

Dingler Dissemination Highlights aus 6 Jahren »Digitalisierung des Polytechnischen Journals«

Marius Hug, M. A.; Martina Gödel, M. A.;
Timo Arndt, B.A.; Una Schäfer, B.A.

Einreichung zum Call for Posters

DHd-Tagung 2015

Von Daten zu Erkenntnissen: Digitale Geisteswissenschaften als Mittler zwischen Information
und Interpretation

7. November 2014

Abstract

Am 28. Februar 2015 endet die Laufzeit des von der DFG geförderten Projekts »Digitalisierung des Polytechnischen Journals« am Institut für Kulturwissenschaft der Humboldt Universität zu Berlin. Damit ist die DHd 2015 genau der richtige Zeitpunkt, um Bilanz zu ziehen. Mit unserem Posterbeitrag möchten wir die wichtigsten Ergebnisse aus der kooperativ angelegten Projektarbeit präsentieren. Highlights sind dabei: Nachhaltige Datenspeicherung und -präsentation, Ausführliche Tagging- und Projektdokumentation per ODD, Zurverfügungstellen der Daten über geeignete Schnittstellen zu Analysezielen bspw. für Computerlinguisten, Aufbereitung der im TEI P5-Format vorliegenden Daten zur wissenschaftlichen Weiterverarbeitung (bspw. Umwandlung historischer Währungen, Visualisierung auf einer Timemap), Bearbeitung des sehr umfangreichen Bildmaterials mittels Image-Markup-Tool, sowie ein vollkommen neuartiger Transfer unserer Daten — aus der virtuellen Welt in die Welt der Objekte — als Grundlage für die Kooperation mit einem Museum.

1 Dingers Polytechnisches Journal (DPJ)

Das »Polytechnische Journal« wurde 1820 vom Augsburger Fabrikanten und Chemiker Johann Gottfried Dingler begründet. Dingler studierte wichtige Zeitschriften (die meisten davon aus England, Frankreich, später aber auch den USA), wählte relevante

Artikel aus, übersetzte und publizierte sie in seinem Journal. Mit einer Laufzeit von 111 Jahren ist diese Zeitschrift ein beispielloses, europaweites Archiv der Technik-, Wissens- und Kulturgeschichte. Besonders bemerkenswert ist die Aktualität der Publikation: So verging kaum Zeit zwischen Erstveröffentlichung der Artikel und Erscheinen der übersetzten Version im DPJ.

2 Das Digitalisierungsprojekt

Im von der DFG geförderten Projekt am Institut für Kulturwissenschaft der Humboldt-Universität zu Berlin wurde der komplette Bestand vom DPJ digitalisiert. Die Bilddigitalisierung wurde an der SLUB-Dresden durchgeführt. Für die Textdigitalisierung und Basisauszeichnung der über 200.000 Seiten war der Dienstleister Editura GmbH zuständig. Alle Bände sind per TEI-P5 kodiert. Das Journal ist online (CC by-nc-sa 3.0) unter www.polytechnischesjournal.de verfügbar.

Nachhaltigkeit, Clarin-D

Ein großes Problem für Digitalisierungsprojekte ist die nachhaltige Verfügbarmachung der Daten, Stichwort: Langzeitarchivierung. Für das Projekt Dingler-Online bedeutet die Zusammenarbeit mit dem BMBF geförderten Verbundprojekt **CLARIN-D** einerseits die Möglichkeit der Dissemination der Projektdaten, andererseits ist dadurch eine langfristige Sichtbarkeit des Projekts garantiert.

Durch die Kooperation mit der Berlin-Brandenburgischen Akademie der Wissenschaften, konkret dem DFG-Projekt **Deutsches Textarchiv** (DTA), profitiert Dingler-Online sehr direkt von deren technischem Know-how. Konkret zu nennen wären hier bspw. die orthographische Normalisierung und linguistische Analyse (POS, Tokenisierung, Lemmatisierung, ...) sowie die Nutzung der elaborierten Recherveschnittstelle.

