

GOEDOC – Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

2019

Briefe kommentieren im Semantic Web
–
Ein Konzept

Stefan Dumont

DARIAH-DE Working Papers

Nr.33

Dumont, Stefan: Briefe kommentieren im Semantic Web : Ein Konzept
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2019
(DARIAH-DE working papers 33)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2019-5>

URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2019-5-8>

Dieser Beitrag erscheint unter der Lizenz [Creative-Commons Attribution 4.0 \(CC-BY\)](https://creativecommons.org/licenses/by/4.0/)



Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

Erschienen in der Reihe
DARIAH-DE working papers

ISSN: 2198-4670

Herausgeber der Reihe
DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

Abstract: Der Aufsatz stellt ein Konzept vor, wie mit Hilfe von Technologien des Semantic Web edierte Briefe semantisch kommentiert werden können. Dabei werden Erwähnungen von Personen, Orten, Schriften etc. nicht nur wie bisher identifiziert, sondern auch typisiert. Diese so notierten Aussagen werden als „named graphs“ mit Hilfe des Web Annotation Data Models mit Metadaten zum Brief und zur Annotation selbst in Verbindung gebracht und dadurch kontextualisiert. Die so maschinenlesbar aufbereiteten Informationen sind als „Linked Open Data“ mit Daten aus anderen Editionen, Forschungsprojekten und Nachschlagewerken automatisch vernetzt und können mit Hilfe von SPARQL abgefragt werden. Der Beitrag stellt die Methode des „Semantischen Kommentierens“ anhand konkreter Beispiele vor und diskutiert deren Grenzen und Nutzen für die Wissenschaft.

Keywords: Semantic Web, Digitale Edition, Briefe, Kommentar

Semantic Web, Digital Scholarly Edition, Correspondence, Commentary

Briefe kommentieren im Semantic Web

Ein Konzept

Stefan Dumont

Berlin-Brandenburgische Akademie der Wissenschaften



Stefan Dumont: „Briefe kommentieren im Semantic Web“. *DARIAH-DE Working Papers* Nr. 33. Göttingen: DARIAH-DE, 2019. URN: [urn:nbn:de:gbv:7-dariah-2019-5-8](https://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2019-5-8).

Dieser Beitrag erscheint unter der
Lizenz [Creative-Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/) (CC-BY).

Die *DARIAH-DE Working Papers* werden von Mirjam Blümm,
Thomas Kollatz, Stefan Schmunk und Christof Schöch
herausgegeben.

Mitherausgeber dieses Working Papers ist Andreas Kuczera
(ADWL Mainz)



Dieser Beitrag ist im Rahmen der Tagung Graphentechnologien entstanden, die am 19. und 20. Januar 2017 an der Akademie der Wissenschaften und der Literatur Mainz (ADWL Mainz) in Zusammenarbeit mit DARIAH-DE und Historical Network Research (HNR) stattfand.

Timo Kissinger (ADWL Mainz) sei gedankt für die überaus konstruktive Unterstützung des Lektorats.

Zusammenfassung

Der Aufsatz stellt ein Konzept vor, wie mit Hilfe von Technologien des Semantic Web edierte Briefe semantisch kommentiert werden können. Dabei werden Erwähnungen von Personen, Orten, Schriften etc. nicht nur wie bisher identifiziert, sondern auch typisiert. Diese so notierten Aussagen werden als „named graphs“ mit Hilfe des Web Annotation Data Models mit Metadaten zum Brief und zur Annotation selbst in Verbindung gebracht und dadurch kontextualisiert. Die so maschinenlesbar aufbereiteten Informationen sind als „Linked Open Data“ mit Daten aus anderen Editionen, Forschungsprojekten und Nachschlagewerken automatisch vernetzt und können mit Hilfe von SPARQL abgefragt werden. Der Beitrag stellt die Methode des „Semantischen Kommentierens“ anhand konkreter Beispiele vor und diskutiert deren Grenzen und Nutzen für die Wissenschaft.

Schlagwörter

Semantic Web, Digitale Edition, Briefe, Kommentar

Keywords

Semantic Web, Digital Scholarly Edition, Correspondence, Commentary

Inhaltsverzeichnis

1	Einleitung	4
1.1	Briefeditionen und Kommentar	4
1.2	Digitales Kommentieren	5
2	Semantisch Kommentieren	9
2.1	Grundlage: Ontologien	9
2.2	Aussagen	10
2.3	Kontexte	11
2.4	Eingabewerkzeuge	13
3	Diskussion	15
3.1	Kommentar oder doch Regest?	15
3.2	Grenzen	15
3.3	Datenanalyse – auch jenseits der Edition	16
4	Fazit	19
	Literaturverzeichnis	21

1 Einleitung

1.1 Briefeditionen und Kommentar

Werden Editionen von Briefwechseln angefertigt, stehen – im Gegensatz zur Edition von Werken – Fragen der Textgenese und der Varianten in der Regel im Hintergrund. Ediert werden Briefe immer nach dem authentischsten Textzeugen: in der Regel also der versendeten Handschrift. Andere Fassungen werden – über die bloße Nennung hinaus – meistens nur dann für die Edition berücksichtigt, wenn die versendete Handschrift nicht mehr vorliegt oder Konzepte bzw. Entwürfe existieren, die signifikant vom versendeten Text abweichen und diesen in seiner Bedeutung verändern. Die Verzeichnung von Varianten findet demgemäß in Briefeditionen nur selten statt (Woesler 1998, 88, 90; Mücke 2001, 99).¹ Auch die zu erfassenden Textbearbeitungen durch den Autor (oder den Empfänger) halten sich in Grenzen. Zwar können in (vor allem freundschaftlichen) Briefen Überarbeitungen enthalten sein, diese bleiben aber im überschaubaren Rahmen. Würde der Brief zu stark bearbeitet sein, wäre dies dem Empfänger gegenüber unhöflich, was schon seit jeher vermieden wird. Seit der Jahrtausendwende nimmt die Editionswissenschaft und die historische Kulturwissenschaft die Gestaltung und die Materialität des Briefes mehr in den Blick und diskutiert deren Bedeutung für den Empfänger und zeitgenössischen Leser. In der editionsphilologischen Alltagsarbeit erlangt die Materialität allerdings nur in Ausnahmefällen Gewicht und verlangt erst dann besondere Aufmerksamkeit, die über die – mittlerweile übliche – Beigabe von digitalen Faksimiles hinausgeht.

Während also Fragen zur Textgrundlage, zur philologischen Erfassung und zur Materialität bei der Edition eines Briefes eine kleinere Rolle spielen, stellt vor allem die inhaltliche Erschließung, genauer gesagt die Kommentierung die Bearbeiterin vor Herausforderungen. Versteht man den „Kommentar“ editionswissenschaftlich gesehen im weiten Sinne, gehört dazu auch die Frage der Textgrundlage (Konzept, Handschrift, Abschrift oder Druck) sowie die der Datierung (Woesler 1998). Letzteres berührt auch schon ein Spezifikum der Textsorte „Brief“: ein Brief wird von jemanden an jemanden gesendet. Er ist – neben anderem – ein Kommunikationsakt zwischen zwei oder mehr, aber bestimmbar Absendern bzw. Empfängern. Seit jeher wird daher für einen Brief in Editionen die Kopfzeile angegeben, die neben den Korrespondenten auch mindestens Schreibort und Datum nennt. Darüber hinaus wird auch der Korrespondenzzusammenhang angegeben, d. h. auf welches Schreiben ein Brief antwortet bzw. mit welchem Brief er beantwortet wird (Woesler 1998, 93). Weiterhin wird unter „Kommentar“ häufig auch die historische Einführung zu einem Band (d. h. zu einem Briefwechsel oder einem Zeitabschnitt in einer Gesamtausgabe) verstanden (Woesler 1993, 13).

Zum Kommentar gehören aber vor allem die sog. „Wort- und Sacherläuterungen“ (oder auch: „Einzelstellenkommentare“), die die Aufgabe haben, der Benutzerin „bestimmte sprachliche Wendungen oder im Text erwähnte inhaltliche Bezüge, deren Kenntnis man bei ihm nicht oder nicht mehr voraussetzen kann, zu erklären [...]“ (Hagen 1988, 211). Bei Briefen tritt neben die historische Distanz der modernen Leserinnen noch der Aspekt, dass Briefe auf einen (oder mehrere) bestimmte Adressaten und deren Wissenshorizont hin geschrieben worden sind. Das hat zur Folge, dass gerade Briefe viele Andeutungen und Nennungen enthalten, die sich der (nicht nur modernen) Leserin nicht erschließen. Hauptaufgabe

¹Strobel (2013) skizziert allerdings für *digitale* Briefeditionen zu Recht auch Erweiterungsmöglichkeiten hinsichtlich der Darbietung aller vorhandenen Textzeugen.

der Sacherläuterungen gerade bei Briefeditionen ist es also, „Texte erklärend zu begleiten, um dem Leser deren Verständnis zu erleichtern.“ (Hagen 1988, 209–10).

