

Erschließung digitaler Ressourcen und Forschungsdaten in den Digital Humanities: Der Digitale Wissensspeicher an der Berlin-Brandenburgischen Akademie der Wissenschaften

,
grabsch@bbaw.de
Berlin-Brandenburgische Akademie der Wissenschaften,
Deutschland

,
juergens@bbaw.de
Berlin-Brandenburgische Akademie der Wissenschaften,
Deutschland

Im Rahmen des DFG-geförderten Projektes „Digitaler Wissensspeicher“ wurde an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) von 2012 bis April 2015 ein zentraler Zugang für sämtliche digitalen Forschungsdaten und Ressourcen der Akademie geschaffen. Damit sind über 170 Projekte mit insgesamt mehr als 1 Mio. digitaler Ressourcen im Volltext und mit Metadaten erfasst. Diese in Inhalt, Format und Sprache äußerst vielfältigen und heterogenen Ressourcen werden erstmals vollständig über ein einheitliches Interface zugänglich gemacht. Den Kern der Ressourcen stellen digitale Editionen und Übersetzungen (in XML, HTML, PDF), elektronische Kataloge, Dokumentationen und Datenbanken, digitale Volltextsammlungen sowie Wörterbücher dar. Durch den Einsatz von Sprachtechnologien (Bing, DONATUS) kann der Digitale Wissensspeicher mehrsprachig (deutsch/englisch/französisch) und morphologisch normalisiert durchsucht werden. Neben manuell erfassten, fließen auch automatisiert erstellte Metadaten in den Datenbestand des Wissensspeichers ein und können von den Nutzerinnen und Nutzern abgefragt werden. Die Metadaten für den gesamten Bestand der digitalen Ressourcen werden dabei auch über eine maschinenlesbare Schnittstelle (OAI-PMH) zur Verfügung gestellt. Damit werden die digitalen Forschungsdaten der BBAW Teil der Linked Open Data Cloud. Neben der Einbeziehung von Semantic-Web-

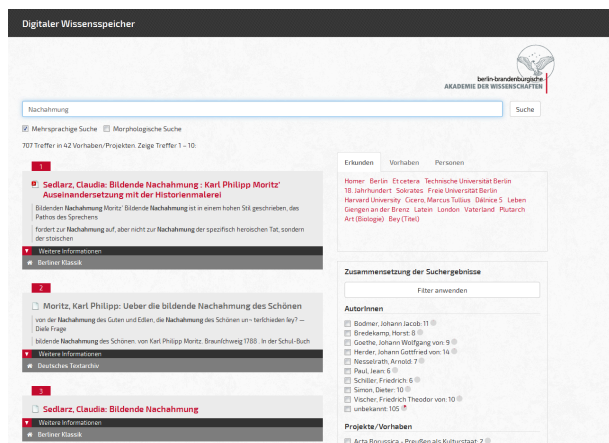
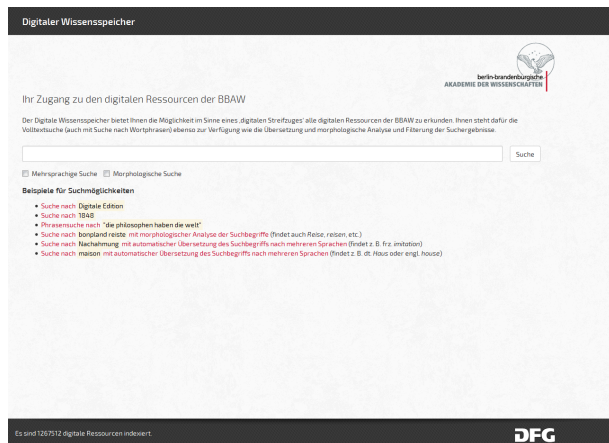
Ressourcen zur Anreicherung der Suchergebnisse stellt der Wissensspeicher so auch Ressourcen und Forschungsdaten der BBAW zur Verfügung.

Die größte Herausforderung beim Aufbau einer disziplinübergreifenden Forschungsdateninfrastruktur wie dem Wissensspeicher stellte die große Heterogenität der an der BBAW in den letzten 20 Jahren entstandenen digitalen Ressourcen dar. Sowohl inhaltlich als auch technisch bilden diese ein unübersichtliches Feld, das für den Digitalen Wissensspeicher strukturiert und zugänglich gemacht werden musste. Gestützt auf einen Volltextindex (Apache Lucene) und ein an die Bedingungen der Akademie angepasstes Metadatenschema (basierend auf OAI-ORE) wurden für die einzelnen Projekte und Ressourcensammlungen Importmodule entwickelt, welche die sehr unterschiedlichen Datenstrukturen der jeweiligen Projekte in den (Meta-)Datenbestand des Wissensspeichers integrieren. Über die Einbindung von Semantic-Web-Technologien (u. a. von DBpedia) und Text-Mining-Tools werden für die Nutzerinnen und Nutzer semantisch an die Suchanfrage gebunden Vorschläge für die Erweiterung und Erkundung des Datenbestandes angeboten.

Die zweite Projektphase des Wissensspeichers wird bis Ende 2017 die Erweiterung der Nutzungsmöglichkeiten und besonders die Nachhaltigkeit des Projektes zum Schwerpunkt haben. Es besteht ein erheblicher Bedarf von wissenschaftlichen Institutionen an einer nachhaltigen Forschungsdateninfrastruktur, welche die spezifische Situation der Geisteswissenschaften – heterogene Ressourcen – adäquat abbilden kann. Ein wichtiges Ziel der nächsten Projektphase ist es daher, den Wissensspeicher für weitere Nutzergruppen zu öffnen. Dafür werden die Softwarekomponenten des Wissensspeichers als installierbares Paket zur Verfügung gestellt. So werden externe Partnerinstitutionen in die Lage versetzt, einen eigenen Wissensspeicher mit eigenen digitalen Ressourcen zu betreiben. Für die Abstimmung mit zukünftigen externen Nutzern und die koordinierte Weiterentwicklung als Open Source Software wird an der BBAW im April 2016 ein offener Workshop stattfinden.

Ein weiterer Hauptpunkt der Arbeit in dieser Projektphase bildet die Erstellung von Guidelines mit strukturellen und inhaltlichen Mindestanforderungen, die Ressourcen und Metadaten erfüllen müssen, um effektiv und zweckmäßig in den Wissensspeicher aufgenommen werden zu können. Damit werden Zielvorgaben für die (technische) Qualität der in den Wissensspeicher aufzunehmenden Ressourcen mit ihren Metadaten formuliert. Diese Best-Practice-Empfehlungen können auch über den konkreten Anwendungsfall des Digitalen Wissensspeichers hinaus einen Empfehlungscharakter für den Aufbau und Betrieb von Ressourcensammlungen in den Digital Humanities bekommen und sollen von den Partnerinstitutionen sukzessive optimiert und an die eigenen Gegebenheiten angepasst werden. Ebenso werden Workflows für die manuelle Erhebung und Einspeisung von Metadaten entwickelt. Weitere Entwicklungsziele stellen der Ausbau von Visualisierungskomponenten sowie die automatisierte

Auswertung und Einbindung von Nutzerfeedback in den Suchprozess dar.



Fußnoten

1. Der derzeitige Entwicklungsstand des Digitalen Wissensspeichers kann unter <http://wspdev.bbaw.de> eingesehen und getestet werden. Je nach Gegebenheiten vor Ort soll dies über einen Laptop auch bei der Posterpräsentation ermöglicht werden.