Ausführliche Tagging- und Projektdokumentation per ODD

Alle editorischen Entscheidungen, die verwendeten Elemente und Attribute wurden in der ODD-Datei ausführlich beschrieben bzw. festgelegt. Für die eingesetzten Attribute wurden geschlossene Listen mit projektspezifischen Werten definiert und ihr Einsatz erklärt. Konkrete Quelltextbeispiele aus dem Projekt veranschaulichen das Vorgehen. Eine Transformation in das HTML-Format mittels des TEI-Tools OxGarage ermöglicht die Sichtbarmachung dieser Dokumentation im Look and Feel der TEI Guidelines selbst.

Das restriktiv formulierte Datenmodell und seine transparente Dokumentation unterstützen die möglichst schwellenarme automatisierte interne und externe Weiterverarbeitung. Die Dokumentation ist online unter <http://dingler.culture.hu-berlin.de/Schema/dingler.html> verfügbar.

Historische Daten

Ein wichtiges Thema – nicht zuletzt aufgrund der stets größer werdenden digital vorliegenden Datenmengen – ist die Visualisierung. Exemplarisch wurde in unserem Projekt ein solches Verfahren anhand von Patentdaten durchgeführt (s. Abb. 1). Diese eignen sich in besonderem Maße, da die Einheit Patentschrift sehr überschaubar ist, aber dennoch die für ein tief granuliertes TEI-Tagging benötigten Elemente enthält: Names, Dates, People, und Places (TEI P5 Guidelines, ch. 13). Der Workflow, mit Hilfe dessen die in Patentlisten vorliegenden Einträge in das zur Darstellung auf einer Timemap benötigte KML (Keyhole Markup Language) transformiert wurde, ist gut dokumentiert.

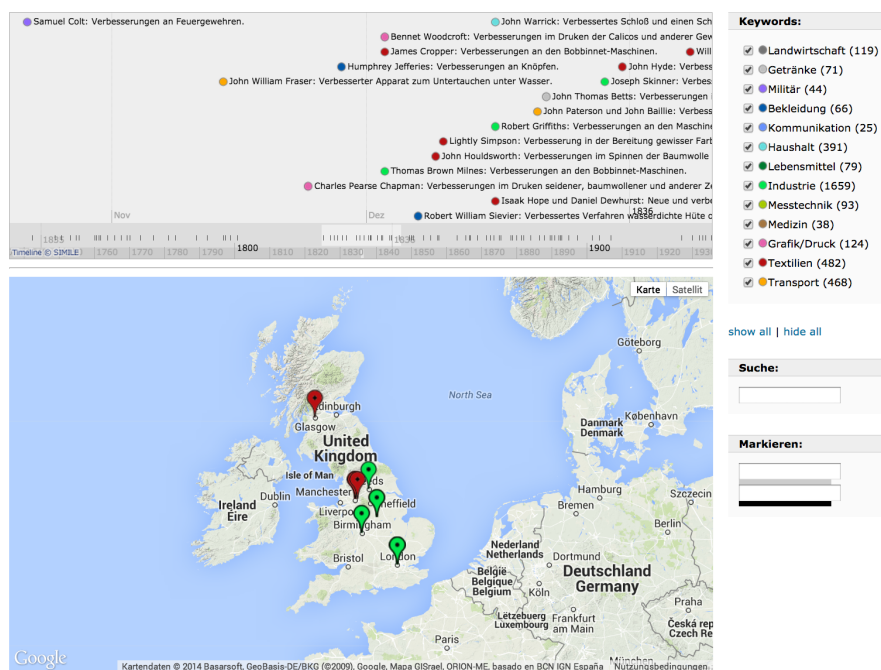


Abbildung 1: Visualisierung der TEI-Daten auf einer Timemap.

