

An den Grenzen der Interoperabilität: Eine kritische Reflexion über digitale Forschungsdaten und -anwendungen in der Online-Edition des Projekts "Die Schule von Salamanca"

Wagner, Andreas

wagner@rg.mpg.de

Max-Planck-Institut für europäische Rechtsgeschichte,
Frankfurt am Main

Glück, David

glueck@rg.mpg.de

Akademie der Wissenschaften und der Literatur, Mainz;
Johann Wolfgang Goethe-Universität Frankfurt am Main

Einleitung

Der Beitrag diskutiert kritisch, welche nicht allein arbeitsökonomischen, sondern vor allem intellektuellen, methodologischen und wissenschaftstheoretischen "Kosten" mit der Digitalisierung und dem dadurch etablierten Fokus auf Fragen der Interoperabilität entstehen. Die deutlichen wissenschaftlichen Vorteile von Interoperabilität müssen nämlich einem Umbau (und z.T. Rückbau) in wissenschaftlichen Modellierungen und im Verständnis wissenschaftlicher Erkenntnisvermehrung gegenüber gestellt werden. Dabei wird im Sinne des doppelten Genitivs "Kritik der digitalen Vernunft" sowohl das naheliegende Phänomen diskutiert, dass die digitale Transformation Hoffnungen weckt und Lösungen wissenschaftlich-methodologischer Schwierigkeiten nahelegt, die sich (z.T. erst bei der Implementierung) als wissenschaftlich nicht akzeptabel erweisen, als auch der umgekehrte Fall, dass die digitale Transformation ein kritisches Umdenken in der Ausrichtung der eigenen wissenschaftlichen Arbeit und eine Umorientierung wissenschaftlicher Ambitionen erzwingt.

Diese Diskussion wird im Durchgang durch einige beispielhafte Entwicklungen im seit vier Jahren laufenden Projekt "Die Schule von Salamanca" geführt, bevor ein Versuch der Verallgemeinerung unternommen wird. Dass Interoperabilität signifikante wissenschaftliche Kosten mit sich bringt, heißt im Übrigen nicht, dass diese nicht womöglich durch die Vorteile aufgewogen würden. Es ist

allerdings in der Orientierung der Projektarbeit wichtig, sich an den „Grenzen der Interoperabilität“ abzuarbeiten und sich regelmäßig beide Seiten dieser Bilanz vorzulegen.

Interoperabilität

Mit der digitalen Transformation sind die Möglichkeiten des überregionalen und interdisziplinären Austauschs und der Weiterverwendung von Forschungsdaten in einer ganz neuen Weise möglich geworden. So hat sich der Begriff der Interoperabilität als zentrales Paradigma in den Digital Humanities etabliert, um den Anspruch zu beschreiben, diese Möglichkeiten methodisch auszubauen und ein Evaluationskriterium für Forschungsleistungen und -ergebnisse anzubieten. Interoperabilität ist auf mehreren Ebenen zu verstehen (vgl. Gradmann 2009) und wird auf den konkreten technischen und syntaktischen Ebenen vor allem durch Standards für Datenformate und Schnittstellen, sowie auf der pragmatischen Ebene u.a. durch den Bezug auf Normdaten realisiert. Mag Interoperabilität auf der semantischen Ebene schließlich vor einiger Zeit noch als ein utopisches Ziel angemutet haben (vgl. Baumann 2011), so haben sich etwa durch die Selbstreflexion digitaler geisteswissenschaftlicher Forschungs-Arbeit (Hughes et al. 2016) oder in der Beschreibung von kulturellen und geistigen Phänomenen (Le Boeuf et al. 2017; Stead et al. 2015) deutliche Fortschritte in der Modellierbarkeit geisteswissenschaftlicher Phänomene und in der Verfügbarkeit und Relationierbarkeit von semantischen Forschungsdaten und #ergebnissen ergeben.

