

Stefan Münnich

Ontologien als semantische Zündstufe für die digitale Musikwissenschaft? Eine Bestandsaufnahme

Zusammenfassung: Ontologien spielen eine zentrale Rolle für die formalisierte Repräsentation von Wissen und Informationen sowie für die Infrastruktur des sogenannten *semantic web*. Trotz früher Initiativen der Bibliotheken und Gedächtnisinstitutionen hat sich die deutschsprachige Musikwissenschaft insgesamt nur sehr zögerlich dem Thema genähert. Im Rahmen einer Bestandsaufnahme werden neben der Erläuterung grundlegender Konzepte, Herausforderungen und Herangehensweisen bei der Modellierung von Ontologien daher auch vielversprechende Modelle und bereits erprobte Anwendungsbeispiele für eine ‚semantische‘ digitale Musikwissenschaft identifiziert.

Schlüsselwörter: Semantic Web; Ontologien; Wissensrepräsentationen; Digitale Musikwissenschaft

Ontologies as a Semantic Booster for Digital Musicology? An Overview

Abstract: Ontologies play a crucial role for the formalized representation of knowledge and information as well as for the infrastructure of the semantic web. Despite early initiatives that were driven by libraries and memory institutions, German musicology as a whole has turned very slowly to the subject. In an overview the author addresses basic concepts, challenges and approaches for ontology design and identifies models and use cases with promising applications for a ‚semantic‘ digital musicology.

Keywords: Semantic web; ontologies; knowledge representation; digital musicology

1 Hinleitung und Motivation

In der von Tim Berners-Lee und anderen um die Jahrtausendwende formulierten Vision eines *web of data*¹, auch bekannt als *semantic web*, kommt bei der Strukturierung und semantischen Anreicherung dieses projektierten Daten- und Wissensgeflechts² eine Schlüsselstellung den Ontologien zu (siehe Abb. 1). Wissen

¹ Sehr früh in Berners-Lee, Tim: What the Semantic Web Can Represent. September 1998.

<https://www.w3.org/DesignIssues/RDFnot.html>. Ausführlicher in Berners-Lee, Tim; Hendler, James; Lassila, Ora: The Semantic Web. In: Scientific American Magazine (17. Mai 2001).

<https://pdfs.semanticscholar.org/566c/1c6bd366b4c9e07fc372771690d5ba31.pdf>.

² Zum Geflecht als Urbild postmoderner Wissensrepräsentation vgl. Umberto Eco, *Opera aperta*, Mailand 1962; dt. Fassung: *Das offene Kunstwerk*, übers. von Günter Memmert. Frankfurt a.M. 1973, S. 201: „[...] die Anerkennung der Tatsache, dass die Welt ein Geflecht von Möglichkeiten ist“, sowie den Rhizom-Begriff bei Deleuze, Gilles; Guattari, Félix: *Rhizome*. Introduction. Paris 1976; dt. Ausgabe: *Rhizom*. Aus dem Französischen von Dagmar Berger, Clemens-Carl Haerle, Helma Konyen, Alexander Krämer, Michael Nowak und Kade Schacht. Berlin 1977.

wird hier nicht durch zusammenschaltete und kontextualisierte, verknüpfte Dokumente strukturiert, sondern durch die zusammenschalteten und kontextualisierten Daten selbst.

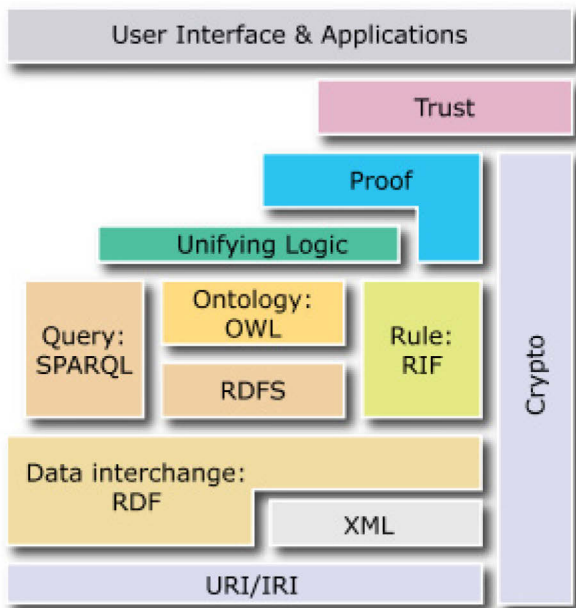


Abb. 1: Schichtenmodell der *semantic web*-Technologien. Quelle: <https://www.w3.org/2007/03/layerCake-small.png>.

