

Roland S. Kamzelak

Digitale Editionen im *semantic web*

Chancen und Grenzen von Normdaten, FRBR und RDF

Alle Editionen sind digital; das ist inzwischen eine Binsenweisheit. Die Erarbeitung und Verarbeitung der Daten, die für eine Edition gesammelt und hinzugefügt werden, als Transkription, als Kommentar und als Metadaten auch technischer Art, geschieht digital, mit Computern im weitesten Sinne. Das Produkt oder die Produkte sind zunächst digitale Daten, später bedruckte Seiten oder eine digitale Publikation auf E-Readern oder im *World Wide Web* (W3). Ebenso eine Binsenweisheit ist es, dass Editionsdaten sehr hochwertige Daten sind. So scheint es im Zeitalter von „Mashups, der intelligenten Wiederverwendung von Webressourcen“¹ sinnvoll zu sein, die digitalen Daten von Editionsprojekten für solche Mashups zur Verfügung zu stellen.

Als Beispiel soll – nicht ganz zufällig – die Edition der Tagebücher von Harry Graf Kessler dienen.² Es sind seit 1994 neun von acht Bänden erschienen. Die Arbeit daran dauerte insgesamt 16 Jahre. Bei der Edition sind hochqualifizierte Daten gesammelt worden, die als erläuterte Register in jedem Band zu finden sind. Dabei sind relevante Mengen zusammengelassen, bedenkt man, dass alle Einträge gut recherchiert und mit IDs der *Gemeinsamen Normdatei* (GND) angereichert sind. Über 4.000 Werke der europäischen Literatur, über 4.000 Körperschaften zumeist aus Europa, aber auch aus den USA, über 5.000 Orte auf der ganzen Welt, teilweise mit historischen Hintergrunddaten und über 12.000 Namen der wichtigsten Kulturschaffenden Europas um die Jahrhundertwende mit Verflechtungen in die ganze Welt.³ Auch die Weltreise Kesslers hat die Register erweitert.⁴ Diese Arbeit wurde zumeist mit öffentlichen Geldern im Umfang von mehr als 2,3 Millionen Euro finanziert. Zwar ist die Ausgabe erfolgreich

¹ Siehe Brigitte Endres-Niggemeyer (Hg.), *Semantic Mashups. Intelligent Reuse of Web Resources*, New York 2013.

² Harry Graf Kessler, *Das Tagebuch 1880–1937* (9 Bde., 2004–2010), hg. von Roland S. Kamzelak und Ulrich Ott, Stuttgart 2004–2010.

³ Für eine ausführliche Darlegung der materiellen Seite dieser Edition siehe Roland S. Kamzelak, *Materi-
alwirtschaft*, in: *editio* 23 (2009), S. 159–168.

⁴ Siehe Roland S. Kamzelak (Hg.), *Harry Graf Kesslers Weltreisealbum*. Mit einem Essay von Ulrich Pohlmann, edition.eliber.de 2013.

und verkauft sich im Vergleich mit anderen, qualitativ nicht im Mindesten dieser Edition nachstehenden Ausgaben sehr gut. Dennoch ist es angesichts der Förder-summe, der Arbeitszeit und vor allem der weiteren Möglichkeiten, diese Daten in weiteren Zusammenhängen zu nutzen, absolut geboten, weiterzudenken, wie diese Daten tatsächlich nachgenutzt werden können.

Bei qualitativ hochwertigen Daten zur Kultur der Jahrhundertwende um 1900, die digital vorliegen, denkt man an das *semantic web*. Also einer automatischen Verknüpfung semantischer Entitäten im W3. Das *World Wide Web Consortium* (W3C) definiert das *semantic web* kurz so:

The term „Semantic Web“ refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.⁵

Das hier zugrundeliegende Verständnis von „Semantik“ dürfte sich allerdings nicht mit dem eines Geisteswissenschaftlers decken, der eher eine Definition wie die folgende erwarten dürfte:

Semantik (Semasiologie, vom griech. *sema*, „Zeichen“; Signifikant, vom lat. *signum*, „Zeichen“), Teildisziplin der Semiotik, deren Forschungsgegenstand die Beziehung zwischen dem Zeichen und dem durch dieses bedeutetem oder bezeichnetem Objekt ist. Allgemein ist die S. auch die Wissenschaft von den Bedeutungen sprachlicher Zeichen, die Theorie der Bedeutungen.⁶

Was in der Linguistik eher als die Bedeutung von Zeichen und von Verbindungen von Zeichen (Wörtern, Sätzen, ganzen Büchern oder Bibliotheken) verstanden wird, versteht die Technik als Verknüpfung von „Gleichem“ und konzentriert sich auf die Verknüpfungsregeln (RDF, OWL),⁷ um dieses Gleiche, das an verschiedenen Orten abgelegt ist, über eine Abfragesprache (SPARQL)⁸ sichtbar zu machen. Und hier beginnen die Schwierigkeiten, denn ein einfacher Satz wie „Ich hab dich lieb bis zum Mond“, sagte der kleine Hase und machte die Augen zu.“⁹ ist an sich unverständlich

⁵ World Wide Web Consortium (W3C), *Semantic Web*, online abrufbar unter <http://www.w3.org/standards/semanticweb/> [Stand: 30. Nov. 2015].