Diese Sacherläuterungen kann man nach Hagen in folgende Kategorien einteilen, die auch für Briefeditionen ihre Gültigkeit haben:

1. Identifizierung von erwähnten Entitäten (Personen, Orte, Werke etc.)
2. Identifizierung von Zitaten oder Paraphrasen
3. Erläuterungen historischer Ereignisse (sowohl weltgeschichtlicher als auch nur für die Korrespondenzpartner von Bedeutung)
4. Übersetzungen und sprachgeschichtliche Hilfen

Für die Benutzerin treten diese Kommentierungen in gedruckten Briefeditionen allerdings nicht immer als Fußnote in Erscheinung. Vielmehr wurde – und wird – versucht, den Einzelstellenkommentar (d. h. Fußnoten) zum einen durch die historische Einführung, zum anderen durch das Register zu entlasten (Mücke 2001, 101; Kruse 2001, 176). So werden übergreifende Ereignisse bzw. Sachverhalte zwischen Korrespondenten schon in der Einführung erläutert. Die Identifizierung von Personen und anderen Entitäten wird dagegen im Register dargeboten, indem die Person dort unter den häufig auftauchenden Namensformen und mitsamt den weitergehenden Informationen (Leistungsdaten, Beruf etc.) vermerkt werden (Stockmar 2005, 128).

1.2 Digitales Kommentieren

In der *digitalen* Briefedition ist Sinn und Zweck des Kommentars grundsätzlich erstmal derselbe geblieben: Identifikation und Erläuterung, um der Leserin das Verständnis zu erleichtern. Verändert haben sich allerdings die Erarbeitung und Darbietung des Kommentars: So ist die für eine Identifizierung notwendige Recherche durch die neuen, umfassenden Suchmöglichkeiten in Lexika, zeitgenössischen Publikationen und moderner Forschungsliteratur natürlich erheblich erleichtert worden.² Zum anderen können Namen und Werktitel in digitalen Editionen einfach als Personennamen markiert und mit einem entsprechenden Registereintrag verknüpft werden. Dadurch kommt es gleichzeitig auch zu einem „Shift“ von Namen hin zu Entitäten: zwar werden immer noch alle möglichen Namensvarianten zu einer Person markiert, identifiziert wird eine erwähnte Person allerdings nun durch eine ID, die im Idealfall sogar projektübergreifend gültig ist – etwa aus der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek.

Dieser Shift ermöglicht auch ein wesentliches Charakteristikum digitaler Briefeditionen: Anhand von Normdateien, BEACON- und CMIF-Schnittstellen können sie untereinander und mit anderen Angeboten im Web automatisiert vernetzt werden (Stadler 2012; 2014; Dumont 2016; 2018a). Das kommt natürlich den heterogenen Inhalten von Briefen und deren Netzwerkcharakter sowie dem selektiven Lesen von Briefeditionen mehr als entgegen.

²Die Arbeit der Identifikation muss jedoch auch heute noch durch die Wissenschaftlerin besorgt werden. Eine automatisierte Identifizierung scheitert an der doch häufigen und nicht maschinell überwindbaren Uneindeutigkeit von Erwähnungen. Dies wird auf absehbare Zeit auch noch so bleiben.

Durch die digitale Vernetzung ist nun eine weitere Entlastung des Kommentars möglich, die der Entlastung des Einzelstellenkommentars durch die Register im Druckzeitalter ähnelt. So können die Benutzerinnen von digitalen Briefeditionen gerade anhand von IDs aus Normdateien weitergehende Informationen zu einer Entität, z. B. einer Person, mit wenigen Klicks in anderen Briefeditionen oder Nachschlagewerken finden. Dadurch kann sich der Registereintrag in einer digitalen Briefedition auf einige wenige Grundinformationen beschränken und darüber hinaus nur spezifische Hinweise vermerken, etwa solche Informationen, die das Verhältnis zur Person betreffen, der die Edition gewidmet ist, oder solche Informationen, die bisherigen, anderen widersprechen.

Trotz des „digital turns“ und der damit verbundenen neuen Möglichkeiten in Erarbeitung und Darbietung, bleibt jedoch der Kommentar in digitalen Briefeditionen – abgesehen von den Kopfdaten des Briefes – bisher noch stark dem Druck-Zeitalter verhaftet. Zwar werden die Daten vordergründig maschinenlesbar abgelegt, tatsächlich ist mit Hilfe der TEI-XML-Auszeichnung und Normdaten-ID allerdings nur die Tatsache der Erwähnung einer Person und deren Identifizierung maschinell lesbar – mehr nicht. Das gilt auch für die sehr wenigen digitalen Editionen neuzeitlicher Texte, die Semantic Web-Technologien einsetzen. Diese setzte z. B. das Akademienvorhaben „Johann Friedrich Blumenbach-Online“ bereits vor ein paar Jahren ein, um Beziehungen zwischen den edierten Texten und den dort erwähnten Objekten zu modellieren (Wettlaufer u. a. 2015; Wettlaufer 2018). Dabei wurde aber nur die Beziehungsart „verweist auf“ im Sinne einer Identifikation verwendet, d. h. die Art der Erwähnung nicht weiter qualifiziert.³ Auch die digitale Briefedition „Burckhardtsource“⁴ setzte zur Annotation auf Semantic Web-Technologien, benutzte diese aber – soweit ersichtlich⁵ – ebenfalls nur zum Identifizieren der erwähnten Namen (Giannaccini 2017). In beiden Fällen wurde Semantic-Web-Technologie also vor allem im Sinne des „Linked Open Data“-Konzeptes eingesetzt. Dadurch sind die Erwähnungen zwar als RDF abfragbar, aber ihre Aussagekraft geht nicht über die Tatsache der Erwähnung – also die bloße Identifizierung – hinaus.⁶

Das wirkt angesichts der Vielfalt der Erwähnungsmöglichkeiten – gerade in Briefen – unbefriedigend. Einige wenige Beispiele⁷ aus dem Briefwechsel des Arztes, Anatomen und Naturforschers Samuel Thomas (von) Soemmerring (1751–1830) mögen das veranschaulichen:

Beispiel 1:

„Erlauben Sie mir Ihnen Herrn Professor Doct. Goldfuss aus Erlangen, einen äusserst braven, bescheidenen, sehr unterrichteten Mann zu einer freundlichen Aufnahme bestens zu empfehlen.“⁸

³Für die Modellierung der Texte und Textteile untereinander sowie für die Metadaten zu den Objekten wurden natürlich noch andere Beziehungstypen definiert. Diese betreffen aber nicht die Art der Erwähnung eines Objekts im Text.

⁴<http://burckhardtsource.org/>

⁵D.h. soweit dies aus den Richtlinien (Müller 2015) und stichprobenhaften Überprüfungen anhand des „Advanced Annotation View with pundit“ (zu pundit siehe weiter unten) hervorgeht.

⁶Im Bereich der digitalen Edition mittelalterlicher Rechnungsbücher werden demgegenüber Semantic Web-Technologien bereits erfolgreich eingesetzt, um die Bedeutungsebene dieses seriellen Quellentyps zu erfassen und der Forschung bereitzustellen. Vgl. [vogeler_warum_2015; tomasek_medea_2016]

⁷Die *unpublizierten* Beispiele stammen aus dem wissenschaftlichen Nachlass von Dr. Franz Dumont, ehemals Mitarbeiter der Soemmerring-Forschungsstelle an der Akademie der Wissenschaften und der Literatur, Mainz.

⁸Samuel Thomas von Soemmerring an Georges de Cuvier. München, 8. März 1816. Museum d'Histoire Naturelle Paris, Ms 629 V N° 11.

Beispiel 2:

„Aber erlauben Sie mir zugleich die dringende Bitte um gütige Mittheilung der versprochenen Anmerkungen zu meiner Schrift vom Nervensaft zu wiederholen.“⁹

Beispiel 3:

„Ich erinnere mich daß Haas die ganze Haut prächtig roth ausgespritzt hat, das läßt doch Gefäße in der Cuticula vermuthen.“¹⁰

Beispiel 4:

„[...] Überbringer dieses Briefs ist der merkwürdige Hermaphrodit Dürge, über den ich vor einigen Tagen gelesen habe. [...]“¹¹

Beispiel 5:

„Hier Mein Theuerster! ein geringes Zeichen meines Andenkens, welches den noch schuldigen Dank für den zweyten Band Ihrer Vermischten Schriften begleiten sollte.“¹²

Allein anhand dieser wenigen Beispiele kann man demonstrieren, wie Erwähnungen genauer spezifiziert und typisiert werden könnten:

- Personen:
 - deren Empfehlung gegenüber Dritten
 - Kontakt
- Schriften:
 - Rezeption (gelesen, zustimmend oder kritisierend)
 - Bitten um Meinungen dazu
 - deren Übersendung

Die Liste ließe sich natürlich fortsetzen. Eine umfassendere Kategorisierung würde aber den Rahmen des vorliegenden Beitrags, der vor allem die Grundzüge des Semantischen Kommentierens skizzieren möchte, bei weitem sprengen.

⁹Samuel Thomas von Soemmerring an Johann Heinrich Ferdinand Autenrieth. München, 18. April 1814. Württembergische Landesbibliothek, Stuttgart, o.S.

