalbibliothek sowie das *Virtual International Authority File (VIAF)*. Im kunstgeschichtlichen Kontext sind hier insbesondere auch die vom Getty Research Institute herausgegebenen Vokabulare *Art and Architecture Thesaurus (AAT)*, *Thesaurus of Geographic Names (TGN)*, *Union List of Artist Names (ULAN)* und *Cultural Objects Name Authority (CONA)* von Bedeutung (www.getty.edu/research/tools/vocabularies). Werden in digitalen Forschungsumgebungen die jeweils anwendungsspezifischen Daten auf die Normdateien abgebildet, können die Anwendungen über die Identifikatoren aus den Normdateien vernetzt werden und miteinander kommunizieren. Die Implementierung von Schnittstellen zu solchen Normdateien ist daher eine wichtige Anforderung an eine digitale Forschungsumgebung.

Ontologien

Darüber hinaus ist die Strukturierung von kunsthistorischen Inhalten heutzutage von vornherein anwendungsübergreifend zu konzipieren, damit auch komplexer strukturierte Informationen zwischen Webanwendungen

ausgetauscht und verlinkt werden können. So sollten z. B. Suchanfragen nach Porträts eines Künstlers dahingehend differenziert werden können, ob der Künstler in seiner Rolle als Hersteller eines Porträts oder aber in seiner Rolle als Porträtiertes gefunden werden soll. Bei einer Suche nach Künstlerelbstbildnissen würden beide Rollen zutreffen. Solche Unterscheidungen, die von Menschen kontextabhängig leicht getroffen werden können, müssen für Computer explizit gemacht werden und in den Datenstrukturen abgebildet sein, damit sie zwischen Webanwendungen ausgetauscht werden können. Hierfür kann im Hinblick auf Objekte des kulturellen Erbes und ihrer Bedeutungszusammenhänge das *CIDOC Conceptual Reference Model (CIDOC CRM)* verwendet werden, das seit 2006 als ISO-Standard 21127 anerkannt ist. Der Standard wurde von der *CIDOC Documentation Standards Group* entwickelt. *CIDOC (Comité international pour la documentation)* steht für das Internationale Komitee zur Dokumentation des internationalen Museumsverbandes *ICOM (International Council of Museums)*. Vereinfacht gesagt, zielt das *CIDOC CRM* darauf ab, das gesamte Wissensgebiet, Objekte des kulturellen Erbes betreffend, abzubilden und in Form einer sogenannten Ontologie bereitzustellen. Webanwendungen im Bereich des kulturellen Erbes, die ihre anwendungsspezifischen Daten auf diese Ontologie abbilden, können dadurch leichter Informationen über ihre Anwendungsobjekte austauschen. Die Unterstützung von Normdateien, Thesauri und kontrollierten Vokabularen sowie die Möglichkeit der Abbildung des anwendungsspezifischen Datenmodells auf das *CIDOC CRM* sind wichtige Anforderungen an die zu entwickelnde digitale Forschungsumgebung



DER AUTOR

Werner Köhler M. A. ist Kunsthistoriker und Kulturinformatiker und als Mitarbeiter der IT-Abteilung in der Arbeitsstelle Marburg für Software, Datenmodellierung und Datenmanagement des Projekts *Corpus der barocken Deckenmalerei* zuständig.

als Arbeits- und Publikationsplattform innerhalb des Corpus-Projekts.

Workflow

Darüber hinaus sollte die digitale Forschungsumgebung als Arbeitsplattform die Forscher in allen Stadien des Forschungsprozesses funktional unterstützen, und zwar möglichst umfassend in sämtlichen Aspekten und Phasen, angefangen bei der Materialsammlung von Notizen, Skizzen, Bestandslisten, Arbeitsfotos etc. Weiter geht es über das Ordnen und Zusammenführen sowohl textlicher als auch visueller Informationen hin zum Entwerfen und Erstellen von strukturierten Texten und deren Verknüpfung mit zusätzlichen, visuellen und textlichen Informationsquellen, bis zur Verwaltung unterschiedlicher Versionen von Texten in unterschiedlichen Stadien. Letztlich folgt die abschließende Veröffentlichung einer finalen Version als Publikation im Webportal. Die digitale Forschungsumgebung sollte kollaboratives Arbeiten solchermaßen unterstützen und die Kommunikation der Forscher sowie das Teilen und Diskutieren der Ergebnisse untereinander online ermöglichen.