⁶ Art. *Semantik* in: Heinrich Schmidt, *Philosophisches Wörterbuch*, neu bearb. von Georgi Schischkoff, Stuttgart 1982, S. 631.

⁷ RDF = Reference Description Framework. Siehe <https://www.w3.org/RDF/> [Stand: 30. Nov. 2015].

OWL = Web Ontology Language. Siehe <http://www.w3.org/TR/owl2-overview/> [Stand: 30. Nov. 2015].

⁸ SPARQL = SPARQL Protocol And RDF Query Language.

⁹ Aus dem Kinderbilderbuch: Sam McBratney, *Weißt Du eigentlich, wie lieb ich dich hab?*, Frankfurt 2015.

bzw. im Wortsinne unverständlich. Bei „bis zum Mond“ fehlt der Vergleichspunkt. Nimmt man den Standpunkt des Sprechers (den eine Maschine ja nicht hat), so könnte sich eine Distanz ergeben, die man messen kann und als Vergleich zum „Liebhaben“ nehmen könnte, doch misst man „Liebhaben“ ja kaum mit Distanzen. Auch der Halbsatz „und machte die Augen zu“ ist mehrdeutig. Entweder der Hase schläft einfach ein oder er entschläft. Dem Wort „einschlafen“ ist das nicht abzulesen, wohl aber dem Kontext des ganzen Buches, wo es ja ums Einschlafen, oder das nicht Einschlafen können/wollen, um das zu Bett gehen, geht. Dieses metaphorische Sprechen ist selbst für geübte Leser und Sprecher eine Herausforderung. In diesem Falle noch mehr durch die Multimodalität und Mehrsprachigkeit, da die Geschichte nicht nur aus Text, sondern auch aus Illustrationen¹⁰ besteht, und der Text eine Übersetzung aus dem Englischen ist.¹¹ Bedeutung ist also kontextabhängig und mehrdeutig.

Dass dies kein konstruiertes Beispiel ist, beweist ein Blog-Beitrag von Lucy Wickham:

[...] I guess I just do not understand why it is that specific distance. The moon is, on average, 238.855 miles away. So technically people are saying they love you 477.710 miles (the distance to the moon and back). You think that's a lot, but in terms of a car, it's not really. A car gets an average of 12.000 miles a year. Saying someone loves you 477.710 miles is like saying someone loves you for 40 years. That is not a very strong guarantee. [...]¹²

Was Wickham versucht, ist eine Explikation dieser Metapher. Linguistisch gesehen stellt sie Implikaturen her, um das Unverständliche in Verständliches aufzulösen.¹³ Sie expliziert das Metaphorische. Grice gibt als Grundmaxime aus, dass Verständigung auf Kooperation beruht. Diese Grundbedingung muss man auch unterstellen, wenn man die Bemühungen um ein *semantic web* beurteilt. Doch die vier Maximen, die dann folgen, sind schwer zu greifen: 1. Qualität zielt darauf ab, nur wahre Dinge zu sagen, Dinge, für die man Beweise besitzt, 2. Quantität spricht von der Menge der Informationsvergabe, die weder zu ausladend, noch zu verkürzt sein darf, 3. Relation und 4. Art und Weise: Hier soll man Mehrdeutigkeit vermeiden oder obskure Ausdrücke. Am schwersten zu greifen ist allerdings die dritte Maxime, die erläutert wird mit

¹⁰ Illustrationen von Anita Jeram.

¹¹ Übersetzung von Rolf Inhauser. Im Englischen ist „I love you to the moon and back“ inzwischen eine sehr populäre Redeweise geworden.

¹² Lucy Wickham, *The Dilemma with the Phrase*, „I Love You to the Moon and Back“, in: *Thought Catalog*, online abrufbar unter <http://thoughtcatalog.com/lucy-wickham/2014/08/the-dilemma-with-the-phrase-i-love-you-to-the-moon-and-back/> [Stand: 30. Nov. 2015].

¹³ Siehe H. Paul Grice, *Logic and Conversation*, in: *Syntax and Semantics*, hg. von P. Cole und J. Morgan, New York 1975, Bd. 3: *Speech Acts*, S. 41–58.

dem Satz „Be relevant.“ Diese dritte Maxime erinnert aber am meisten an die Sprache des W3C, wo es eben um Verknüpfungen von Daten geht, um Relationen, und zwar um relevante, um semantisch zugehörige Daten.