¹⁰Alexander von Humboldt an Samuel Thomas Soemmerring. Bayreuth, 8. Mai 1796, hg. v. Ulrich Päßler unter Mitarbeit von Klaus Gerlach und Ingo Schwarz. In: edition humboldt digital, hg. v. Ottmar Ette. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin. Version 2 vom 14.09.2017. URL: <https://edition-humboldt.de/v2/H0008511>

¹¹Friedrich Tiedemann an Samuel Thomas Soemmerring. Landshut, 29. Juli 1806. Akademie der Wissenschaften und der Literatur Mainz, ehem. Soemmerring-Forschungsstelle, ohne Signatur.

¹²Samuel Thomas von Soemmerring an Gottfried Reinhold Treviranus. München, 24. Oktober 1817. Universitätsbibliothek Bremen, Nachlass Treviranus, TS 136, XXIV.

Manche Erwähnungen könnten schon in einer TEI-XML-Auszeichnung weiter qualifiziert werden, etwa nach dem Muster:

```
<persName key="P000009" role="#recommended">GoIdfuss</persName>
```

Die Kodierung in TEI-XML erscheint unter Aspekten der Langzeitarchivierung sinnvoll. Betrachtet man aber die Qualifizierung der Erwähnungen genauer, sieht man, dass es sich dabei eigentlich um die Beziehungen zwischen Entitäten handelt: Person A empfiehlt Person B. Person C kritisiert Werk D. Dass Soemmerring August Goldfuß empfiehlt, geht aus der oben genannten Kodierung nur implizit hervor und funktioniert nur in Kombination mit den Metadaten des Briefes. Das ist zum einen fehleranfällig, zum anderen können manche Aussagen auf diese Weise gar nicht kodiert werden, z. B. wenn Kritik von Dritten an Dritten wiedergegeben wird. Das legt es nahe, eine Technologie zu wählen, die nicht das Dokument im Fokus hat, sondern eben Beziehungen zwischen Entitäten: Graphentechnologien.

Dabei stehen prinzipiell zwei Wege offen: zum einen Graphdatenbanken im engeren Sinne und zum anderen das „Semantic Web“ auf Basis von RDF. Graphdatenbanken mit ihren „property graphs“, also Graphen, bei denen Knoten und Kanten Eigenschaften aufweisen können, wären sicherlich keine ungeeignete Wahl. Im Rahmen eines einzelnen Projekts wären sie sogar vielleicht die technisch und konzeptionell leichter einsetzbare Wahl. Berücksichtigt man allerdings die grundsätzliche Unabgeschlossenheit und Offenheit einer Briefedition, scheint es wünschenswert, die Daten auch möglichst offen breitzustellen. So „fehlen“ zum einen in einer Briefedition (ob Gesamt- oder Briefwechselfassung) per se immer diejenigen Briefe, die nicht von oder an die Person(en) geschickt wurden, denen die Edition gewidmet ist. Dabei sind Briefe nicht nur Teile eines Briefwechsels, sondern Teil eines größeren Korrespondenznetzes (Bunzel 2014, 236–37).

Zum anderen werden in Briefen häufig viele verschiedene Personen, Schriften, Ereignisse etc. erwähnt, die aus ganz unterschiedlichen und auch „editionsernen“ Perspektiven von Interesse sind (Biermann und Lange 1962, 226; Mücke 2001, 89–90). Betrachtet man das obige Beispiel, erschließt sich das Problem rasch: Die Wissenschaftsgeschichte ist ja nicht nur daran interessiert, dass Soemmerring „Dürre“, also Maria Dorothea Derrier¹³, getroffen hatte, sondern möchte auch die umgekehrte Frage stellen: Welche Personen hat denn Derrier wiederum getroffen? Dadurch wird es nötig, dass die Daten möglichst gut projektübergreifend abfragbar bereitstehen. Was aber bei der Masse an Briefen nur dezentral geschehen kann. Hier hat der Einsatz von RDF im Sinne von Linked Open Data einen klaren Vorteil gegenüber Graphdatenbanken. Hinzu kommt, dass nur Semantic-Web-Technologien „inferencing services“ bieten. Diese stecken zwar in der konkreten Anwendung in vielen Bereichen noch in den Kinderschuhen, versprechen aber für die Zukunft ein großes Potential. Schlussendlich kann man natürlich auch beide Technologien kombinieren: Graphdatenbanken für die Analyse im eigenen Projektrahmen und RDF als projektübergreifendes Bereitstellungsformat. Der Aufsatz interessiert sich aber vor allem für den letzteren Aspekt, weshalb er auch in der Folge nur auf den Einsatz von Semantic-Web-Technologien fokussiert.

¹³Carl Dürregé, auch Maria Dorothea Derrier (1780–1835), ein damals sog. „Hermaphrodit“, d.h. ein intersexueller Mensch.

2 Semantisch Kommentieren

2.1 Grundlage: Ontologien

Der Einsatz von Semantic-Web-Technologien macht es nötig, die oben angedeuteten Qualifizierungen in richtige Vokabulare (oder auch: Ontologien¹⁴) umzusetzen. Dabei werden Klassen (classes) und Eigenschaften (properties, d. h. Relationen) sowie deren Verhältnis zueinander spezifiziert. Das kann im Semantic Web mit Hilfe von Ontologiesprachen geschehen. Im vorliegenden Fall wird die Web Ontology Language (OWL) verwendet, die besonders ausdrucksstark ist und gerade im wissenschaftlichen Bereich viel Anwendung findet. Bisher fehlt es allerdings an einer Ontologie zu wissenschaftsgeschichtlich relevanten Handlungen. Auch mit diesem Beitrag, der ein Konzept darlegen möchte, kann und soll kein abschließendes Vokabular vorgelegt werden. Anhand der oben aufgeführten, kurzen Quellenzitate sollen lediglich beispielhaft einige Klassen und Eigenschaften definiert werden, um das Konzept zu demonstrieren. Anhand der Beispiele könnte man zu folgenden Definitionen kommen:

<i>Klasse</i>	<i>Definition</i>
hos:Actor	Handelnde Person oder Körperschaft
hos:Lecture	Vorlesung
hos:Opinion	Meinung

<i>Eigenschaft</i>	<i>Definition</i>
hos:askFor	fragt nach
hos:recommends	empfiehlt [Person]
hos:gives	hält [Vorlesung]
hos:about	[Vorlesung Meinung] über [Schrift Person Thema]
hos:reads	[Person] liest [Publikation Brief]

Die Definitionsbeispiele sind den aufgeführten, *wissenschaftshistorisch* relevanten Briefbeispielen entnommen. In Briefen, die für andere Disziplinen relevant sind (Literaturgeschichte, Musikwissenschaft etc.), mögen andere oder weitere Relationen zu finden sein. Einen Teil wird man dagegen auch domänenübergreifend definieren können, so etwa korrespondenz-spezifische Handlungen wie das Überbringen oder Weiterleiten von Briefen. Auch dazu liegt noch keine Ontologie vor. Ebenfalls nur zu Demonstrationszwecken sollen anhand der Quellenbeispiele folgende Definitionen vorgenommen werden:

<i>Klasse</i>	<i>Definition</i>
cs:Letter	Brief als physisches Objekt
cs:Communication	Kommunikationsvorgang
cs:Actor	Person oder Körperschaft im Kommunikationsvorgang
cs:Person	Person
cs:Body	Körperschaft

¹⁴Die Begriffe „Ontologie“ und „Vokabular“ werden hier im Sinne des W3C synonym verwendet. Siehe <https://www.w3.org/standards/semanticweb/ontology>.

<i>Eigenschaft</i>	<i>Definition</i>
cs:isObjectIn	[cs:Letter] ist Objekt in [cs:Communication]
cs:hasSender	[cs:Communication] hat als Absender [cs:Actor]
cs:hasReceiver	[cs:Communication] hat als Empfänger [cs:Actor]
cs:hasCarrier	[cs:Communication] hat als Überbringer [cs:Actor]

Bestimmte Klassen und Eigenschaften wiederum wird man gar nicht erst definieren müssen, da es für manche Informationstypen schon ausreichende Vokabulare gibt. So könnten z. B. Publikationen und deren Eigenschaften und Beziehungen über Ontologien wie Bibframe oder DCMI Metadata Terms repräsentiert werden.

In einer so erstellten Ontologie könnten und sollten darüber hinaus Klassen und Beziehungen des spezifischen, eigenen Vokabulars auf allgemeinere Vokabulare abgebildet und dadurch für ganz andere Fragestellungen nachgenutzt werden – ohne dass dies in jeder einzelnen Instanz notiert werden muss. So könnte z. B. die Klasse `cs:Actor` als Subklasse von „E 39 Actor“ im Erlangen CRM¹⁵ definiert werden. Dadurch wird im konkreten Fall die Person oder Organisation nicht nur bei Abfragen nach deren spezifischer Rolle als Absender, Empfänger oder Überbringer gefunden, sondern auch als einfach Handelnder, der sich z. B. gerade an einem bestimmten Ort aufhält.

