

Die Schule von Salamanca. Ansätze für vernetzte und visualisierbare Daten

,
andreas.wagner@em.uni-frankfurt.de
Akademie der Wissenschaften und der Literatur | Mainz,
Deutschland

,
caesar@rg.mpg.de
Akademie der Wissenschaften und der Literatur | Mainz,
Deutschland

Das 2013 gestartete Projekt *Die Schule von Salamanca. Eine digitale Quellensammlung und ein Wörterbuch ihrer juristisch-politischen Sprache* umfasst eine Sammlung von über 100 Texten iberischer Theologen und Juristen des 16. und 17. Jahrhunderts zu politischen und rechtlichen Themen. Diese Sammlung wird ergänzt durch ein Handbuch, in dem neben biografischen Informationen zu den in der Edition vertretenen Autoren die Entwicklung zentraler Begriffe der europäischen Rechts- und politischen Ideengeschichte in diesem Diskussionszusammenhang erschlossen wird. Beide Teile werden vollumfänglich und frei online zugänglich sein; zu Beginn des Jahres 2016 werden die ersten Daten der Quellensammlung und des Wörterbuchs zur Verfügung gestellt. Dies sind insbesondere die TEI-XML Dateien der Volltexte der ersten Werke sowie die dazugehörigen Digitalisate. Parallel besorgen wir die Bereitstellung der Daten als Linked Open Data in Verbindung mit einem SPARQL-Endpoint.

Zu Beginn des DHd-Vortrags wird das Projekt und seine Website kurz vorgestellt. Eine wichtige Facette besonders der künftigen Projektentwicklung sehen wir aber in der Einbindung unserer Daten in eine Linked Data Infrastruktur. Der Vortrag soll daher vor allem dafür genutzt werden, um einen Einblick in die Arbeitsschritte, Entscheidungen und Implementierungen zu geben, die in dieser Hinsicht erfolgt sind und noch ausstehen.

1. Wie werden die TEI-Daten in einer Linked-Data-Umwelt angeboten: welche Elemente oder Attribute werden welchen Objekten und Prädikaten welcher Ontologien zugeordnet? Wie wird diese Zuordnung in Anschlag gebracht, um die Daten als semantische Daten anzubieten? Wie ist die zeitliche Dimension vieler biographischer Angaben abzubilden? Wie ist mit alternativen Werten, z. B. konkurrierenden Angaben zu Geburtsdaten, umzugehen? Gibt es Rückwirkungen auf unser „Kerngeschäft“, i. e. das TEI Schema und die Datenerfassung?

Wir werden semantische Daten zu den Werken und zu den in der Edition vertretenen Autoren anbieten, dabei

hauptsächlich auf die foaf-, bio-, relationship-, und SPAR-Ontologien zurückgreifen und die TEI Daten über den xtriples Webservice (xTriples 2015) in RDF umwandeln.

2. Wie werden die Daten in unseren eigenen Diensten vernetzt genutzt, wie sollen sie zur externen Nachnutzung und Vernetzung angeboten werden: Welche Dienste und Ressourcen sollen – direkt oder indirekt – über das Projekt angeboten werden? Welche Möglichkeiten ergeben sich durch die Vernetzung unserer Daten mit denen anderer Anbieter, Forschungsfragen in neuer Weise zu bearbeiten? In welcher Form können oder sollen solche erweiterten Möglichkeiten auf der Projektsite öffentlich angeboten werden, wie verhält es sich mit Rechtemanagement und Qualitätssicherung in föderierten Abfragen/Daten? Welche anderen Datenanbieter sind interessante Kooperationspartner, welches sind die einschlägigen Normdatenbanken? Was sind die Erfahrungen mit den verfügbaren Schnittstellen dieser Anbieter?

Wir werden sowohl auflösbare Entitäten URIs als auch einen SPARQL Endpoint anbieten. Föderierte Abfragen zur Bearbeitung spezieller Forschungsfragen werden im Vortrag vorgestellt. Auch für uns bislang noch ungelöste Probleme, die sich z. B. aus der heterogenen Struktur und Qualität der Daten ergeben, werden angesprochen.

Über die infrastrukturelle Seite dieser Vernetzung hinaus gilt es hier die Benutzerperspektive nicht aus den Augen zu verlieren: Wie und wo sollen welche Art von Benutzern Vernetzungen und Abfragen selbständig definieren können? Weit mehr als eine Frage der technischen Möglichkeiten ist dies ein Problem der Widmung und des Einsatzes von Entwicklungsressourcen, denn sie betrifft die Entwicklung von Oberflächen, deren Nutzung weit über die Kernaufgaben der meisten Projekte hinausgeht.

3. Visualisierungsmöglichkeiten, aufsetzend auf verschiedene Services: Ein Problem bei der Gewinnung vieler Daten aus dem Semantic Web ist ihre Aufbereitung in einer Form, die den Anwendenden dazu befähigt, neue Entdeckungen zu machen, vielleicht sogar neue wissenschaftliche Erkenntnisse zu generieren. Je größer dabei die Quantität und die Heterogenität der Daten ist, desto anspruchsvoller wird die Aufgabe, eine gute Visualisierung zu erzeugen.