Da Maschinen noch keine Implikaturen herstellen können, müssen die Relationen eindeutig sein. Das gelingt am zielsichersten durch die Verwendung von Normdaten. Ist eine Person durch eine GND-Nummer¹⁴ identifiziert, lassen sich leicht Verknüpfungen zu anderen Angeboten herstellen, die diese Nummer ebenfalls führen. Das sehr beliebte und stetig qualitativ verbesserte Wikipedia¹⁵ führt diese GND-Nummer zum Beispiel mit. Es wäre also möglich – und wird auch praktiziert – statt selbst Personenerläuterungen zu recherchieren und zu schreiben, über die GND-Nummer auf Angebote wie die Deutsche Biographie¹⁶ oder den Wikipedia-Eintrag zu verlinken oder diesen Link alternativ zur eigenen Erläuterung anzubieten. Das wird von der *academic community* aber nur angenommen, wenn die Verbindungen auch eindeutig sind und nicht durch falsche Verknüpfungen auf andere Personen zeigen, als die, die man gemeint hat. Trotz sorgfältiger Redaktion an der Deutschen Nationalbibliothek (DNB) gibt es doch oft uneindeutige Individualisierungen. Das rührt daher, dass die Personen durch Croudsourcing in der Bibliotheksgemeinde über die Katalogverbünde in die DNB eingespeist werden. Zweck der Eingabe ist durchaus die eindeutige Identifizierung eines Autors, um ihn mit einem Werktitel zu verknüpfen, doch bleibt dafür im Zusammenhang mit der Erschließung wenig Zeit. So ist es doch oft einfacher, einen neuen GND-Satz anzulegen.¹⁷ Nimmt man hier noch in Betracht, dass es international mehrere ernst zu nehmende Normdatenpools zu Personen gibt, verwässert sich der Wert des Normdatums zunehmend, da sich das Problem der Dubletten potenziert. Antworten auf dieses Problem hat das Projekt VIAF¹⁸ gesucht und teilweise gefunden. VIAF hat einen Algorithmus entwickelt, der Normdaten verschiedener Normdatenpools abgleicht und da, wo es durch die Beschreibungen möglich ist, zusammenführt. So kann VIAF z. B. sechs Ansetzungen von Harry Graf Kessler auf einen Datensatz reduzieren. Das gelingt aber leider nicht zuverlässig.¹⁹ Hier besteht also Handlungs-

¹⁴ Siehe die Informationsseite der Deutschen Nationalbibliothek, die die GND pflegt: http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html [Stand: 30. Nov. 2015].

¹⁵ *Wikipedia. Die freie Enzyklopädie*, URL: <https://www.wikipedia.de/> [Stand: 30. Nov. 2015].

¹⁶ <http://www.deutsche-biographie.de/> [Stand: 30. Nov. 2015]. Die Deutsche Biographie hat gegenüber der Wikipedia den Nachteil, dass die Artikel ‚veralten‘, während die Wikipedia ‚lebt‘ und stetig verbessert werden kann. Dafür lässt sich die Deutsche Biographie eindeutiger referenzieren, weil sie stabiler ist.

¹⁷ Es gibt 29.675 Müller, 80.282 Heinrich, 1.947 Veit in der GND. Bei solchen Mengen ist selbst die kaskadierende Suche noch sehr aufwändig.

¹⁸ *Virtual International Authority File (VIAF)*, URL: <https://viaf.org/> [Stand: 30. Nov. 2015].

¹⁹ Sucht man in VIAF im Feld „Namen“ nach „Veit“, erhält man 3.057 Treffer. Bei „Veit, Joachim“ schränkt sich die Auswahl ein auf 86 Treffer. Der gesuchte Joachim Veit mit der GND-Nummer 103988327 ist jedoch nicht darunter. Die Suche erfolgte am 29. September 2015.

bedarf. Dennoch: Die GND-Nummer ist die verlässlichste Identifikation, die wir im deutschsprachigen Raum haben und sollte unbedingt zum Pflichtfeld aller Metadaten gehören.

Gleiches oder Ähnliches gäbe es über Normdaten für Körperschaften, Orte, Werke, Periodika, Sachen und Themen zu sagen. Bei diesen Gruppen ist die Identifizierung in der Regel noch viel schwieriger zu vereindeutigen. Bei Orten verwendet man zunehmend Georeferenzen oder Geoinformationssysteme (GIS) und kann auf diese Weise zuverlässig einen Punkt auf einer vermassten Karte anzeigen, aber keine Flächen und keine historische Dimension eines Ortes. Die historische Dimension muss über andere Quellen, als textliche Metadaten, hinzugespielt werden.²⁰ Dies in aller Breite auszuführen würde den Rahmen jedoch sprengen.²¹ Hier sollen die Werke noch betrachtet werden, die in Editionen von Lebensdokumenten eine besondere Rolle spielen. Eine besondere Rolle deshalb, weil in Briefen und Tagebüchern zwar über publizierte Werke, aber auch über Publikationsprojekte gesprochen wird. Projekte haben einen besonderen Charakter, denn sie werden im Idealfall einfach zu publizierten Werken, können aber auch unvollendet bleiben oder es können mehrere Werke aus diesen Projekten resultieren.²² Im ersten Fall ist die Wahrscheinlichkeit groß, ein Normdatum zu finden, im zweiten Falle in der Regel nicht. Wie geht man im *semantic web* mit solchen Projekten um? Wenn Thomas Mann über seinen *Joseph* schreibt, dann ist überhaupt nicht klar, um welchen Romanteil es geht, und oft ist ein Motiv gemeint, das in mehreren Bänden vorkommt. Dieses Motiv jedoch als Thema oder Schlagwort zu begreifen, würde zu kurz greifen, da es ja wiederum mit dem Werk verbunden ist. Gibt man also alle Werke an, die in Frage kommen, dann wird die Verknüpfung inflationär und führt zu einer Verwässerung der Suchergebnisse.