Erfreulicherweise erkannten Bibliotheken und Gedächtnisinstitutionen bereits relativ früh das Potenzial von Ontologien für ihre Zwecke. Ihre Bestrebungen nach besseren Auffindbarkeits-, Integrations- und Verknüpfungsstrategien für ihre Metadaten, ihre diesbezüglichen Erfolge und Irrwege sowie steten Diskussionsbeiträge³ blieben nicht ohne Konsequenz. Inzwischen wurden wiederholt Chancen und Grenzen semantischer Technologien im Bereich des kulturellen Erbes⁴ und sogar im Hinblick auf digitale (Text-)Editionen⁵ im deutschsprachigen Raum diskutiert. Spätestens mit dem Freien Symposium „Wissenssystematiken im digitalen Zeitalter“ auf der Jahrestagung 2017 der Gesellschaft für Musikforschung haben Fragen von Wissensstrukturen und Ontologien auch endgültig die deutschsprachige (digitale) Musikwissenschaft insgesamt erreicht.

³ Siehe u.a. Mittermeier, Bernhard (Hg.): *Vernetztes Wissen – Daten, Menschen, Systeme*. 6. Konferenz der Zentralbibliothek Forschungszentrum Jülich, 5.–7. November 2012. Jülich 2012 (Reihe Bibliothek/Library; 21). <http://hdl.handle.net/2128/4699>. Danowski, Patrick; Pohl, Adrian (Hg.): *(Open) Linked Data in Bibliotheken*. Berlin/Boston 2013 (Bibliotheks- und Informationspraxis; 50). <http://dx.doi.org/10.1515/9783110278736>. Mittelbach, Jens: *Modernes Datenmanagement: Linked Open Data und die offene Bibliothek*. In: *o-bib - das offene Bibliotheksjournal* 2(2) (2015) S. 61–73. <http://dx.doi.org/10.5282/o-bib/2015H2S61-73>.

⁴ U.a. bei Hohmann, Georg; Fichtner, Mark: *Chancen und Herausforderungen in der praktischen Anwendung von Ontologien für das Kulturerbe*. In: *Digitales Kulturerbe. Bewahrung und Zugänglichkeit in der wissenschaftlichen Praxis*. hg. von Caroline Y. Robertson-von Trotha und Ralf H. Schneider. Karlsruhe 2015, S. 115–128. <http://dx.doi.org/10.5445/KSP/1000044869>.

⁵ Wettlaufer, Jörg: *Der nächste Schritt? Semantic Web und digitale Editionen*. Referat auf der Tagung „Digitale Metamorphose. Digital Humanities und Editionswissenschaft“, 2.–4.11.2015, Herzog August Bibliothek, Wolfenbüttel, und Kamzelak, Roland S.: *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 Kristina Richts und Peter Stadler für den Virtuellen Forschungsverbund Edirom. München 2016, S. 423–436.

Was aber genau sind Ontologien? Wie kann mit ihrer Hilfe Wissen systematisch aufbereitet, modelliert und verknüpft werden? Welche Herausforderungen bestehen bei einer solchen Modellierung? Und welche Modelle und Anwendungen existieren für eine digitale Musikwissenschaft? Im Folgenden sollen im Sinne einer Bestandsaufnahme Antwortmöglichkeiten auf diese Fragen aufgezeigt werden.

2 Was sind Ontologien?

Im informationstechnischen Sinn werden unter ‚Ontologien‘⁶ verschiedene Arten formaler Begriffssysteme zusammengefasst, die eine explizite und somit maschinell verarbeitbare Zuweisung von Bedeutung zu gemeinschaftlich definierten sprachlichen Konzepten ermöglichen.⁷ Es geht also einfach gesprochen darum, wie menschliches Wissen regelbasiert beschrieben und repräsentiert werden kann, um es einer maschinenverarbeitbaren Speicherung, Auswertung und Nachnutzung zugänglich zu machen.

Die unter diesem Begriff versammelten Ansätze werden oftmals entweder nach dem Grad ihrer semantischen Mächtigkeit oder nach der Spezifik ihres Anwendungsbereiches klassifiziert.⁸ Im ersten Fall ergibt sich ein semantisches Spektrum von einfacheren *lightweight ontologies* wie Verzeichnissen und Katalogen mit kontrollierten Vokabularen (wie sie für Normdateien Anwendung finden) und Taxonomien über Thesauri hin zu semantisch ausdrucksstarken *heavyweight ontologies*, die die Anwendung formaler und logischer Operationen erlauben. Im zweiten Fall erhält man eine Gegenüberstellung von generischeren *top-level ontologies* und spezifischeren *application ontologies* sowie gegenstands- bzw. aufgabenbezogenen *domain* bzw. *task ontologies*.