Wie Forschungsprojekte den derart gewachsenen Interoperabilitätserwartungen gerecht werden können, kann jedoch kaum allgemein, sondern nur im Rahmen ihrer jeweils spezifischen Arbeit ermittelt werden (zu einer projektunabhängig formulierten Handreichung vgl. aber Beer et al. 2014). Dabei stellt sich die Frage nach dem Nutzen, den Grenzen und den Kosten interoperabler Techniken im Projekt "Die Schule von Salamanca" insofern auf besondere Weise, als dieses einerseits über einen langen Zeitraum und mit Blick auf eine langzeitverfügbare Erschließung von relativ großen (Text-)Datenmengen operiert, andererseits aber auch die Erforschung innovativer Textaufbereitungsverfahren und Webanwendungsfunktionen zur Ermöglichung neuartiger Forschungserkenntnisse für die beteiligten Fachwissenschaften anstrebt.

Das Projekt

Im durch die Akademie der Wissenschaften und der Literatur | Mainz geförderten und insgesamt auf 18 Jahre angelegten Projekt "Die Schule von Salamanca. Eine digitale Quellensammlung und ein Wörterbuch ihrer juristisch-politischen Sprache" (Duve et al. 2013) werden voraussichtlich insgesamt etwa 120 Texte der gleichnamigen Schule iberischer Theologen und Juristen

des 16. und 17. Jahrhunderts nach und nach digitalisiert und als Volltexte erfasst. Die in TEI-XML ausgezeichneten und aufwändig normalisierten Texte werden nicht nur strukturell, sondern auch im Hinblick auf als Linked Open Data (LOD) referenzierbare Entitäten – etwa Personennamen – erschlossen; dabei werden die Text-Digitalisate, die Volltexte und die LOD-Datensammlungen online bereit gestellt. Hinzu kommt ein digitales (und schließlich auch gedrucktes) Wörterbuch, in dem sowohl biographische Informationen zu den in der Edition vertretenen Autoren als auch zentrale Begriffe der Rechts- und politischen Ideengeschichte und deren Entwicklung im Diskussionszusammenhang der "Schule von Salamanca" erfasst werden. In diesen beiden Säulen des Projekts ist die Erschließung und Repräsentation der Struktur der internen Verweise (Autoren, die sich wechselseitig zitieren, Wörterbuchartikel, die auf Textstellen verweisen) eines der zentralen wissenschaftlichen Ergebnisse.

Die sowohl für die Benutzung als auch für die Bereitstellung der Daten als zentrales Portal dienende Webseite des Projekts (<https://salamanca.school/>) ist technisch in Form einer komplex modularisierten Webanwendung implementiert, die fachwissenschaftlichen BenutzerInnen eine Vielzahl von Funktionen bieten soll, z.B. eine geräteübergreifende und performante Darstellung der Editionstexte, eine intelligente Suchfunktion, die den frühmodern-lateinischen und -spanischen Volltexten angepasst ist, und eine feinkörnige Referenzierung der Texteinheiten. Durch technische Mechanismen (content negotiation, API, RESTful Microservice-Architektur), Exportfunktionen (abschnitts-, text- oder corpusweise, plaintext-, TEI- oder andere Formate) und die Orientierung an Formatstandards wird Anforderungen der Interoperabilität explizit Rechnung getragen. Der Quelltext der Webanwendung wird bis Januar 2018 ebenfalls veröffentlicht und fortan in Open Source weiterentwickelt werden.

Interoperabilitäts-"Konflikte": Einige Beispiele

Einhergehend mit der Open Source-Veröffentlichung der Webanwendung soll über einige repräsentative Aspekte der Projektarbeit reflektiert werden, in denen sich an vermeintlich technischen Herausforderungen Konflikte um die wissenschaftlichen Implikationen von Interoperabilität entzündeten:

a) Textrepräsentation:

Während in diesem Kernbereich der Projektarbeit gegenwärtig noch mit einem eigens spezifizierten und den bisherigen Forschungsanforderungen entsprechenden "idiosynkratischen" TEI-Format gearbeitet wird und die Datenmodellierung somit eher als "research-driven" (vgl.