Aber selbst bezüglich der publizierten Werke stellt sich oft die Frage, was referenziert werden soll. So sprechen Schriftsteller – so wie wir auch – meist nur vom Faust. Wir gehen davon aus, dass Goethes *Faust I* gemeint ist,²³ was aber durchaus nicht gesichert ist. Es gibt *Faust I* und *II*, den *Urfaust* und die Figur Faust. Es gibt einen *Faust* von Christopher Marlowe, dann den *Dr. Faustus* von Thomas Mann usw. Wir müssen also, wenn wir Faust referenzieren, Verweisungen setzen zu all diesen Möglichkeiten. Hilft das oder wird auch hieraus eine Flut von Verknüpfungen, die verwässert?

²⁰ Eine wichtige Quelle hierfür ist *Geonames*, das einige historische Informationen mitführt. Siehe <http://www.geonames.org/> [Stand: 30. Nov. 2015].

²¹ Hier sollte es einen interdisziplinären Workshop mit Geographen, Historikern und vielleicht Ethnologen gemeinsam mit Fachleuten aus der Informatik und den Digital Humanities geben.

²² So wird aus Thomas Manns *Joseph* eine Tetralogie. Siehe z. B. Thomas Mann, *Joseph und seine Brüder* (3 Bde.), Frankfurt am Main 1988.

²³ Im Grunde auch nur eine Grice'sche Implikatur, um die Dunkelheit der Sprache, wie Adorno sagt, aufzulösen.

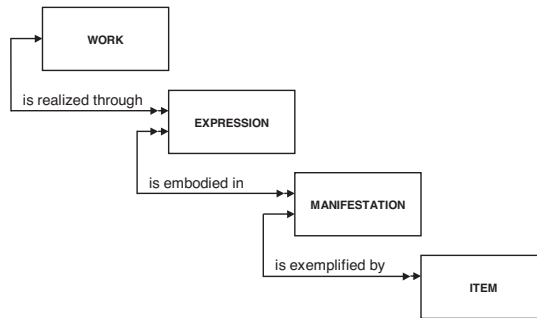


Abbildung 1: FRBR-Modell, *Group 1 Entities and Primary Relationships*²⁴

Eine Antwort auf diesen Problemkomplex versucht die *International Federation of Library Associations and Institutions* (IFLA) mit dem System FRBR²⁵ zu geben (vgl. Abb. 1). Die Bibliotheken haben erkannt, dass die Verbindungen zwischen Entitäten für jeden Nutzer eines Bibliothekskatalogs sichtbar sein sollten. Dabei ist *item* ein spezifisches Exemplar einer Medieneinheit, das in einer Bibliothek zu finden ist. Dieses Exemplar ist unikal, aber eine Bibliothek kann mehrere Exemplare eines Werkes besitzen. So zum Beispiel Ausgaben verschiedener Verlage oder verschiedene Auflagen, also Exemplare (*items*) verschiedener *manifestations* des Werkes. Das Werk selbst ist dabei als Werkidee (*work*) zu verstehen, die einen spezifischen Ausdruck findet (*expression*), zum Beispiel als Roman oder als Komposition. Oder aber, um im Beispiel zu bleiben, wäre Thomas Manns Projekt *Joseph*, von dem er in Briefen spricht, *work*, der erste Teil des Romans *Die Geschichten Jaakobs* *expression* und die Fischerausgabe davon die *manifestation*. Das scheint logisch zu sein und abzudecken, was in Lebensdokumenten an Referenzpunkten vorkommt. Wie sieht es also mit dem zweiten Beispiel – *Faust* – aus? Nehme ich den Goetheschen *Faust* als *work*, so sind *Urfaust*, *Faust I* und *Faust II* *expressions*? Oder nehme ich den Stoff *Faust* als *work* und habe dann als *expressions* Goethes *Urfaust*, *Faust I* und *Faust II*, aber auch Marlowes *Dr. Faustus* und vielleicht sogar Thomas Manns *Dr. Faustus* als *expressions*? Wie sieht es mit Übersetzungen aus? Sind sie *expressions* neben den Originaltexten? Das würde dann ja bedeuten, dass sie direkt aus der Werkidee stammen, was sie faktisch nicht tun. Und wenn man nun die Originalmanu- oder typoskripte hinzunimmt, die ja vor dem 20. Jahrhundert teilweise wesentlich von den gesetzten Texten abweichen, dann

²⁴ Abbildung entnommen aus <http://archive.ifla.org/VII/s13/frbr/frbr1.htm#3> [Stand: 30. Nov. 2015].