In vielen Anwendungsfällen lassen sich solche Abgrenzungsversuche jedoch nicht strikt aufrechterhalten, da eine Verflechtung verschiedener Anwendungsbereiche stattfindet. Auch werden in einem engeren Begriffsverständnis häufig erst oberhalb von Thesauri klassifizierte Formalisierungen als ‚Ontologien‘ angesprochen.

⁶ Der Begriff hat eigentlich keinen Singular und auch keine adjektivische Form, da unter „Ontologie“ bzw. „ontologisch“ im Allgemeinen die entsprechende philosophische Teildisziplin verstanden wird, von der die Pluralform abgeleitet wurde.

⁷ Diese Beschreibung folgt der Definition von Studer, Benjamins und Fensel (1998): „An ontology is a formal, explicit specification of a shared conceptualisation.“ (Studer, Rudi; Benjamins, Richard; Fensel, Dieter: Knowledge Engineering: Principles and Methods. In: Data & Knowledge Engineering 25(1–2) (1998) S. 161–198, hier S. 186). Sie spezifiziert die sonst identische, häufig anzutreffende Definition von Thomas Gruber durch die Begriffe „formal“ und „shared“. Siehe Gruber, Thomas: A Translation Approach to Portable Ontology Specifications. In: Knowledge Acquisition 5(2) (1993) S. 199–220, hier S. 199.

⁸ Diese beiden häufigsten Klassifizierungsmodelle gehen auf Ora Lassila und Deborah McGuinness (2001) bzw. auf Nicola Guarino (1998) zurück. Siehe Rehbein, Malte: Ontologien. In: Digital Humanities. Eine Einführung, hg. von Fotis Jannidis, Hubertus Kohle und Malte Rehbein. Stuttgart 2017, S. 162–176, hier S. 165–166, und Hohmann/Fichtner (Anm. 4) S. 116–118.

3 Formalisierung von Wissen durch Ontologien

Ein wesentlicher Vorteil von Wissensrepräsentationen mittels Ontologien besteht in der grundlegenden Trennung von abstrakter Terminologie (*terminological knowledge*), individuellen Zuweisungen (*assertional knowledge*) sowie logischen Ableitungen (*inferred knowledge*):⁹

Die formale, abstrakte Struktur der betrachteten Konzepte und ihrer Beziehungen zueinander beschreibt das terminologische Wissen. Hier wird gefragt: „Wie, also mit welchen Begriffen, spreche ich über den betrachteten Weltausschnitt?“ Je nach Perspektive, Fragestellung usw. kann es hier ganz unterschiedliche Begrifflichkeiten und Konzepte für ein und denselben Gegenstandsbereich geben. Grundsätzlich lassen sich bestimmte ‚Grundbausteine‘ der terminologischen Beschreibung beobachten: Zum einen Klassen (*classes, entities*), die Konzepte oder Dinge der realen Welt abstrakt repräsentieren, wie z.B. ‚Person‘, ‚Rolle‘, ‚Ereignis‘, ‚Werk‘¹⁰. Diese Klassen besitzen charakteristische Eigenschaften (*attributes, properties*), die sie näher spezifizieren. Attribute können auf einen einfachen Datenwert (*literal*) oder wiederum auf eine Klasse verweisen (*relations*). Gerade diese attributgesteuerten Relationen sind also grundlegend für die Verknüpfung der Klassen untereinander. Ebenfalls in den Bereich des terminologischen Wissens zählen formale Definitionen von Grundannahmen (*axioms*) oder Einschränkungen (*constraints*), die eine logisch konsistente Verwendung von Attributen und Eigenschaften garantieren sollen.

Den Konzepten einer solchen abstrakten Struktur können konkrete Individuen, sogenannte Instanzen, zugewiesen werden. Die Gesamtheit aller Instanzen lässt sich als assertionales, zugewiesenes Wissen beschreiben, hinter dem die Frage steht: „Welche konkreten Behauptungen über den betrachteten Weltausschnitt kann ich mittels meiner Terminologie treffen?“. Einer Instanz („Anton Webern“) der Klasse ‚Person‘ könnte z.B. die konkrete Rolle „Komponist“ zugewiesen werden, mit der individuelle Kompositionereignisse verknüpft sind, die auf verschiedene Kompositionen (Instanzen der Klasse ‚Werk‘) verweisen (vgl. Abb. 2). Der Vorgang könnte mit weiteren Rollen (z.B. „Dirigent“, „Leser“), Ereignissen usw. immer weiter ausdifferenziert werden. Derlei Aussagen über Entitäten der realen Welt sind immer subjektiven Zuweisungsakten unterworfen und müssen von einem anderen Standpunkt aus evtl. modifiziert, revidiert oder gar negiert werden.

