

Kollaborative Forschung über Linked Open Data Forschungsdatenbanken der Universitätsgeschichte Implementierung des Heloise Common Research Model

Riechert, Thomas

thomas.rieichert@htwk-leipzig.de
Hochschule für Technik, Wirtschaft und Kultur (HTWK)
Leipzig, Deutschland

Beretta, Francesco

francesco.beretta@ish-lyon.cnrs.fr
Laboratoire de recherche historique Rhône-Alpes
(LARHRA), Frankreich

Motivation

Die Beantwortung von Forschungsfragen über bestehende Forschungsdatenquellen im Linked Open Data Web ist von besonders hoher Relevanz für nachhaltige Forschungsarbeiten in den Geisteswissenschaften und insbesondere in der Geschichtswissenschaft (vgl. Meroño Peñuela et al. 2014: 1-27). Eine ständig wachsende Menge an Archivmaterialien, Literaturquellen, Forschungsergebnisse und Forschungsdatenbanken ist online verfügbar. Durch Standardisierungsbestrebungen u. a. durch den Einsatz von RDF und OWL als Beschreibungssprache von Daten ist es möglich diese zu verknüpfen, und Inferenz-Algorithmen auf diesen Ressourcen anzuwenden. Eine typische Herangehensweise an die damit einhergehenden Herausforderungen der Datenintegration ist die Verwendung von Standard-Vokabularen, wie sie z. B. im E-Business erfolgreich praktiziert wird (vgl. Domingue et al. 2011). Im Bereich der Geisteswissenschaften haben sich zum Beispiel mit GND und Europeana Data Model (EDM) solche Vokabulare entwickelt. Darüber hinaus wird CIDOC-CRM von unterschiedlichen Projekten als Modell für die Veröffentlichung von Daten verwendet (vgl. Kurtz et al. 2009). Die Autoren sind in verschiedenen Forschungsprojekten in Deutschland und Frankreich im Bereich der Informatik und der Geschichtswissenschaften aktiv und müssen konstatieren, dass der Weg der Standardisierung in der historischen Forschung schwieriger ist. Dies ist vor allem auf den hohen Grad

der domänenspezifischen Eigenheiten der vorliegenden Daten, und auf die besondere Rolle projektbezogener Forschungsfragestellungen bei der Datenerstellung und Datenerhebung zurückzuführen (vgl. Beretta 2009).

Das im Jahr 2012 gegründete Europäische Netzwerk Héloïse zur Vernetzung von online verfügbaren Datenbanken im Bereich der Universitätsgeschichte, sieht sich dieser Herausforderung gegenübergestellt. Inhalte regelmäßig stattfindender Workshops sind die Präsentation von verfügbaren Forschungsdatenbanken und deren Verwendung bei auf diese heterogenen Datenbanken aufbauenden Forschungsfragestellungen. Parallel mit der Nutzung der Repositorien in zukünftigen Forschungskontexten erfolgt eine kollaborative Daten-Kuration und sichert daher die Langzeitlebigkeit von historischen Forschungsdaten - nicht nur im Sinne von Datensicherung, sondern auch von ständiger Anreicherung und Qualitätsverbesserung der Daten.

Heloise Common Research Model

Mit dem Heloise Common Research Model (HCRM) haben die beiden Autoren dem Konsortium eine Methode präsentiert (vgl. Beretta und Riechert 2016), welches durch die Forschungspartner gemeinsam verfeinert wird und in der Forschungsdomäne eingesetzt werden soll. Das HCRM ist als Schichtenarchitektur konzipiert, bestehend aus drei Schichten: dem Repository-Layer, dem Application-Layer und dem Research-Interface-Layer. In einem parallelen Entwicklungsprozess entsteht neben der detaillierten Definition von Modulen (vgl. Abbildung 1, Beretta und Riechert 2016) im HCRM eine Implementierung der Methodologie in Form der Héloïse-Plattform.

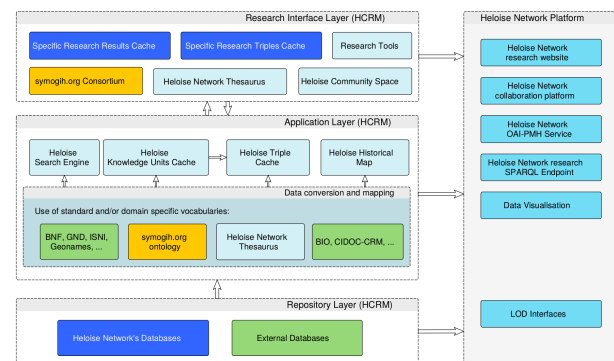


Abbildung 1: Heloise Common Research Modell – Überblick über die Schichten und Module (links), Héloïse-Plattform (rechts) (vgl. Beretta und Riechert 2016)

Repository-Layer Die Datenbanken der Partnern des Héloïse-Netzwerkes, genauso wie externe Informationsquellen, stellen die Basis zur Beantwortung übergeordneter Forschungsfragen. Die Publikation der

Daten als LOD erfolgt mit Hilfe etablierter Werkzeuge. So werden u. a. die Werkzeuge OntoWiki (Frischmuth et al. 2013) und D2RQ (Bizer und Cyganiak 2006) zum Publizieren eingesetzt. Die Verlinkung erfolgt innerhalb etablierter Authoring-Prozesse in den Forschungsprojekten.