2.2 Aussagen

Die oben aufgeführten Beispiele könnten also auf Basis der vorgeschlagenen Modellierung und Ontologie von der Editorin zu folgenden semantischen Kommentaren führen¹⁶:

Beispiel 1:

„Erlauben Sie mir Ihnen Herrn Professor Doct. Goldfuss aus Erlangen, einen äusserst braven, bescheidenen, sehr unterrichteten Mann zu einer freundlichen Aufnahme bestens zu empfehlen.“

<<http://d-nb.info/gnd/118805193>> hos:recommends <<http://d-nb.info/gnd/100151310>>

Beispiel 2:

„Aber erlauben Sie mir zugleich die dringende Bitte um gütige Mittheilung der versprochenen Anmerkungen zu meiner Schrift vom Nervensaft zu wiederholen.“

¹⁵<http://erlangen-crm.org/>

¹⁶In den Beispielen werden die Namensräume durch entsprechende Kürzel abgekürzt: „soe“ für eine fiktive digitale Soemmerring-Edition; „hos“ und „cs“ für die oben eingeführten Klassen und Eigenschaften. Alle anderen Kürzel entstammen existierenden Vokabularen, z.B. „bf“ für Bibframe.

```
<http://d-nb.info/gnd/118805193> hos:askFor soe:419c9b34 .
soe:419c9b34 rdf:type hos:Opinion .
soe:419c9b34 hos:about soe:WN00058 .
soe:WN00058 rdf:type bf:Work .
soe:WN00058 bf:title "Über den Saft, welcher aus den Nerven wieder
eingesaugt wird, im gesunden und kranken Zustande des menschlichen
Körpers".
```

Beispiel 3:

„Ich erinnere mich daß Haas die ganze Haut prächtig roth ausgespritzt hat, das läßt doch Gefäße in der Cuticula vermuthen.“

```
<http://edition-humboldt.de/H0012069/> hos:reads <http://edition-humboldt/SJ7RKVCJ/> .
```

Beispiel 4 (Teil):

„[...] Hermaphrodit Dürrge, über den ich [Tiedemann] vor einigen Tagen gelesen habe.
[...]"

```
<http://d-nb.info/gnd/118867733> rdfs:label "Friedrich Tiedemann"
<http://d-nb.info/gnd/118867733> hos:gives soe:e76d08f0 .
soe:e76d08f0 rdf:type hos:Lecture .
soe:e76d08f0 hos:about <http://www.wikidata.org/entity/Q53730446> .
<http://www.wikidata.org/entity/Q53730446> rdfs:label "Maria Dorothea Derrier" .
```

Die Beispiele 4 und 5 beschreiben zudem die Übersendung eines Briefes bzw. einer Publikation. Dies wird im nächsten Abschnitt aufgegriffen.

2.3 Kontexte

Die so projektübergreifend qualifizierten Erwähnungen sind aber natürlich nicht allgemeingültig bzw. unabhängig von einem Kontext. Auch wenn sie in eine maschinenlesbare Form überführt worden sind, ist es immer noch äußerst wichtig zu wissen, dass sie einer historischen Quelle entnommen sind. Im Gegensatz zur (direkten) Kodierung in TEI-XML stehen diese semantischen Aussagen erst einmal allein. Daher muss die als Triple formulierte Aussage mit Metadaten über diese Quelle verbunden werden, die ebenfalls als RDF vorliegen – das kann natürlich auf Basis der TEI-Metadaten geschehen.

Auch für die Modellierung von Korrespondenzen in RDF gibt es noch kein allgemein anerkanntes Modell. In Nutzung der oben aufgeführten, beispielhaften Definitionen könnten die korrespondenz-spezifischen Metadaten für das Beispiel 4 (inkl. der Nennung des Überbringers) wie folgt notiert werden:

```
soe:L00001 rdf:type cs:Letter .
soe:L00001 cs:isObjectIn soe:C01205 .
soe:C01205 rdf:type cs:Communication .
soe:C01205 cs:hasSender <http://d-nb.info/gnd/118867733> .
soe:C01205 cs:hasReceiver <http://d-nb.info/gnd/118805193> .
soe:C01205 cs:hasCarrier <http://www.wikidata.org/entity/Q53730446> .
```

Anzumerken ist, dass in der hier vorgeschlagenen Modellierung zwischen dem (evtl. gar nicht mehr vorhandenen) physischen Objekt Brief und dem Kommunikationsvorgang getrennt wird. Dies liegt u. a. darin begründet, dass Briefe weitergeleitet, d. h. mehrmals versendet werden können. Ohne den „Kommunikationsvorgang“ (`cs:Communication`) wäre es in einem solchen Fall nicht mehr möglich zu spezifizieren, in welcher Reihenfolge Absender, ggf. Überbringer und Empfänger zu sortieren wären. Darüber hinaus ermöglicht es die Trennung von physischen Objekt und Kommunikationsvorgang, das Versenden von Gegenständen und Publikationen im semantischen Kommentar wiederzugeben. Das Beispiel 5 würde dann analog zum oben genannten notiert.

Bei der Modellierung von Korrespondenzmetadaten könnte man sich auch stärker an das Konzept von `<correspDesc/>` aus den Richtlinien der Text Encoding Initiative anlehnen. Dann würde unterhalb von `cs:Communication` noch eine zweite Klassen-Ebene (`cs:CommunicationAct`) erscheinen, die – analog zum TEI-Element `<correspAction/>` – Personen, Orte und Datumsangaben nach Senden, Empfangen, Überbringen etc. gruppieren.

Neben dem Kontext des Kommunikationsvorgangs ist es auch notwendig, den wissenschaftlichen Kontext notieren zu können, d. h. wer diesen maschinenlesbaren Kommentar überhaupt bereitgestellt hat. Beides muss nun mit der notierten Aussage verbunden werden. Dafür kann im Semantic Web das Konzept der „named graphs“ verwendet werden. Zur konkreten Modellierung der Verbindung von Kommentar, Metadaten der Quelle und Metadaten über die Annotation an sich kann darüber hinaus auf ein standardisiertes und etabliertes Datenmodell zurückgegriffen werden: Web Annotation Data Model (WADM, ehemals „Open Annotation Data Model“) (Sanderson, Ciccarese und Young 2017). In diesem Datenmodell gliedert sich die Annotation – stark vereinfacht gesprochen – in zwei Teile: den `body`, also den Kommentar an sich und das `target`, also das, was kommentiert wird.

Der `body` einer solchen Annotation kann dabei nicht nur ein in reinem Text oder HTML abgelegter Kommentar sein, sondern wiederum ein Graph. Das ist mit Hilfe des in der SPARQL-Spezifikation (Harris, Seaborne und Prud'hommeaux 2013) niedergelegten Konzepts der „named graphs“ möglich, wodurch Aussagen als ein eigener, benannter Graph gespeichert und bei der Abfrage adressiert werden können.

Das `target` ist in unserem Fall mindestens ein Brief. Wünschenswert ist natürlich, die Textstelle so genau wie möglich adressieren zu können. Das WADM bietet hierfür zwar verschiedene Möglichkeiten, die allerdings alle voraussetzen, dass der edierte Text dann unverändert bleibt. Abgesehen vom `body` und

target kann die Annotation durch weitere Angaben ergänzt werden, z. B. wer zu welcher Zeit die Annotation angelegt hat.

Zusammen genommen entsteht nun die gewünschte Verbindung zwischen Kommentar, Quellen- und Annotationsmetadaten. Das oben aufgeführte Beispiel 3 würde dann folgendermaßen aussehen:

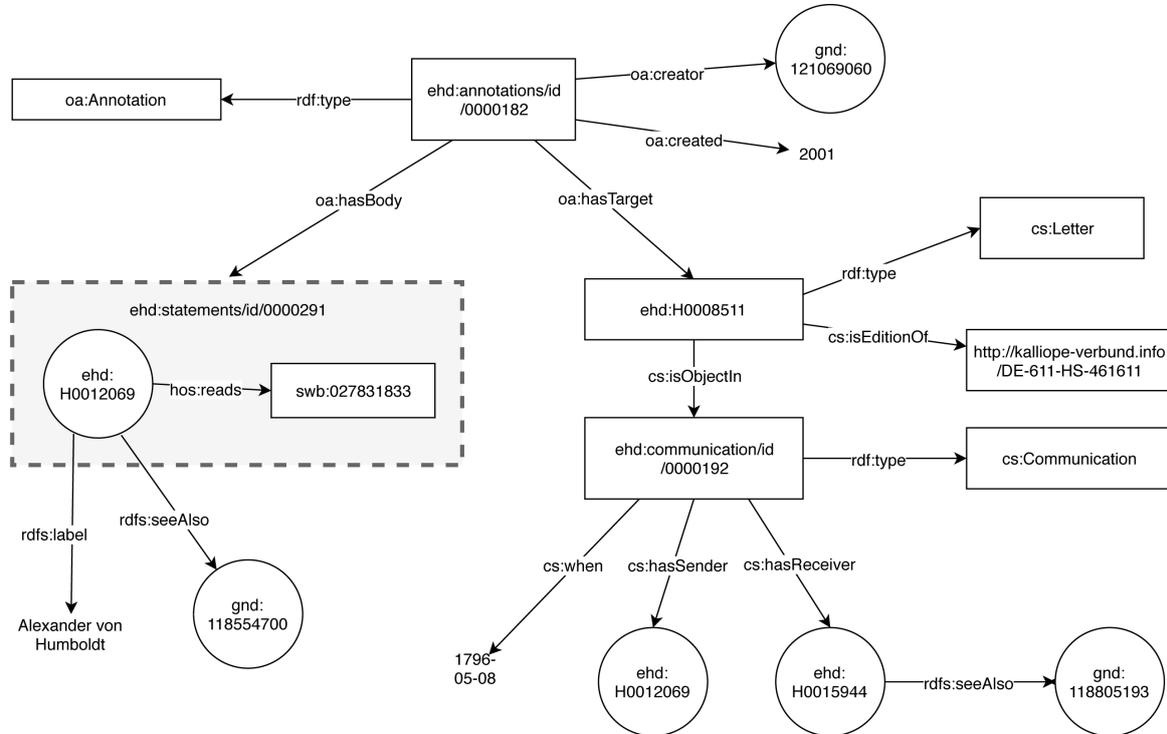


Abbildung 1: Semantischer Kommentar mit Kontext

In der Abbildung 1 ist links in der Mitte ist die Aussage „A. v. Humboldt liest ‚De vasis cutis...‘“ mit Hilfe von Triples festgehalten. Sie ist als eigener Graph im Triplestore gespeichert, der über die URI `ehd:statements/id/000029` adressierbar ist. Er ist body der dargestellten Annotation. Das Ziel der Annotation ist der edierte Brief „Alexander von Humboldt an Samuel Thomas Soemmerring. Bayreuth, 8. Mai 1796“, der ebenfalls über eine URI adressierbar ist. Die Korrespondenzmetadaten sind entsprechend dem oben vorgeschlagenen Modell notiert. Zugleich ist noch festgehalten, dass der Brief eine Edition einer Archivalie darstellt, die über eine Kalliope-URI identifiziert wird. Die Annotation insgesamt ist mit Metadaten (Ersteller und Erstellungsdatum) versehen.

2.4 Eingabewerkzeuge

Voraussetzung für dieses Kommentierungsverfahren ist, dass entsprechende Werkzeuge zur Eingabe zur Verfügung stehen. Tatsächlich existiert schon ein solches Tool: pundit Annotator Pro¹⁷, das vor einigen Jahren von der italienischen Firma Net7 entwickelt wurde. Bei pundit Annotator Pro handelt

¹⁷<http://thepund.it/annotator-pro-web-annotation/>

es sich um eine Client-Software, mit deren Hilfe semantische Annotationen in Triples notiert werden können. Naturgemäß funktioniert das nur mit einer entsprechenden Datenbank auf dem Server, in der die Annotationen mitsamt Metadaten gespeichert werden. Diese wurde von Net7 ebenfalls zur Verfügung gestellt. Das Software-Paket wurde im Rahmen von „Burckhardtsource“ entwickelt. Pundit böte einen guten Startpunkt für ein Werkzeug, um eine Briefedition mit semantischen Kommentaren anzureichern.

Möglich wäre es aber auch, diese Kommentare im Rahmen einer TEI-XML-basierten Arbeitsumgebung vorzunehmen. Dies könnte beispielsweise mit Hilfe der TEI-Elemente `<listRelation/>` und `<relation/>`¹⁸ in der jeweiligen Quelle geschehen. Ein anderer Ansatzpunkt, Annotationen entsprechend dem WADM auch in der TEI zu ermöglichen, wird gerade im TEI-Council diskutiert.¹⁹ So könnte eine TEI-XML-basierte Editionsarbeitsumgebung wie z. B. ediarum²⁰ mit überschaubarem Aufwand erweitert werden. Die so in TEI-XML kodierten „Triples“ müssten nach der Eingabe nur noch mit Skripten nach RDF konvertiert werden, wofür schon generische Werkzeuge bereitstehen.²¹

Nicht zu vergessen ist, dass durch die Editorin nur die Aussagen eingegeben werden müssten (also in Abbildung 1 der named graph, d. h. der graue Kasten). Die Kontexte, d. h. Annotations- und Quellenmetadaten, würden vom System bereitgestellt bzw. aus den TEI-Header-Daten extrahiert werden. Dadurch reduziert sich die von der Editorin zu leistende Eingabe pro Textstelle nur auf das Wesentlichste.²²

Vor diesem Hintergrund ist die Frage zu diskutieren, ob das vorgeschlagene Kommentierungsverfahren zu Mehraufwand bei der täglichen Editionsarbeit führt. Dies wird wohl in geringem Maß der Fall sein. Allerdings ist die Identifizierung, Auszeichnung und Verknüpfung von Namen und Literaturtiteln heutzutage immer noch Standard in digitalen Briefeditionen. Eine Qualifizierung der Erwähnung wird erstmal nur unwesentlich mehr Zeit kosten, vorausgesetzt, dass aus einer bereits definierten Menge an Typen ein Typ ausgewählt werden soll. Allerdings kann es sein, dass mehr Namen als bisher ausgezeichnet werden. Reichte es bislang vielleicht aus, in einem Brief nur die erste Erwähnung auszuzeichnen, muss die Bearbeiterin nun darauf achten, ob bei mehreren Erwähnungen nicht unterschiedliche Aussagen zu diesem Namen getroffen werden. Wenn das der Fall ist, sollten diese Erwähnungen natürlich auch ausgezeichnet werden, weil es eben nicht mehr nur darum geht, die Person etc. zu identifizieren und für das Register zu erschließen. Das wird aber mutmaßlich nur in geringem Maß mehr Zeit kosten. Der Hauptaufwand bei der Kommentierung fließt auch heute noch in die Recherche zur Identifizierung von erwähnten Personen, Orten etc.

¹⁸<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-relation.html>

¹⁹Siehe <https://github.com/TEIC/TEI/issues/1745>

²⁰<http://www.bbaw.de/telota/software/ediarum>

²¹So z. B. der Webservice „XTriples“, <http://xtriples.spatialhumanities.de>, siehe Grüntgens und Schrade (2016).

²²Insofern ist auch Kamzelak (2016) zu widersprechen, dass dies für die Editoren zu kompliziert wäre.

3 Diskussion

3.1 Kommentar oder doch Regest?

Vielleicht wird man sich an dieser Stelle – wie andere²³ – fragen, ob man das wirklich noch „Kommentar“ nennen kann oder die vorgeschlagene Art der „Annotation“ nicht doch eher dem Erstellen eines Regests gleich kommt. Die Ähnlichkeiten sind dabei offensichtlich: Zum einen werden eigentlich keine zusätzlichen Informationen ergänzt, sondern lediglich die Informationen der Quelle expliziert. Zum anderen werden diese Aussagen aus der Quelle abstrahiert und kategorisiert („empfiehlt“, „sendet an“ etc.) dargestellt. Das ist richtig und zutreffend. Gleichwohl gibt es zwei markante Unterschiede zum Regest: So erhebt das Regest den Anspruch, zumindest alle wichtigen Informationen wiedergeben zu wollen. Wie im nächsten Abschnitt diskutiert, wird man aber mit der vorgestellten Methode niemals alle Aussagen aus den Quellen wiedergeben können, auch nicht alle (vermeintlich) wichtigen. Darüber hinaus ist das semantische Kommentieren nicht als Ersatz für eine Transkription gedacht, sondern soll sie nur begleiten und auf eine bestimmte Art analysierbar machen. Und wie bei allem quellenkritischen Arbeiten soll die Daten-Rezipientin stets die Möglichkeit haben, die erstellte Kommentierung anhand der Quelle zu überprüfen. Schlussendlich knüpft die Methode auch an die Identifizierung von erwähnten Personen, Orten, Schriften etc. an und qualifiziert diese nur weiter. Sie gehört also zumindest im Ursprung dann doch zum Kommentar. Die Gleichstellung mit dem klassischen Briefregest erscheint daher problematisch und könnte auch Missverständnisse befördern.²⁴ Daher erscheint es dem Verfasser sinnvoller, weiterhin – trotz gewisser Ähnlichkeiten mit dem Regest – vom „Kommentar“ zu sprechen.

3.2 Grenzen

Wie jeder Einsatz von (nicht nur digitalen) Methoden hat auch diese ihren Nutzen, aber auch ihre Grenzen. So ist die notwendige Erstellung einer entsprechenden Ontologie für das Projekt bzw. den fachlichen Bereich aufwendig. Soll das volle Potential des Semantic Web genutzt werden, muss diese Ontologie wiederum auf allgemeinere Vokabulare abgebildet werden, was ebenfalls Zeit braucht und auch problematisch sein kann (Dröge 2010). Schließlich werden nicht nur Vokabulare für die Beziehungen zwischen und die Klassen von Entitäten benötigt. Auch die Entitäten selbst müssen adressierbar sein, und zwar am besten projektübergreifend im Sinne von Linked Open Data. Das mag für Personen und Orte mit VIAF, GND, GeoNames & Co. noch relativ unproblematisch sein. Schon für Schriften und Werke sieht die Situation dagegen noch nicht so gut aus. Insbesondere fehlt häufig die Möglichkeit, das „Werk“ (und nicht die Ausgabe oder gar das konkrete Exemplar) zu adressieren. Zwar kann ersatzweise auf Einträge in Katalogen zurückgegriffen werden (siehe unten), aber oft hat die Editorin keine Möglichkeit, anhand des Briefes eine Ausgabe zu identifizieren, sondern kann nur konstatieren, dass über eine nicht näher bestimmbare Entstehungsstufe eines Werks gesprochen wird. Bisher werden Technologien des Semantic Web vor allem in Museen und Bibliotheken eingesetzt und werden vor allem auf realweltliche

²³Der Dank des Verfassers für diesen interessanten Einwand gilt Gerrit Brüning.