²⁵ *Functional Requirements for Bibliographic Records* (FRBR). Siehe <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records> [Stand: 30. Nov. 2015].

eindeutig dem Werk zuordnen. Um es positiv zu wenden, kann man vielleicht sagen, dass die Einteilung von Werken nach *work*, *expression*, *manifestation* und *item* in die richtige Richtung weist und hilft, ein Bewusstsein für die Komplexität des Werkbegriffs zu entwickeln. Die Beschreibung mit Pfeildiagrammen nach dem Schema Subjekt, Prädikat und Objekt eröffnet vielfältige Möglichkeiten. Dass es sich bei diesem Schema um RDF handelt, ist unschwer zu erkennen. Bibliothekskataloge künftig mit RDF aufzubauen, ist ein großer Schritt, um digitale Editionen mit der Bibliotheks- und Archivwelt zu verschmelzen.²⁷

Wünschenswert wäre ein einfaches System, das wie eine Genealogie funktioniert. Von jeder Person gibt man ausschliesslich Vater und Mutter ein. Daraus allein erschließt sich die gesamte Familiengeschichte mit allen Seitenlinien: Nachzeitig sind Mutter und Vater, Großmütter und Großväter, Urgroßeltern usw., vorzeitig sind Töchter und Söhne, Enkelinnen und Enkel, Urenkel usw., zur Seite Cousins, Cousinen 2. Grades, 3. Grades usw., oder zur Seite vorzeitig finden sich Nichten und Neffen 2. Grades (also die Kinder von Cousins) usw. Nur die Begriffe gehen aus, nicht die Eindeutigkeit der Beziehung zur Ausgangsperson.

Kehrt man zu dem Beispiel Faust zurück, wäre ein Graph wie in Abb. 3 dargestellt vorstellbar. Dies wäre das Ergebnis einer Suche mit SPARQL, der RDF-Abfragesprache.²⁸ RDF muss man formal beschreiben, wobei man verschiedene Notationen verwenden kann: Turtle, N-Triples, N-Quads, JSON-LD, Notation3 oder RDF/XML. Der Datensatz zu Harry Graf Kessler aus der Deutschen Biographie ist im Listing 1 wiedergegeben. Auf diese Weise formal beschrieben kann ein Projekt wie RelFinder,²⁹ das die Daten aus der Wikipedia in Form der DBPedia³⁰ abfragt, diese in Graphen umsetzen. Sucht man nach Goethe und Faust, erhält man den Graphen aus Abb. 4. Weder *Faust I* noch *Faust II* kommen in dem Graphen vor. Auch die Verbindung zu Ovid wird nicht dargestellt. Sucht man nach Goethe und Ovid, ergibt sich der Graph aus Abb. 5. Zunächst ist der Graph verwirrend und enthält leider auch falsche Angaben, wie zum Beispiel, dass Goethe Shakespeare beeinflusst habe. Das ist aber sicherlich nur eine Frage der Justierung. Interessant ist an dem Graphen, dass Namen zum Vorschein kommen, an die man bei der Eingabe der Stichwörter vielleicht nicht gedacht hat. Gibt man nun ein drittes Stichwort ein, ergeben sich neue Fragestellungen (vgl. Abb. 6).

²⁷ Die bisherige Katalogisierungsrichtlinie RAK (*Regeln für die alphabetische Katalogisierung*) soll abgelöst werden durch das auf FRBR und RDF aufbauende RDA (*Resource Description and Access*). Siehe <http://www.rda-jsc.org/archivedsite/rda.html> [Stand: 30. Nov. 2015]. Ziel ist es dabei auch, die Katalogisierung bereits im Verlag zu beginnen, indem dort die Werke bereits in RDA angelegt werden.

²⁸ Siehe <http://www.w3.org/TR/rdf-sparql-query/> [Stand: 30. Nov. 2015].

²⁹ *RelFinder: Interactive Relationship Discovery in RDF Data*, URL: <http://www.visualdataweb.org/relfinder.php> [Stand: 30. Nov. 2015].

³⁰ Siehe <http://wiki.dbpedia.org/> [Stand: 30. Nov. 2015].