Historische Varianz stellt eine weitere Herausforderung für unsere Daten dar. Dies betrifft nicht nur die Texte, sondern auch Zahlen und Einheiten. Hier wurden die in den digitalisierten Daten vorkommenden Währungen und Einheiten gesammelt und über Listen entsprechende Umrechnungen zugewiesen, wobei sich bspw. für den Wiener Fuß folgendes Bild ergibt:

Image Markup Tool

Neben dem Textbestand (etwa 420 Mio. Zeichen) ist v. a. die große Menge an Bildern bzw. Zeichnungen hervorzuheben, wobei neben rund 3500 Falttafeln – eine Tafel enthält bis zu 114 Einzelfiguren – auch zahlreiche Figuren von Text umflossen gedruckt wurden. Um der Bedeutung des Bildmaterials im technischen Kontext gerecht zu werden, wurde

```
wienerfuss:
  name: Wiener Fuß
  unit: Wiener Fuß
  wp: http://de.wikipedia.org/wiki/Fuß_(Einheit)
  conversions:
    zentimeter: x * 31.608
    millimeter: x * 316.08
```

Abbildung 2: Syntax des projekteigenen Einheitenrechners.

hier bei der Erschließung besonders großer Aufwand betrieben.

Die Tafelwerke des »Polytechnischen Journals« wurden auf der Ebene der Einzelfiguren mit Koordinaten versehen. Der Mehrwert dieser Auszeichnungsstrategie besteht für den Nutzer zum einen in der konkreten Referentialisierung von Textpassage und Einzelfigur sowie individueller Anordnungs-, Betrachtungs- und Ausgabemöglichkeiten.

3 Museum

Hintergrund der von uns auf Grundlage unserer im Projekt aufbereiteten Daten angelegten Kooperation mit einem Museum – erste Prototypen sind gerade im Einsatz – ist folgende These: Weder ist Sammlung jenseits von Wissenschaft noch Forschung jenseits der Dinge möglich. Von einer Zusammenarbeit profitieren demnach beide Seiten. Ziel ist es, ausgewählte Objekte eines Museums multimedial erlebbar zu machen und damit den Besuchern ein neuartiges, interessantes Objekterlebnis zu ermöglichen. Andererseits erzeugt diese *Nachnutzung* unserer Daten eine längerfristige Sichtbarkeit der in den Projektdaten gespeicherten Informationen.

Mit Hilfe einer kostenlos zur Verfügung stehenden und speziell für das Projekt entwickelten App, werden dem Museumsbesucher zu ausgewählten Objekten vertiefende Informationen angeboten.

Denkbar sind hier bspw.:

- weiterführende Informationen zum Ausstellungsgegenstand und seiner Geschichte
- Informationen zu beteiligten Akteuren (Erfinder, Produzenten, Firmen...)
- veranschaulichende Bilder/Figuren
- verwandte Themenfelder, bspw. per Schlagwortwolke

Die App-Entwicklung wird bis zum Februar soweit sein, dass wir diese neuartige Nutzung unserer Daten tatsächlich hands-on vorführen können.¹

¹An dieser Stelle möchten wir darauf hinweisen, dass die Einreichung von Christian Kassung für einen Vortrag mit dem Titel »Making Things Chatter« eine weitere Technologie zur Verlinkung von Daten und Objekten präsentiert. Im Unterschied zu unserem Poster liegt der Fokus dort auf dem Museum.

4 Fazit

Zusammengefasst verfolgt unsere Posterpräsentation eine doppelte Strategie: 1) Das in der Community teilweise schon bekannte Projekt kann in verschiedenster Hinsicht Erfahrungen der letzten Jahre weitergeben und so für andere (kleinere) Digitalisierungsprojekte inspirierend sein. 2) Wir würden uns freuen, mit unserer neuartigen Idee einer Verknüpfung von Text- und Objektdaten (aus dem Museum) mit der Community in Diskussion zu treten und für das weitere Vorgehen von zu erwartenden Synergieeffekten zu profitieren.

5 Webressourcen

- CLARIN-D: <http://de.clarin.eu/de/>
- DinglerOnline: <http://www.polytechnischesjournal.de>
- Dingler ODD: <http://dingler.culture.hu-berlin.de/download>
- DTA: <http://www.deutschestextarchiv.de>
- Google Timemap: <https://code.google.com/p/timemap/>
- Image Markup Tool: http://tapor.uvic.ca/~mholmes/image_markup/
- KML: <https://developers.google.com/kml/>