Application-Layer: Die Applikations schicht unterstützt das Auffinden von Ressourcen und deren Verlinkung. Hierfür wird auf generische Werkzeuge und standardisierte Vokabulare aufgebaut, damit einhergehend ist der Zugang auf die Informationen auf diese Standards beschränkt. Als erste Anwendungen im Forschungskonsortium wird gegenwärtig die Implementierung einer Personen-Suche, basierend auf der BIO-Ontologie (Davis und Galbraith 2010), sowie die Darstellung relevanter geographischer Daten, über die Repositorien der Partner, entwickelt.

Research Interface Layer Die Forschungsschicht des Modells bietet einen Zugang auf die Forschungsdaten für neue wissenschaftliche Fragestellungen. Hierfür wird basierend auf einer Meta-Ontologie, wie sie das SyMoGIH Projekt (Beretta 2015, Beretta 2016) für die Geschichtswissenschaft entwickelt hat, ein fachspezifisches Meta-Vokabular entwickelt. Dieses Vokabular verbindet einen kollaborativen, domänenspezifischen Ansatz mit einer größtmöglichen Unabhängigkeit von spezifischen, Projekt-bezogenen Forschungsfragestellungen und ermöglicht den Datenaustausch im Héloïse-Netzwerk.

Héloïse-Plattform

Zentral für zukünftige Forschungsprojekte im Héloïse-Konsortium ist die Implementierung der Methodologie. Die heterogene Zusammensetzung des Konsortium spricht gegen die Etablierung einer zentralen Administration einer solchen Plattform. Vielmehr ist die Forschungsk Kooperation auf der Ebene der Historiker in einer vergleichbaren Art und Weise bei den assoziierten IT-Partnern zu finden. Die Autoren schlagen daher die Realisierung von Diensten der Plattform durch Microservices vor. Microservice stellen unabhängige Dienste zur Verfügung und sind im Sinne der angestrebten Cloud-basierten Plattform virtualisierbar (vgl. Newman 2015).

Das Poster stellt die Resultate der iterativen Implementierung des HCRM im Héloïse-Forschungsnetzwerk detailliert vor. Es werden die Ergebnisse der fachlichen Diskussion über zwei Heloise Workshops (Madrid, 2015 und Perugia, 2016) vorgestellt. Der komplette Katalog, der durchgängig im Kontext der Linked-Open-Data-Philosophie, als Open-Source verfügbaren Microservices (Docker Container), wird präsentiert und deren Anwendung innerhalb der Heloise-Plattform online gezeigt.

Fußnoten

1. GND Ontology <http://d-nb.info/standards/elementset/gnd> [25/08/2016]
2. Europeana Data Model: <http://pro.europeana.eu/page/edm-documentation> [25/08/2016]
3. CIDOC Conceptual Reference Model: <http://www.cidoc-crm.org> [25/08/2016]
4. Héloïse - European Workshop on academic Database: <http://heloisenetwork.eu/> [25/08/2016]
5. Héloïse-Partner: <http://heloisenetwork.eu/repositories> [25/08/2016]
6. Ontologie und Instanzen des Vokabulars sind auf der Webseite des Projektes online zugänglich: <http://symogih.org> [25/08/2016]
7. Docker Virtualisierung: <http://docker.com> [25/08/2016]
8. Héloïse-Network-Plattform: <http://heloisenetwork.eu/platform> [25/08/2016]

Bibliographie

Beretta, Francesco (2015): „Publishing and sharing historical data on the semantic web: the SyMoGIH project–symogih.org“, in: *Workshop: Semantic Web Applications in the Humanities II*.

Beretta, Francesco (2016): „L’interope#rabilite# des donne#es historiques et la question du mode#le: l’ontologie du projet SyMoGIH“, in: Minel, Jean-Luc (ed.): *Quels enjeux nume#riques pour les me#diations scientifique et culturelle*. Presses universitaires de Paris Ouest (im Erscheinen).

Beretta, Francesco / Riechert, Thomas (2016): „Collaborative Research on Academic History using Linked Open Data: A Proposal for the Heloise Common Research Model“, in: *CIAN-Revista de Historia de las Universidades, Norteamérica* 19 (Juni).

Bizer, Christian / Cyganiak, Richard (2006): „D2r server-publishing relational databases on the semantic web“, in: *5th International Semantic Web Conference* 294–309.

Davis, Ian / Galbraith, David (2010): *BIO: A vocabulary for biographical information*.

Domingue, John / Fensel, Dieter / Hendler, James A. (eds.) (2011): *Handbook of semantic web technologies*. Berlin / Heidelberg: Springer Science / Business Media.

Frischmuth, Philipp / Martin, Michael / Tramp, Sebastian / Riechert, Thomas / Auer, Sören (2014): „OntoWiki - An Authoring, Publication and Visualization Interface for the Data Web“, in: *Semantic Web Journal (IOS Press)*.

Kurtz, Donna / Parker, Greg / Shotton, David / Klyne, Graham / Schroff, Florian / Zisserman, Andrew / Wilks, Yorick (2009): „Claros-bringing classical art to a global public“, in: *Fifth IEEE International Conference on e-Science* 20–27.

Meroño-Peñuela, Albert / Ashkpour, Ashkan / van Erp, Marieke / Mandemakers, Kees / Breure, Leen (2014): „Semantic technologies for historical research: A survey“, in: *Semantic Web Journal (IOS Press)*.

Newman, Sam (2015): *Building Microservices: Designing Fine-Grained Systems*. O'Reilly Media.