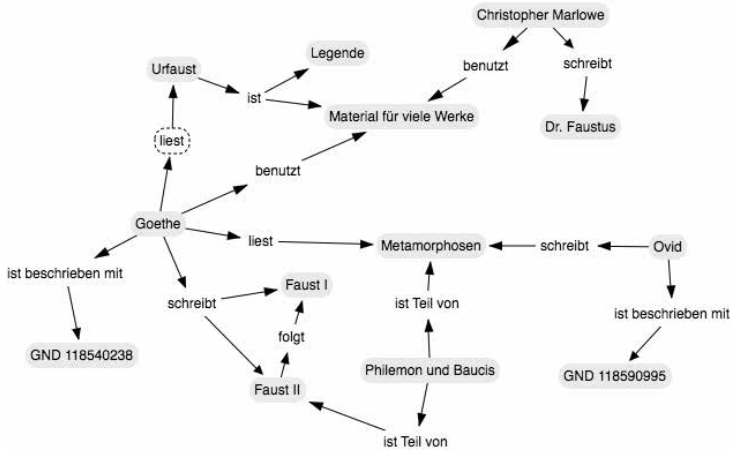


Abbildung 3: Natürlichsprachliche Triple

Das Ergebnis ist sicherlich für die Faustforschung unergiebig. Es fehlen dezidiert literaturwissenschaftliche Relationen – wie zum Beispiel der intertextuelle Bezug zu Philemon und Baucis. Das ist jedoch nicht dem Projekt RelFinder zuzuschreiben, sondern der Datenbasis, die eben noch zu dünn ist. Doch selbst jetzt bereits wird man durch die Visualisierung auf Dinge hingewiesen, die man vielleicht beim Beginn der Recherche nicht im Blick hatte³¹ – wie zum Beispiel den angezeigten Filmen.

Als Quintessenz kann man festhalten, dass die Entwicklung hin zu einem *semantic web* mit Hinzufügung von Editionsdaten auf einem guten Weg ist. Schaut man sich DBPedia an, also ein Croudsourcing Projekt wie die Wikipedia selbst, bei dem semantische Entitäten im Sinne des W3C aus der Wikipedia extrahiert werden, so stellt man fest, dass die guten Effekte, wie sie etwa der RelFinder zeigt, durch falsche Schlüsse überlagert werden: Dass Shakespeare durch Goethe beeinflusst wurde, kann schlicht nicht sein. Es wäre deshalb sinnvoller, die RDF-Tripel selbst herzustellen oder die Editionsdaten so aufzubereiten, dass die RDF-Tripel sich eindeutig daraus ableiten. Ein Aspekt ist dabei besonders wichtig, dass nämlich die Experten für ein Gebiet selbst die Relationen beschreiben; aber auch nur die Relationen ihres eigenen Fachgebietes und nicht auf Vollständigkeit aller möglichen Relationen zielen. Das kann nur zu Ungenauigkeiten, zum Beispiel zu Dubletten führen. Vielleicht reicht

³¹ Dies ist das Ziel des vom Bundesministerium für Bildung und Forschung BMBF geförderten Projektes „Vernetzte Korrespondenzen | Exilnetz33. Erforschung und Visualisierung sozialer, räumlicher, zeitlicher und thematischer Netze in Briefkorpora“, siehe <http://exilnetz33.de/de/> [Stand: 30. Nov. 2015].

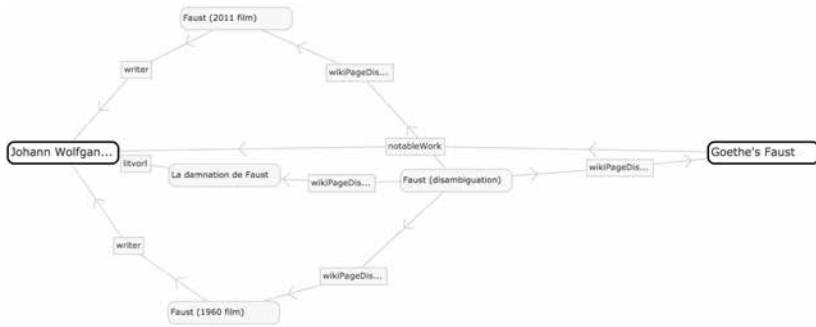


Abbildung 4: RelFinder Graph zu Goethe und Goethes *Faust*.

Siehe <http://www.visualdataweb.org/refinder/refinder.php> [Stand: 30. Nov. 2015] mit Eingabe der zwei Stichwörter „Goethe, Johann Wolfgang von“ und „Goethes Faust“

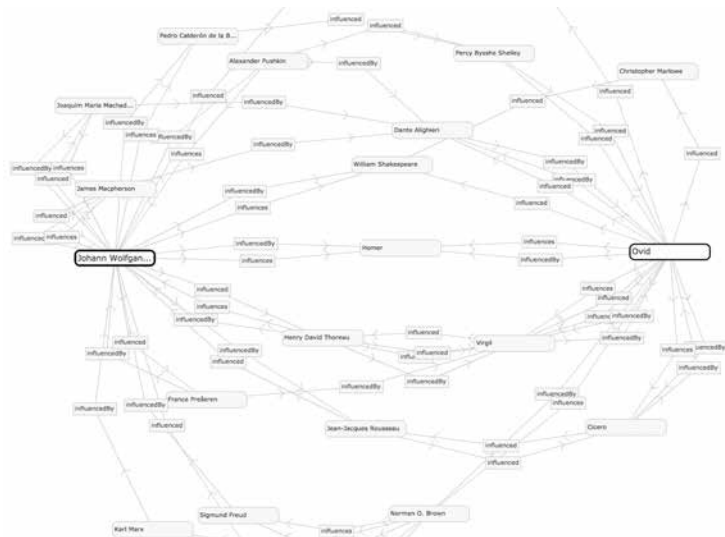


Abbildung 5: RelFinder Graph zu Goethe und Ovid.