²⁴Für die Regesten mittelalterlicher Urkunden, die insbesondere Rechtsakte wiedergeben, stellt sich das Problem vielleicht weniger: Hier könnte mit Hilfe entsprechender Ontologien Herrscherhandeln ausreichend umfassend qualifiziert werden. Vgl. auch Kuczera (2017), 185f.

Objekte angewendet, z. B. auf Bücher oder Münzen. Die in Briefen erwähnten Werke sind oft schwer zu definieren, die diskutierten Objekte längst verloren und besprochene Themen zu schwammig. Hier wartet also noch einiges an Arbeit.

Nicht alle Sachverhalte können mit Hilfe der vorgeschlagenen Methode kommentiert werden. Vieles wird man auch weiterhin als Text annotieren müssen, weil die Sachverhalte zu komplex sind, als dass sie sinnvoll in (auch mehreren) Triples abbildbar wären. Doch sollten auch diese Einzelstellenerläuterungen mit Hilfe des Web Annotation Data Model gespeichert und bereitgestellt werden (Lordick u. a. 2016). In diesem Fall besteht der Annotationsbody nicht aus einem oder mehreren Triples, sondern aus reinem oder bestenfalls mit leichtem Markup versehenen Text. So können alle Einzelstellenkommentare – ob semantische oder normale – auf die gleiche Art und Weise ausgetauscht werden. Das ist insbesondere von Vorteil, wenn ein Brief nicht nur in einer einzigen Edition vorliegt, sondern in zwei (oder gar mehreren). Das ist z. B. der Fall, wenn beiden Korrespondenten jeweils eine eigene Edition (vielleicht gar Gesamtausgabe) gewidmet ist. In solchen Fällen kann ein Brief folglich aus mehreren Perspektiven kommentiert vorliegen. Dabei kommt es normalerweise zu Überschneidungen, aber auch zu sich ergänzenden Kommentaren, die in der jeweils anderen Edition fehlen. Möglicherweise kommt es sogar zu sich widersprechenden Kommentaren. Sind diese Sacherläuterungen im Web Annotation Data Model notiert und stehen über die entsprechende API bereit, kann man diese Annotationen projektübergreifend nachnutzen und in der jeweils anderen Edition einblenden (Dumont 2018b).

3.3 Datenanalyse – auch jenseits der Edition

Die mit der vorgeschlagenen Methode erstellten Annotationen müssen allerdings auch abgefragt werden können. Dazu muss eine Nutzerin zum einen die Abfragesprache SPARQL beherrschen, zum anderen die benutzten bzw. aufeinander abgebildeten Vokabularien kennen. Einfach zu bedienende Interfaces zum Explorieren oder Abfragen existieren bisher noch nicht. Für die Formulierung von Abfragen ist daher im Moment noch eine Einarbeitung in die Abfragesprache SPARQL und die in den Daten verwendeten Vokabulare notwendig. Die Frage ist aber, ob dies wirklich ein Hindernis darstellt, diese Technologie einzusetzen. Auch für die Realisierung von digitalen Editionen oder zur historischen Netzwerkanalyse sind entsprechende technische Kenntnisse notwendig, die über Studiengänge und Schulungen erworben werden können. Es ist also zu hinterfragen, warum die Analyse von RDF-Daten mit SPARQL überhaupt völlig intuitiv durchführbar sein muss. Zudem ist das Konzept von RDF durchaus überschaubar und die Abfragesprache SPARQL leichter zugänglich als etwa das Pendant SQL. Schlussendlich können Recherche-Interfaces i. d. R. auch nur für Interaktionen erstellt werden, die man antizipiert. Für völlig neue Fragen ist das nicht möglich.

Der große Nutzen des semantischen Kommentierens liegt darin, dass inhaltliche Aussagen über das Handeln und Interagieren von Akteuren im großen Maßstab analysierbar werden. Dabei werden Annotations- und Quellenkontext stets festgehalten. Beides ist aus Gründen der Wissenschaftlichkeit im Allgemeinen und der historischen Quellenkritik im Besonderen notwendig. So sind die semantischen Annotationen ja erstmal keine per se gültigen Fakten, sondern Aussagen, die in einer bestimmten Quelle gefunden wurden. Im Rahmen der Quellenkritik müssten sie nun geprüft und gewichtet werden. Mit der vorgeschlagenen Modellierung ist das allerdings auch weiterhin möglich. Man könnte sogar eigens festhalten,

dass Aussagen anderen vorliegenden Informationen widersprechen, weil die festgehaltenen Aussagen als „named graphs“ ja selbst adressierbar und somit wiederum Ziel einer Annotation werden können.

Darüber hinaus liegt gerade in Briefeditionen der Reiz darin, dass Aussagen aus den Briefen zusammen mit ihrem Kontext – dem Kommunikationsvorgang – abfragbar werden und es möglich wird zu analysieren, wer wem gegenüber welche Aussagen trifft. Denn Briefe sind auf einen (oder auch mehrere) bestimmten Absender hin geschrieben und nur unter dieser Voraussetzung zu verstehen (Woesler 1977, 47–48; Mücke 2001, 98; Gerlach 2009, 22). Durch die vorgestellte Methode könnten Briefe in größerer Menge unter dem Aspekt Adressatenbezogenheit besser verstanden werden. So könnte man beispielsweise (ab-)fragen, *welchem Adressaten gegenüber* eine Person empfohlen oder auch *nicht* empfohlen wird. Das ist insbesondere bei Akteuren interessant, die eine *organisierende* Funktion im jeweiligen Feld (Wissenschaft, Musikbetrieb etc.) wahrnahmen.

Vorteil ist darüber hinaus, dass semantische Annotationen, als Linked Open Data bereitgestellt und mit projektübergreifenden Identifizierern versehen, eine Abfrage über Projektgrenzen hinaus ermöglichen. Theoretisch könnten so nicht nur Aussagen aus Briefen, sondern auch aus Tagebüchern und anderen Quellen mit in eine Analyse einbezogen werden.²⁵ Auch können überlieferte Bücher, Objekte aus Museumssammlungen (Wettlaufer 2018) oder Daten aus Professorenkatalogen automatisch verknüpft werden. Dies soll folgendes Beispiel veranschaulichen:

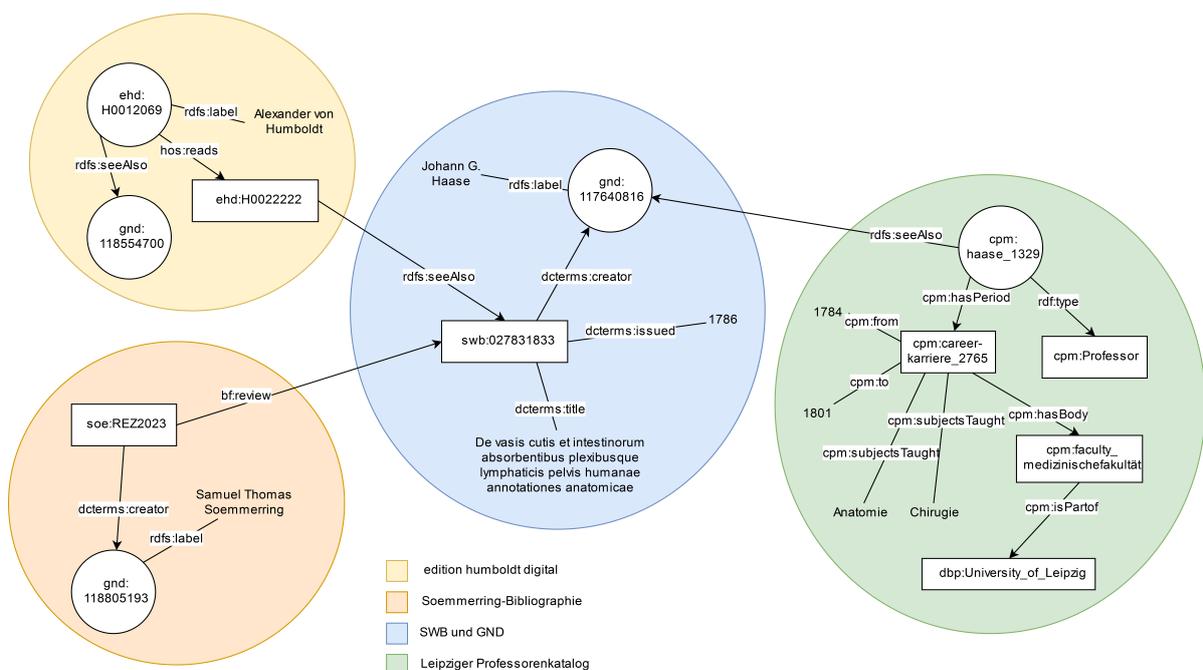


Abbildung 2: Semantischer Kommentar aus einer Edition als Teil eines größeren Linked Open Data-Graphen

²⁵Voraussetzung dafür ist natürlich, dass die Daten unter einer freien Lizenz bereitgestellt werden. Dies sollte allerdings bei einer digitalen wissenschaftlichen Publikation heute mittlerweile zum Standard gehören. Davon zu trennen ist die notwendige Deklaration der wissenschaftlichen Herausgeber- und Autorschaft der Daten, der ja (nicht nur durch Verwendung des WADM) problemlos nachgekommen werden kann.