Es gibt bereits verschiedene Open Source Programme, die sich als Visualisierungstools eignen und bei denen es sich lohnt, über einen Einsatz nachzudenken. Neben generischen Visualisierungs-Tools (z. B. NodeBox 2015, Sgvizler 2015, d3SPARQL 2015), *Google Heatmaps* (Heatmaps 2015) und Zeitverlaufsanzeigen (z. B. über Timeline 2015; HDAT 2015) etwa durch Abfragen von Koordinaten aus dem TGN, lohnt es auch, einen Blick auf *RelFinder* (2015) zu werfen. RelFinder visualisiert die Beziehungen zwischen zwei Ressourcen und gibt auch den jeweiligen Inhalt der Ressource aus. Für alle diese Werkzeuge gilt es jedoch das mögliche Einsatzszenario sorgfältig zu entwerfen. Überlegungen und Kriterien für ein allgemeines Verfahren solcher Entwürfe sollen vorgestellt werden.

Die Zukunftsperspektiven für die Bereiche Vernetzung und Visualisierung werden im Salamanca-Projekt stets mitbedacht, und so umreißt der Vortrag bereits erbrachte Leistungen und Implementierungen, getroffene Entscheidungen aber auch noch offene und zu erprobende Themen: Insbesondere in Punkt 1 gehen viele Aspekte schon im Frühjahr 2016 in „production use“, in Punkt 2 werden einige konzeptuelle Fragen angesprochen, die wir für uns noch nicht gelöst haben, in Punkt 3 werden wir beschreiben, wie wir von einem eher zufälligen Spiel mit verschiedensten technischen Möglichkeiten zu einem systematischen Prozess der Reflexion über aktuelle und neue Entwicklungsziele übergehen.

Fußnoten

1. Z.B. Auflösung von Ressourcen-URLs, content negotiation, triple-store dump, Linked Data Fragments, SPARQL Endpoint usw.
2. Projekte: z. B. *Scholasticon* (Schmutz 2009) mit biographischen Informationen, die Sammlung *Post-Reformation Digital Library Scholastica* (PRDL 2013) mit Informationen zu Universitäten, Fakultäten und Lehrstühlen der Frühen Neuzeit oder *Early Modern Letters Online* (EMLO 2015) mit Daten zu frühneuzeitlichen Korrespondenzen. Normdatenbanken: z. B. *Getty Thesaurus of Geographic Names* (TGN 2015), *Gemeinsame Normdatenbank* der Deutschen Nationalbibliothek (GND 2015) oder der *CERL Thesaurus* (CERL 2015).
3. Die Anwendung *Linked Data Fragments Client* (LDFC 2015) zeigt etwa Beispiele, wie sich Nutzende Ressource(n) und Abfragen zusammenklicken können. Eine graphische Oberfläche für (begrenzte) Abfragen ermöglicht einen effizienten Umgang mit Linked Open Data, ohne dafür die Abfragesyntax beherrschen zu müssen. Auch über *LodLive* (LodLive 2012) ließe sich ein entsprechendes Abfrage-Interface realisieren.

Bibliography

- CERL** (2015): *CERL Thesaurus* <http://thesaurus.cerl.org> [letzter Zugriff 14. Oktober 2015].
- d3SPARQL** (2015): *d3sparql.js* Utilities for visualizing SPARQL results with the D3 library <http://biohackathon.org/d3sparql/> [letzter Zugriff 13. Oktober 2015].
- EMLO** (2015): *Early Modern Letters Online* <http://emlo.bodleian.ox.ac.uk/> [letzter Zugriff 14. Oktober 2015].
- GND** (2015): *Gemeinsame Normdatei* <http://www.dnb.de/DE/Header/Hilfe/kataloghilfe.html#doc207526bodyText73> [letzter Zugriff 14. Oktober 2015].
- HDAT** (2015): *Historical Dutch-Asiatic Shipping* <http://app.thebrownmap.nl/> [letzter Zugriff 13. Oktober 2015].

- Heatmaps** (2015): *Google Maps Javascript API* <https://developers.google.com/maps/documentation/javascript/heatmaplayer> [letzter Zugriff 13. Oktober 2015].
- LDLC** (2015): *Linked Data Fragments client* <http://client.linkeddatafragments.org/> [letzter Zugriff 13. Oktober 2015].
- LodLive** (2012): *LodLive* Browsing the Web of Data <http://en.lodlive.it/> [letzter Zugriff 13. Oktober 2015].
- NodeBox** (2015): *Meet NodeBox 3* <https://www.nodebox.net/node/> [letzter Zugriff 13. Oktober 2015].
- PRDL** (2013): *Scholastica* Early Modern Academies & Universities <http://www.prdl.org/schools.php> [letzter Zugriff 14. Oktober 2015].
- RelFinder** (2015): *RelFinder* Interactive Relationship Discovery in RDF Data <http://www.visualdataweb.org/relfinder.php> [letzter Zugriff 13. Oktober 2015].
- Schmutz, Jacob** (2009): *Scholasticon* Ressources en ligne pour l'étude de la scolastique moderne (1500-1800): auteurs, sources, institutions <http://scholasticon.ish-lyon.cnrs.fr> [letzter Zugriff 14. Oktober 2015].
- Sgvizler** (2015): *Sgvizler* <http://dev.data2000.no/sgvizler/> [letzter Zugriff 13. Oktober 2015].
- TGN** (2015): *The Getty Thesaurus of Geographic Names* (TGN) <http://vocab.getty.edu/> [letzter Zugriff 13. Oktober 2015].
- Timeline** (2015): *Leaflet.timeline* <https://github.com/skeate/Leaflet.timeline> [letzter Zugriff 13. Oktober 2015].
- xTriples** (2015): *Xtriples* A generic webservice to extract RDF statements from XML resources <http://xtriples.spatialhumanities.de/index.html> [letzter Zugriff 13. Oktober 2015].