Siehe <http://www.visualdataweb.org/refinder/refinder.php> [Stand: 30. Nov. 2015] mit Eingabe der zwei Stichwörter „Goethe, Johann Wolfgang von“ und „Ovid“

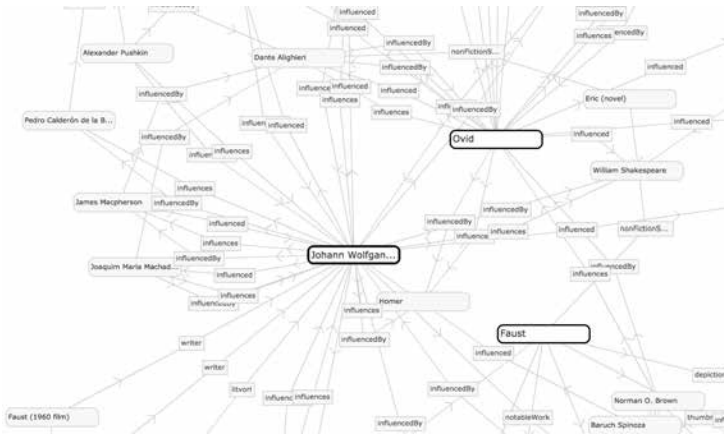


Abbildung 6: RelFinder Graph zu Goethe, Ovid und *Faust*.

Siehe <http://www.visualdataweb.org/refinder/refinder.php> [Stand: 30.Nov.2015] (Ausschnitt) mit Eingabe der drei Stichwörter „Goethe, Johann Wolfgang von“ und „Ovid“ und „Faust“

es künftig tatsächlich für einen Bibliothekar, den Werktitel nach Vorlage sowie den Autorennamen nach Vorlage zu erfassen und diese beiden Entitäten miteinander zu verknüpfen. Die Abfragesprache berücksichtigt dann den gesamten Datenbestand im W3, um die Beziehungen zwischen Werk und Person eindeutig zu definieren. Dies würde aber bedingen, dass alle Experten ihre Daten als RDF-Tripel eingeben. Das scheint zu kompliziert, wie das Beispiel am Datensatz der Deutschen Biographie zu Harry Graf Kessler gezeigt hat. Kaum scheint es gelungen zu sein, die Editoren auf XML mit den Richtlinien der TEI³² eingeschworen zu haben, sollen sie RDF/XML erlernen? Und vielleicht dann auch RDF/XML mit Richtlinien der TEI? Das führt doch allmählich zu weit weg vom philologischen Terrain, auch wenn eine grundsätzliche computertechnische Kompetenz inzwischen vorausgesetzt werden muss und kann.

Auch Bibliotheken setzen nicht auf die Eingabe von Tripeln nach RDF, sondern bauen weiterhin auf an RDA angepasste Katalogsysteme, aus denen RDF in einer der möglichen Sprachen erzeugt wird. Doch wie überprüfen sie die Richtigkeit der übersetzten Syntax? Für Editoren gibt es bereits einen epistemischen Grund, Algorithmen nach den Regeln der Wissenschaftlichkeit zu verifizieren oder wenigstens zu plausibilisieren, d. h. blindes Vertrauen auf rechnergestützte Verfahren darf es in der Wissenschaft

³² Text Encoding Initiative (TEI), URL: <http://www.tei-c.org> [Stand: 30.Nov.2015].

nicht geben. Editoren müssten also die Transferleistung von ihrer eigenen Erfassung in RDF überprüfen und gegebenenfalls korrigieren, was wiederum die Kenntnis von RDF/XML bedingen würde. Es stellt sich die Frage, warum in der heutigen Zeit mit den heutigen technischen Möglichkeiten keine natürlichsprachlichen Tripel realisierbar sind. Dabei denke ich in der Tat an die Form wie oben in Abb. 3 beschrieben. Wenn Computerlinguisten es schaffen, aus einem Werk wie den Thomas Mannschen *Buddenbrooks* mit NER-Taggern³³ sämtliche Romanfiguren zu erkennen, dann kann es nicht unmöglich sein, in gut aufbereiteten natürlichsprachlichen Tripeln zielsicher Anschlussmöglichkeiten herzustellen. Auch die Sprache darf dabei keine Rolle spielen, wenn man zwei Grundregeln für die Prädikate anwendet: 1. Alle Prädikate werden im Präsens verfasst. 2. Alle Prädikate werden aktivisch gefasst. Auf diese Weise könnten alle Experten hochqualifizierte Daten-Tripel erstellen.