Abbildung 2 zeigt den semantischen Kommentar aus Beispiel 3 (in der Abbildung oben links)²⁶ als Teil des globalen Linked Open Data-Graphen. Dabei ist die hier dargestellte Vernetzung nur ein kleiner Ausschnitt des tatsächlich möglichen. Die Aussage ist über das von Humboldt gelesene Buch „De vasis cutis ...“ mit dessen Katalogdaten im Südwestdeutschen Bibliothekverbundes (SWB) vernetzt. Die dort notierte GND ist wiederum mit den Daten zu Haase im Professorenkatalog der Universität Leipzig²⁷ (catalogus professorum lipsiensium) verknüpft. Schlussendlich sind bibliographische Daten aus der Soemmerring-Forschung aufbereitet und ebenfalls mit dem Katalogdatensatz verknüpft. Anhand dieses – noch überschaubaren Beispiels – kann man sehen, was „Linked Open Data“ bedeutet. Stehen die Daten als korrektes RDF und mit projektübergreifenden URIs versehen (GNDs u. a., im Beispiel meistens mit `rdfs:seeAlso` angegeben) bereit, müssen keine Vorbereitungen getroffen werden, um diese zu vernetzen. Sie sind vernetzt, sobald man die Daten in einen Triplestore lädt oder die SPARQL-Schnittstellen gemeinsam abfragt. Diese vernetzten Daten kann man nun aus ganz verschiedenen Perspektiven befragen: Eine Humboldt-Forscherin etwa könnte mit der Fragestellung an die Daten herangehen, welche Literatur Alexander von Humboldt wann und zu welchem Thema gelesen und in seinen Briefen diskutiert hat; eine andere Wissenschaftlerin könnte untersuchen, wie medizinische Werke um 1800 gelesen, rezensiert und diskutiert worden sind; eine dritte Wissenschaftlerin, die sich mit der Geschichte der Universität Leipzig befasst, könnte fragen, wie groß der Einfluss der medizinischen Fakultät Ende des 18. und Anfang des 19. Jahrhunderts gewesen ist. Dies alles auf Grundlage derselben Daten.

Dabei können die SPARQL-Abfragen unterschiedlich komplex werden. Für letztere Frage sähe die SPARQL-Abfrage zum Beispiel so aus:

```

1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX cpm: <http://catalogus-professorum.org/cpm/2/>
4 PREFIX cpmb: <http://catalogus-professorum.org/lipsiensium/body/>
5 PREFIX dcterms: <http://purl.org/dc/terms/>
6 PREFIX dbp: <http://dbpedia.org/resource/>
7 PREFIX hos: <http://history-of-science.eu/vocab/dev/0.1>
8
9 SELECT DISTINCT ?reader
10 WHERE {
11     ?prof rdf:type cpm:Professor ;
12         cpm:hasPeriod ?careerPeriod .
13     ?careerPeriod cpm:periodBody cpmb:faculty-medizinischefakultaet .
14     cpmb:faculty-medizinischefakultaet cpm:isPartOf dbp:University_of_Leipzig .
15     ?prof rdfs:seeAlso ?gnd .
16     ?gnd ^dcterms:creator ?publication .
17     ?publication ^rdfs:seeAlso ?id .
18     ?id ^hos:reads ?reader .
19 }

```

Abbildung 3: Projektübergreifende SPARQL-Abfrage, wer alles Publikationen Leipziger Medizin-Professoren gelesen hat

²⁶Hier zur besseren Übersicht allerdings ohne die dazugehörigen Annotations- und Quellmetadaten.

²⁷<https://research.uni-leipzig.de/catalogus-professorum-lipsiensium>

Die gezeigte Abfrage kann anhand des in Abbildung 2 gezeigten Datenmodells Schritt für Schritt nachvollzogen werden. Technisch gesehen könnte man die Abfrage reduzieren, je nachdem wie die vorhandenen Ontologien Identifizierungskonzepte wie `owl:sameAs` oder `skos:narrowMatch` verwenden, um URIs direkt zu verknüpfen.²⁸ Auch könnte die abfragende Wissenschaftlerin eine eigene Ontologie für wiederholende Teile der Abfrage erstellen. Schließlich steht der Forscherin noch offen, aus den von den unterschiedlichen Projekten erstellten Daten neue, reduzierte Daten zu konstruieren, um zum Beispiel eine Netzwerkanalyse vorzunehmen.

Grundsätzlich anzumerken ist, dass das in Abbildung 2 und 3 gezeigte Beispiel lediglich eine Eigenschaft (property) aus dem für diesen Beitrag erstellten Vokabular verwendet, nämlich `hos:reads`. Alle weiteren Klassen und Eigenschaften entstammen anderen Vokabularen. So werden die Katalogdaten vom SWB u. a. mit dem Vokabular DCMI Metadata Terms (`dcterms`) angeboten, der Leipziger Professorenkatalog hat ein eigenes Vokabular (CPM) entwickelt, das entsprechend online dokumentiert ist (Riechert 2010; Morgenstern u. a. 2018). Der kleine Datensatz aus einer (fiktiven) digitalen Soemmerring-Bibliographie wurde ebenfalls mit Standardvokabular realisiert, nämlich `dcterms` und `Bibframe (bf)`. Das zeigt auch, dass zu diesem Zeitpunkt nicht alles mit einer eigenen neuen Ontologie modelliert werden muss, sondern vielfach auf bewährte Vokabulare zurückgegriffen werden kann.

Darüber hinaus ist es sinnvoll, die eigene Ontologie, die man zur Erstellung des semantischen Kommentars verwendet hat, auf passende, allgemeinere Vokabulare abzubilden. Das hilft Wissenschaftlerinnen dabei, die Daten nicht aus der ursprünglich notierten, projekt- oder domänenspezifischen Sicht abzufragen, sondern aus einer übergreifenden Perspektive. Zum Beispiel könnte eine einfache Frage sein, wer Martha Dorothea Derrier (Beispiel 4) überhaupt gekannt hat. Dafür stünde beispielsweise die Relation `knows` im bekannten FOAF-Vokabular zur Verfügung. Wird die oben definierte kommunikationsspezifische Beziehungskette `cs:hasSender` – `cs:Communication` – `cs:hasCarrier` mit Hilfe von `owl:propertyChainAxiom` entsprechend auf `foaf:knows` gemappt, wird diese (verkettete) Beziehung auch bei einer Abfrage anhand der allgemeinen Beziehung `foaf:knows` gefunden. Dazu muss der verwendete Triplestore lediglich OWL2-RL-Reasoning unterstützen.

4 Fazit

Die skizzierte Methode des semantischen Kommentierens hat den großen Vorteil, dass sie in den Briefen enthaltene inhaltliche Aussagen maschinell lesbar und verstehbar machen würde – weit über das bisher übliche Maß hinaus. Insbesondere ist das Potential hervorzuheben, dass die in Editionen aufwendig erarbeiteten Daten als Linked Open Data zur Verfügung stehen und zusammen mit Daten aus anderen Quellen befragt werden können. Im Hinblick auf Briefeditionen schließt das semantische Kommentieren damit an die Bemühungen der letzten Jahre an, edierte Briefe anhand ihrer Metadaten stärker zu vernetzen.

Im Hinblick auf digitale Editionen im Allgemeinen würde die Methode helfen, den Blick stärker wieder auf die inhaltlichen Aspekte zu lenken. In der letzten Dekade widmete sich nämlich die Editionswissenschaft

²⁸Zum Potential und Problematik von `owl:sameAs` und Alternativen vgl. Halpin und Hayes (2010) und Wettlaufer u. a. (2015), 195–96.

eher der Wiedergabe der Manuskriptseite. Die diplomatische Transkription wurde also nicht nur – zu Recht – in der digitalen Edition fortgesetzt, sondern es wurde versucht, das Textbild umfassend, also über die Textstrukturen hinaus, zu erfassen und in der HTML-Darstellung wiederzugeben (Hanrahan 2015). Dabei wird häufig übersehen, dass durch die Beigabe von digitalen Faksimiles die Transkription sich auf die grundlegenden und logischen Textstrukturen²⁹ beschränken und die Editorin sich ansonsten der inhaltlichen Erschließung widmen könnte. Wie oben gezeigt, spielt das digitale Medium mit seiner Maschinenlesbarkeit doch eigentlich gerade hier seine Stärke aus – nutzen wir sie!

²⁹Deren überwiegender Teil z.B. auch durch die Korpuslinguistik maschinell in großen Mengen ausgewertet werden kann – im Gegensatz zu einer gleichsam mimetischen, fast schon „hyperdiplomatischen“ Erfassung des Textbildes.