WissKI als digitale Forschungsumgebung für die Kunstgeschichte

Als „Wissenschaftliche Kommunikationsinfrastruktur“ erfüllt die Software WissKI die oben genannten spezifischen Softwareanforderungen weitgehend (<http://wiss-ki.eu>). Sie wird daher innerhalb des Corpus-Projekts als digitale Forschungsumgebung produktiv eingesetzt und über einen Zeitraum von zwei Jahren erprobt. WissKI wird gegenwärtig mit Förderung der Deutschen Forschungsgemeinschaft als Joint Venture von der Digital Humanities Research Group der Informatik an der Friedrich-Alexander-Universität Erlangen-Nürnberg, der Museum-Informatik-Abteilung im Germanischen Nationalmuseum in Nürnberg und der Biodiversity Informatics Group am Zoologischen Forschungsmuseum Alexander Koenig in Bonn in einer neuen Version weiterentwickelt. Die Vorgängerversion ist bereits in verschiedenen Museumsprojekten und Forschungsprojekten im Einsatz.

WissKI versucht, komplementär zwei Konzepte zu vereinen. Es baut einerseits auf der Idee des Wiki auf, das ortsunabhängiges, kollaboratives Arbeiten via Webbrowser ermöglicht. Zum anderen basiert es auf Grundkonzepten aus der Forschung zur künstlichen Intelligenz (KI) und wird dahingehend weiterentwickelt, dass auf Grundlage der in WissKI erfassten wissenschaft-

lichen Daten ein maschinelles Reasoning über die Daten erfolgen kann. Auf diese Weise können eigenständige, neue Erkenntnisse und Schlussfolgerungen aus den Daten errechnet werden.

Funktionen und Anpassungen

Die Datenhaltung in WissKI verwendet aktuelle Semantic-Web-Technologien und erfolgt ontologiebasiert in einem Triple Store, sodass die Forschungsergebnisse mittels Linked Data weltweit mit anderen Datenrepositorien verknüpft werden können. Das implementierte Datenmodell ist zum *CIDOC CRM* kompatibel, und die Anbindung kontrollierter Vokabulare und Normdaten beim Erfassen mit dem XML-Editor ist möglich (Abb. 2). WissKI bringt also bereits viele der benötigten Grundfunktionalitäten mit und bildet daher für kunsthistorische Anwendungen ein sehr gut nutzbares Framework, muss allerdings im Hinblick auf die sich jeweils aus der spezifischen kunsthistorischen Thematik ergebenden Fragestellungen angepasst werden.

WissKI ist eine Open Source-Software und besteht im Wesentlichen aus PHP-Modulen, die auf dem PHP-Framework Drupal aufsetzen, einem weit verbreiteten Content Management System. So lassen sich in technischer Hinsicht Anpassungen – unter Voraussetzung der Verfügbarkeit entsprechender Ressourcen – durchaus projektintern durch Web-Programmierung durchführen.

Weiterentwicklung der Infrastruktur

Softwareumgebungen befinden sich in ständiger Weiterentwicklung. Daher ist es notwendig, auch WissKI als grundlegendes Framework kontinuierlich anzupassen. Die hierfür zu entwickelnde Infrastruktur befindet sich noch im Aufbau und muss in naher Zukunft eine institutionalisierte Form gewinnen. Eine digitale Forschungsumgebung für die Kunstgeschichte zu entwickeln ist eine Notwendigkeit und steht außer Frage. Das Projekt *Corpus der barocken Deckenmalerei in Deutschland* unterstützt diese Entwicklung und hilft beim Aufbau des dazu benötigten Netzwerkes (Abb. 3). ■

Abb. 3: WissKI, die digitale Forschungsdatenbank des Corpus der barocken Deckenmalerei in Deutschland.

Literatur

G. Görz, WissKI. Semantische Annotation, Wissensverarbeitung und Wissenskommunikation in einer virtuellen Forschungsumgebung, in: *Kunstgeschichte. Open Peer Reviewed Journal*, 2011 (urn:nbn:de:bvb:355-kuge-167-7).