Flanders/Jannidis 2016: 233) bezeichnet werden kann, wird eine langfristige Verfügbarmachung der Texte in einem interoperablen, "curation-driven" Format wie etwa *TEI Simple* (Text Encoding Initiative Consortium 2016) erwägt. Dies ist nicht zuletzt auch durch den Blick auf die Wissenschaftsförderung motiviert (vgl. etwa DFG 2015, Anhang), die die Interoperabilität von Textauszeichnungen dem erst noch nachzuweisenden Erkenntnisgewinn 'reicher' Annotationen gegenüberstellt. Nun beinhaltet das XML-Ökosystem der Edition allerdings auch Informationen wie etwa Metadaten zur Zeichenkodierung (die nicht zuletzt Interoperabilitäts-Funktionen erfüllen), die sich nach derzeitigem Stand nicht ohne weiteres in *TEI Simple* abbilden lassen. So ergibt sich eine nur durch einigen Aufwand aufzulösende Spannung zwischen detaillierter wissenschaftlicher Gegenstandsbeschreibung und Maßnahmen zur Förderung der Nutzbarkeit der eigenen Ressourcen (z.B. des Angebots mehrerer alternativer Datenformate), und es stellt sich die Frage, wann und durch wen jener Aufwand erbracht werden soll. Selbst eine gegenüber den "Experten"-Annotationen tolerantere Zielvorgabe, die sich etwa am Konzept des 'Interchange' (als einer durch menschliche Interpretation vermittelten Nachnutzung, vgl. Holmes 2017, Baumann 2011) orientiert und im Vergleich zum Modell einer bruchlosen Weiterverwertbarkeit der Daten durch automatische Prozesse gemäßigte Ansprüche erhebt, sieht sich ähnlichen Fragen der Standardisierung von Schnittstellen und Datenformaten ausgesetzt.

b) Modulare Infrastruktur:

Im Zuge der Einrichtung einer Linked Open Data-Infrastruktur haben wir *content-negotiation*-Mechanismen, Weiterleitungen und die Adressierung unterschiedlicher Funktionen über verschiedene Server und Server-Adressen eingeführt. Diese Entwicklung legt eine Fortsetzung nahe, die den Umbau der Web-Anwendung insgesamt in ein Ensemble von Microservices bedeuten würde (vgl. Wolff 2016), in dem Daten und Dienste aufs Engste verschränkt sind. Während dies die Interoperabilität, d.h. konversions- und barrierearme Nutzbarkeit der Daten enorm verbessert (etwa dadurch, dass auf verschiedenen Ebenen in verschiedenen Formaten und verschiedenen Granularitäten Daten und Dienste verfügbar sind, vgl. Turska et al. 2016) hat es jedoch auch den Nachteil, dass der Daten- und Anwendungszusammenhang – der eben auch als *Zusammenhang* eine Forschungsleistung darstellt – nur in einer sehr viel aufwändigeren Weise repliziert und ggf. archiviert werden kann: Während erste Infrastrukturen die Archivierung und langfristige Zugänglichkeit von Javascript-Anwendungen erlauben sollen (vgl. Bingert/Buddenbohm 2016; kritischer: Brunelle 2016), so ist dies für eine solche Infrastruktur mit mehreren kooperierenden und kommunizierenden Servern nur sehr viel schwerer vorstellbar.

c) Adressierung, Versionierung und Persistenz:

Im Zusammenhang mit der Adressierung von einzelnen Textpassagen haben wir ein System eingeführt, das sowohl semantische Aussagen über Text-Entitäten als auch die Realisierung der Verweisstrukturen auf implementierungs- und plattformunabhängige, intellektuell intuitive Weise erlaubt (im Anschluss an das Canonical Text Services-Schema, Blackwell/Smith 2014; vgl. Wagner 2016). Diese Struktur der komplexen Verknüpfung von Ressourcen und Entitäten untereinander ver trägt sich aktuell nicht mit etablierten Methoden, Dokumente persistent zu identifizieren und die Überarbeitungshistorie der Ressourcen in einer Versionierung transparent und nachvollziehbar zu machen. Diese Methoden spezifizieren nämlich zumeist den Umgang mit einem ganzen Dokument und vernachlässigen den Bedarf, Entitäten innerhalb des Dokuments (persistent) zu referenzieren oder die Einbettung jenes Dokuments in ein Corpus zu verwalten. Wenn beispielsweise ein Dokument verändert wird, muss es unter anderem normalerweise eine neue persistente ID erhalten - damit entsteht ein Aktualisierungs- oder mindestens Kontrollbedarf bei den Querverweisen innerhalb des Dokuments sowie in allen weiteren Dokumenten des Corpus (und in unserem Falle in allen Artikeln des Wörterbuchs), welche Verweise auf das aktuelle Dokument enthalten. Mechanismen wie das Memento framework (van de Sompel/Nelson 2015) bieten eine transparente und flexible Versionierung, die für Web-Dokumente wie für Semantische Ressourcen gleichermaßen funktioniert, sind jedoch in ihrer Integration mit PID-Systemen ebenfalls noch nicht erprobt. Versteht man diese Motive in einem Interoperabilitäts-Zusammenhang, lesen sie sich wie ein Konflikt zwischen etablierten Lösungen auf verschiedenen Interoperabilitäts-Ebenen und es fragt sich, wie kanonische Verweissysteme (funktional-pragmatische Ebene nach Gradmann 2009) und Persistent Identifiers (technische Ebene) miteinander harmonisieren können.

Diskussion

Die hier am Beispiel digitaler Projektarbeit aufgezeigten Probleme etwa einer gesteigerten Spannung zwischen „Interoperabilität und Expressivität“ (Baumann 2011) weisen auf grundlegende wissenschafts- und erkenntnistheoretische Fragestellungen hin, die nicht zuletzt das Selbstverständnis der Digital Humanities und die mit ihnen verbundenen Hoffnungen berühren: Ist Expertenwissen am Ende vielleicht gar nicht "interoperabel"? Erweisen sich die Möglichkeiten, durch Verknüpfung digitaler Forschungsdaten und #anwendungen verschiedenster Kontexte neuartige Einsichten zu generieren, als stark begrenzt? Eine wissenschaftspolitische Pointe des Beitrags besteht so darin, auf die entscheidende Rolle aufmerksam

zu machen, die die Erwartung darüber spielt, an welchem Ort – zwischen den Disziplinen oder "tief" in den fachwissenschaftlichen Spezialdiskursen – wissenschaftlicher Fortschritt erzielt wird.

Bibliographie

Baumann, Syd (2011): "Interchange vs. Interoperability", in: *Proceedings of Balisage: The Markup Conference 2011* 7, <https://doi.org/10.4242/BalisageVol7.Bauman01> [letzter Zugriff: 13.01.2018].

Beer, Nikolaos / Herold, Kristin / Kolbmann, Wibke / Kollatz, Thomas / Romanello, Matteo / Rose, Sebastian / Walkowski, Niels-Oliver (2014): "Interdisciplinary Interoperability", DARIAH-DE Working Papers Nr. 3, <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2014-1-0> [letzter Zugriff: 18.01.2018].

Bingert, Sven / Buddenbohm, Stefan (2016): "Die HDC-Anwendungskonservierung - ein Dienst zur Archivierung und Bereitstellung komplexer Forschungsergebnisse", in: *GWDG-Nachrichten* 11/2016: 7-9 https://www.gwdg.de/documents/20182/27257/GN_11-2016_www.pdf [letzter Zugriff: 23.09.2017].

Blackwell, Christopher / Smith, Neel (2014): "The Canonical Text Services protocol, version 5.0.rc.2", https://cite-architecture.github.io/cts_spec/ [letzter Zugriff: 25.09.2017].

Brunelle, Justin F. / Kelly, Mat / Weigle, Michele C. / Nelson, Michael L. (2016): "The impact of JavaScript on archivability", in: *International Journal on Digital Libraries* 17/2: 95–117, <https://doi.org/10.1007/s00799-015-0140-8> [letzter Zugriff: 25.09.2017].