Nun bleibt noch die Frage offen, wie denn das *semantic web* zu einem sinnvollen, Bedeutung transportierenden Informationsmedium wird. Der RelFinder kann die DBpedia verwenden, weil sie einen SPARQL-Endpoint bietet, der gezielt abgefragt wird. Damit sind aber nur die Datenquellen im Blick, die auch bekannt sind. Das ist zu wenig. Die Holschuld muss in eine Bringschuld umgewandelt werden, der Fetch- in einen Pushdienst. So wie alle mobilen Daten potenziell überall verfügbar sind, müssen auch alle semantischen Tripel überall verfügbar sein. Es muss eine Technologie entwickelt werden, die jede RDF-Quelle mit einem Sender ausstattet, statt mit einer Schnittstelle. Der Receiver fragt dann in den virtuellen Raum und empfängt relevante Daten, die dann gefiltert, übersetzt und zusammengesetzt werden müssen. Erst dann entstehen Graphen, die von Experten gespeist werden und diese dann wieder durch die Zusammenschau, die Visualisierung zu neuen Fragestellungen führen. Erst dann entsteht ein wirkliches, hochqualifiziertes *semantic web*.

³³ NER = Named Entity Recognition.

```

<rdf:RDF
  ↪ xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  ↪ xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  ↪ xmlns:gnd="http://d-nb.info/standards/elementset/gnd#"
  ↪ xmlns:rda="http://rdvocab.info"
  ↪ xmlns:rel="http://purl.org/vocab/relationship/"
  ↪ xmlns:shv="http://ns.aksw.org/spatialHierarchy/"
  ↪ xmlns:acal="http://cal.aksw.org/"
  ↪ xmlns:ndb="http://data.deutsche-biographie.de/"
  ↪ xmlns:ndbvoc="http://data.deutsche-
  ↪ biographie.de/vocabulary/"
  ↪ xmlns:tei="http://www.tei-c.org/ns/1.0"
  ↪ xmlns:rdaGr2="http://RDVocab.info/ElementsGr2/"
  ↪ xml:base="http://data.deutsche-biographie.de/">
  <rdf:Description rdf:about="http://data.deutsche-
  ↪ biographie.de/Person/sfz40697">
    <rdf:type rdf:resource="http://data.deutsche-
  ↪ biographie.de/vocabulary/Person"/>
    <rdfs:seeAlso
  ↪ rdf:resource="http://d-nb.info/gnd/118561685"/>
    <rdfs:seeAlso rdf:resource="http://linkedhisto-
  ↪ ry.aksw.org/pnd/118561685"/>
    <rdfs:seeAlso
  ↪ rdf:resource="http://beacon.findbuch.de/seealso/
  ↪ pnd-aks?format=sources&id=118561685"/>
    <rdaGr2:identifierForThePerson
  ↪ sfz40697</rdaGr2:identifierForThePerson>
    <ndbvoc:page
  ↪ http://www.deutsche-biographie.de/sfz40697.html</ndbvoc:page>
  ↪
  ↪
  ↪ <rdaGr2:gender>Male</rdaGr2:gender>
    <gnd:gender rdf:resource="http://d-
  ↪ nb.info/standards/vocab/gnd/Gender#male"/>
    <ndbvoc:gender rdf:resource="http://data.deutsche-
  ↪ biographie.de/vocabulary/Male"/>
    <gnd:preferredNameForThePerson>Kessler, Harry Graf
  ↪ von</gnd:preferredNameForThePerson>
    <rdfs:label>Kessler, Harry Graf von</rdfs:label>
    <rdaGr2:periodOfActivityOfThePerson>1868 bis
  ↪ 1937</rdaGr2:periodOfActivityOfThePerson>
    <gnd:periodOfActivityOfThePerson>1868 bis
  ↪ 1937</gnd:periodOfActivityOfThePerson>
    <rdaGr2:professionOrOccupation>Schriftsteller;
  ↪ Diplomat</rdaGr2:professionOrOccupation>
    <gnd:professionOrOccupation>Schriftsteller;
  ↪ Diplomat</gnd:professionOrOccupation>
    evangelisch 1868 186818681868;1868*1868-05-23Paris 1937
  ↪ 193719371937;1937*1937-11-30Lyon
  ↪ <!-- pg -->
    <rel:childOf rdf:resource="http://data.deutsche-
  ↪ biographie.de/Person/sfz40686"/>
  ↪ <!-- 1 -->
  </rdf:Description>
</rdf:RDF>

```

Listing 1: RDF-Datensatz zu Harry Graf Kessler aus der *Deutschen Biographie*,
 URL: <http://data.deutsche-biographie.de/rest/sfz40697.rdf> [Stand: 30. Nov. 2015]