Literaturverzeichnis

- Biermann, Kurt-Reinhard und Fritz G. Lange. 1962. Die Alexander-von-Humboldt-Briefausgabe. *Forschungen und Fortschritte* 36, Nr. 8: 225–230.
- Bunzel, Wolfgang. 2014. Briefe, Briefnetze, Briefnetzwerke. Überlegungen zur epistolaren Interkonnektivität. In: *Fontanes Briefe ediert*, 232–245. Fontaneana 12. Würzburg: Königshausen & Neumann.
- Dröge, Evelyn. 2010. Leitfaden für das Verbinden von Ontologien. *Information in Wissenschaft und Praxis* 61: 143–147.
- Dumont, Stefan. 2016. correspSearch – connecting scholarly editions of letters. *Journal of the Text Encoding Initiative*, Nr. Issue 10 (7. Dezember). doi:[10.4000/jtei.1742](https://doi.org/10.4000/jtei.1742), (zugegriffen: 5. März 2018).
- Dumont, Stefan. 2018a. Interfaces in digital scholarly editions of letters. In: *Digital scholarly editions as interfaces*, hg. von Roman Bleier, Martina Bürgermeister, Helmut Klug, Frederike Neuber, und Gerlinde Schneider, 109–131. Schriften des instituts für dokumentologie und editorik 12. Norderstedt: Books on Demand. <http://nbn-resolving.de/urn:nbn:de:hbz:38-91146> (zugegriffen: 16. Mai 2018).
- Dumont, Stefan. 2018b. Ein Brief – zwei Perspektiven. Stellenkommentare in digitalen Briefeditionen über APIs austauschen. Poster. Köln: Poster, 2. März. doi:[10.5281/zenodo.1248081](https://doi.org/10.5281/zenodo.1248081), (zugegriffen: 16. Mai 2018).
- Gerlach, Klaus. 2009. Zur Textkritik von Handschriften. Ihre Notwendigkeit und hermeneutische Dimension, dargestellt am Beispiel des Briefwechsels zwischen Böttiger und Duvau. In: *Was ist Textkritik?*, hg. von Winfried Woesler, 19–28. Beihefte zu editio 28. Niemeyer.
- Giannaccini, Costanza. 2017. Burckhardtsource.org: Where scholarly edition and semantic digital library meet. In: *Advances in digital scholarly editing*, hg. von Peter Boot, Anna Cappellotto, Wout Dillen, Franz Fischer, Adohán Kelly, Andreas Mertgens, Anna-Maria Sichani, Elena Spadini, und Dirk Van Hulle, 293–298. Leiden: Sidestone Press.
- Grüntgens, Max und Torsten Schrade. 2016. Data repositories in the Humanities and the Semantic Web: modelling, linking, visualising. In: *WHiSe 2016 Humanities in the Semantic Web. Proceedings of the 1st Workshop on Humanities in the Semantic Web (WHiSe)*, hg. von Alessandro Adamou, Enrico Daga, und Leif Isaksen, 53–64. CEUR Workshop Proceedings 1608. Aachen. <http://ceur-ws.org/Vol-1608/#paper-07> (zugegriffen: 16. September 2017).
- Hagen, Waltraud. 1988. Von den Erläuterungen. In: *Vom Umgang mit Editionen. Eine Einführung in Verfahrenweisen und Methoden der Textologie*, hg. von Siegfried Scheibe, Waltraud Hagen, Christel Laufer, und Uta Motschmann, 202–224. Berlin: Akademie Verlag.
- Halpin, Harry und Patrick J. Hayes. 2010. When owl:sameAs isn't the Same: An Analysis of Identity Links on the Semantic Web. In: *LDOW-2010. Linked Data on the Web 2010*, hg. von Christian Bizer, Tom Heath, Tim Berners-Lee, und Michael Hausenblas. CEUR Workshop Proceedings 628. http://ceur-ws.org/Vol-628/ldow2010_paper09.pdf.
- Hanrahan, Elise. 2015. „Over-tagging“ with XML in digital scholarly editions. In: *DHd2015. Von daten zu erkenntnissen. Book of abstracts*. 193–196. Graz, 27. Februar.

- Harris, Steve, Andy Seaborne und Eric Prud'hommeaux, Hrsg. 2013. SPARQL 1.1 query language. W3C, 21. März. <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/> (zugegriffen: 8. Juni 2017).
- Kamzelak, Roland S. 2016. Digitale Editionen im semantic web. Chancen und Grenzen von Normdaten, FRBR und RDF. In: „*Ei, dem alten Herrn zoll' ich Achtung gern*“. *Festschrift für Joachim Veit zum 60. Geburtstag*, hg. von Peter Stadler und Kristina Richts, 423–435. München: Allitera.
- Kruse, Joseph A. 2001. Heines Briefe. Literarische Qualität und historische-biographische Quelle. In: *Briefkultur im Vormärz*, 165–177. Vormärz-Studien 9. Bielefeld: Aisthesis.
- Kuczera, Andreas. 2017. Graphentechnologien in den Digitalen Geisteswissenschaften. *ABI Technik* 37, Nr. 3: 179–196. doi:[10.1515/abitech-2017-0042](https://doi.org/10.1515/abitech-2017-0042), (zugegriffen: 21. September 2017).
- Lordick, Harald, Rainer Becker, Michael Bender, Luise Borek, Canan Hastik, Thomas Kollatz, Beata Mache, Andrea Rapp, Ruth Reiche und Niels-Oliver Walkowski. 2016. Digitale Annotationen in der geisteswissenschaftlichen Praxis. *Bibliothek Forschung und Praxis* 40, Nr. 2: 186–199. doi:[10.1515/bfp-2016-0042](https://doi.org/10.1515/bfp-2016-0042), (zugegriffen: 25. September 2017).
- Morgenstern, Ulf, Christian Rau, Christian Augustin, Frederike Bulka, Lydia Mueller, Thomas Riechert, Simeon Ackermann, u. a. 2018. Catalogus Professorum Model - Version 2.1 (CPM). 26. April. <https://catalogus-professorum.org/cpm/2/index-en.html> (zugegriffen: 31. Mai 2018).
- Mücke, Marion. 2001. Briefeditionen als Quellenwerke für die medizinhistorische Forschung. *editio* 15: 86–121.
- Müller, Susanne. 2015. Semantic annotations protocol. In: *EUROCORR project documentation. Burckhardt source*, hg. von Maurizio Ghelardi. Pisa. <http://wiki.burckhardtsource.org/protocols/semantic-annotation-protocol/>.
- Riechert, Thomas. 2010. Catalogus Professorum Model: Ein Vokabular zur Beschreibung von geschichtswissenschaftlichen Professorenkatalogen für das Semantic Web. In: *Catalogus Professorum Lipsiensis*, hg. von Ulf Morgenstern und Thomas Riechert, 33–48. Leipziger Beiträge zur Informatik 21. Leipzig.
- Sanderson, Robert, Paolo Ciccarese und Benjamin Young. 2017. Web annotation data model. W3C recommendation. W3C, 23. Februar. <https://www.w3.org/TR/annotation-model/> (zugegriffen: 25. September 2017).
- Stadler, Peter. 2012. Normdateien in der Edition. *editio* 26: 174–183.
- Stadler, Peter. 2014. Interoperabilität von digitalen Briefeditionen. In: *Fontanes Briefe ediert*, hg. von Hanna Delf von Wolzhagen, 278–287. Fontaneana 12. Würzburg: Königshausen & Neumann.
- Stockmar, René. 2005. *Private Briefe - freie Wissenschaft. Briefe edieren am Beispiel von Friedrich Nietzsches Briefwechsel 1872-1874*. EditionTEXT 7. Frankfurt am Main / Basel: Stroemfeld.
- Strobel, Jochen. 2013. Der Brief als Prozess. Entwurf und Konzept in der digitalen Edition. In: *Brief-Edition im digitalen Zeitalter*, hg. von Anne Bohnenkamp, 133–146. Beihefte zu editio 34. Berlin/Boston: Walter de Gruyter.

Wettlaufer, Jörg. 2018. Der nächste Schritt? Semantic Web und digitale Editionen. In: *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*, hg. von Roland S. Kamzelak und Timo Steyer. Sonderband der Zeitschrift für digitale Geisteswissenschaften 2. 15. März. http://dx.doi.org/10.17175/sb002_007.

Wettlaufer, Jörg, Christopher Johnson, Martin Scholz, Mark Fichtner und Sree Ganesh Thotempudi. 2015. Semantic blumenbach: Exploration of text–object relationships with semantic web technology in the history of science. *Digital Scholarship in the Humanities* 30, Nr. suppl 1 (1. Dezember): i187–i198. http://dsh.oxfordjournals.org/content/30/suppl_1/i187 (zugegriffen: 17. Januar 2017).

Woesler, Winfried. 1977. Der Brief als Dokument. In: *Probleme der Brief-Edition*, hg. von Wolfgang Frühwald, Hans-Joachim Mähl, und Walter Müller-Seidel, 41–59. Kommission für Germanistische Forschung. Mitteilung II. Boppard: Harald Boldt.

Woesler, Winfried. 1993. Zu den Aufgaben des heutigen Kommentars. *editio* 7: 18–35. <https://www.degruyter.com/view/j/edit.1993.7.issue-1/9783110241983.18/9783110241983.18.xml?rskey=Bwb2iA&result=7&q=brief> (zugegriffen: 17. Mai 2018).

Woesler, Winfried. 1998. Richtlinienvorschläge für Briefkommentare. In: *Wissenschaftliche Briefeditionen und ihre Probleme*, hg. von Hans-Gert Roloff, 87–96. Berliner Beiträge zur Editionswissenschaft 2. Berlin: Weidler.