DFG (2015): "Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft", http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/foerderkriterien_editionen_literaturwissenschaft.pdf [letzter Zugriff: 12.1.2018].

Duve, Thomas / Lutz-Bachmann, Matthias / Birr, Christiane / Niederberger, Andreas (2013): "Die Schule von Salamanca: eine digitale Quellensammlung und ein Wörterbuch ihrer juristisch-politischen Sprache. Zu Grundanliegen und Struktur eines Forschungsvorhabens". Mainz: Akademie der Wissenschaften und der Literatur. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:hebis:30:3-324011> [letzter Zugriff: 12.1.2018].

Flanders, Julia / Jannidis, Fotis (2016): "Data Modeling", in: Schreibman, Susan / Siemens, Ray / Unsworth, John (eds.): *A New Companion to Digital Humanities*. Chichester: Wiley Blackwell 229-237.

Gradman, Stefan (2009): "Interoperability. A key concept for large scale, persistent digital libraries". *Digital Preservation Europe* <http://citeseerx.ist.psu.edu/viewdoc/download?>

doi=10.1.1.363.6311&rep=rep1&type=pdf [letzter Zugriff: 25.09.2017].

Holmes, Martin (2017): "Whatever happened to Interchange?", in: *Digital Scholarship in the Humanities* 32 (Issue suppl_1): i63–i68 <https://doi.org/10.1093/llc/fqw048> [letzter Zugriff: 25.09.2017].

Hughes, Lorna / Constantopoulos, Panos / Dallas, Costis (2016): "Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines", in: Susan Schreibman / Ray Siemens / John Unsworth (eds.), *A New Companion to Digital Humanities*, Wiley & Sons, 150–170, <https://doi.org/10.1111/b.9781118680643.2016.00013.x> [letzter Zugriff: 13.1.2018].

Le Boeuf, Patrick / Doerr, Martin / Ore, Christian Emil / Stead, Stephen et al. (2017): "Definition of the CIDOC Conceptual Reference Model", Version 6.2.2, http://www.cidoc-crm.org/sites/default/files/2017-09-30%23CIDOC%20CRM_v6.2.2_esIP.pdf [letzter Zugriff: 13.1.2018].

Schmidt, Desmond (2014): "Towards an Interoperable Digital Scholarly Edition", in: *Journal of the Text Encoding Initiative* 7, <https://doi.org/10.4000/jtei.979> [letzter Zugriff: 13.1.2018].

Stead, Stephen / Doerr, Martin et al. (2015): "CRMinf: the Argumentation Model. An Extension of CIDOC-CRM to support argumentation", <http://www.cidoc-crm.org/crminf/sites/default/files/CRMinf-0.7%28forSite%29.pdf> [letzter Zugriff: 13.1.2018].

Text Encoding Initiative Consortium (2016): "TEI Simple" <https://github.com/TEIC/TEI-Simple> [letzter Zugriff: 25.09.2017].

Turska, Magdalena / Cummings, James / Rahtz, Sebastian (2016): "Challenging the Myth of Presentation in Digital Editions", in: *Journal of the Text Encoding Initiative* 9 <http://jtei.revues.org/1453> [letzter Zugriff: 25.09.2017].

van de Sompel, Herbert / Nelson, Michael L. (2015): "Reminiscing about 15 Years of Interoperability Efforts", in: *D-Lib Magazine* 21 (11/12) <http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html> [letzter Zugriff: 25.09.2017].

Wagner, Andreas (2016): "What's in a URI? Part I: The School of Salamanca, the Semantic Web and Scholarly Referencing" <https://blog.salamanca.school/2016/11/15/whats-in-a-uri-part-1/> [letzter Zugriff: 22.09.2017].

Wolff, Eberhard (2016): "Microservices. Grundlagen flexibler Softwarearchitekturen". Heidelberg: dpunkt.verlag.