Ihr volles Potenzial entfalten Ontologien, wenn zusätzlich inferenziertes, implizites Wissen durch automatisierte logische Schlußfolgerungen abgeleitet werden kann. Dies geschieht häufig durch den Einsatz sogenannter *reasoner*¹¹, die auf Grundlage der dem Modell zugrundeliegenden Strukturen derartige

⁹ Hierzu und zu den folgenden Ausführungen vgl. Rehbein (Anm. 8) S. 164–165.

¹⁰ Das auch in der Musikwissenschaft problematische Konstrukt ‚Werk‘ wird in der Praxis mittlerweile sehr viel offener und komplexer modelliert, z.B. unter Einbeziehung verschiedener Abstraktionsebenen wie *work, expression, manifestation, item* (FRBR) oder durch den triadischen Zusammenschluss von *Individual Work, Expression Creation, Self-Contained Expression* (DOREMIUS basierend auf CIDOC CRM).

¹¹ Näheres zur Funktionsweise in Sattler, Uli; Stevens, Robert; Lord, Phillip: How does a reasoner work? In: *Ontogenesis* (12. August 2014). <http://ontogenesis.knowledgeblog.org/1486>.

Ableitungen ermöglichen. Sie haben bei der Konsistenzprüfung von terminologischem und assertionalem Wissen eine wichtige Funktion: So können z.B. Zuordnungen von Instanzen zu sich gegenseitig ausschließenden Klassen aufgedeckt werden; ein durchaus trivialer, dennoch häufiger Fehler. Im Gegensatz zum assertionalen Wissen mit seinen bewussten Zuordnungen im Rahmen der Wissensinferenz treten hier vor allem die un(ter)bewussten Implikationen des Modells zu Tage, die kritisch hinterfragt und überprüft werden sollten.

4 Kritische Aspekte

Jede Modellierung und Nutzung von Ontologien steht vor grundlegenden Herausforderungen, von denen im Folgenden nur einige kurz erörtert werden sollen:

4.1 AAA-Pradigma („Zwei mal drei ist vier...“)

Für im Web veröffentlichte Daten und Informationen gibt es keinen Verbraucherschutz. Es gilt das fundamentale Prinzip: „Anyone can say Anything about Any topic“ (AAA-Paradigma).¹² Laut Allemang und Hendler wird der römische Wachsamkeitsgrundsatz *caveat emptor* („möge der Käufer sich in Acht nehmen“) hier auf die Spitze getrieben zu einem *caveat auditor*, bei dem die Verantwortlichkeit für die Einordnung einer beliebigen Information auf deren Empfänger übertragen wird. Im Zusammenhang mit Ontologien bedeutet dies, dass bei der Wiederverwendung existierender Modelle die dort beschriebenen Konzepte sehr sorgsam und kritisch auf eine Anwendbarkeit und Übereinstimmung mit eigenen Konzepten überprüft werden müssen. Im Falle der Selbstmodellierung wiederum ist eine detaillierte Dokumentation der Bedeutung und Anwendungsfälle, mitunter auch der Nichtanwendungsfälle, geboten, um potenziellen Nutzern eine Einordnung und Bewertung zu erleichtern.

4.2 Surrogatcharakter

Es gibt mehr als einen „richtigen“ Weg, Wissensstrukturen mit Hilfe von Ontologien zu repräsentieren. Dies ist immer von dem Untersuchungsgegenstand, den darauf anzuwendenden Fragestellungen sowie der Perspektive, die der Untersuchende diesen gegenüber einnimmt, abhängig. Wenn jedwede Wissensrepräsentation per se nur ein reduziertes, vereinfachtes und somit unvollkommenes Modell des betrachteten Weltausschnitts darstellt, sind auch Ontologien stets von einer bestimmten, mehr oder weniger unvollkommenen Sicht- und Herangehensweise geprägt, sie können nicht allumfassend und abschließend sein.¹³ Eine Modellierung sollte daher auch nicht diese unvermeidbare Reduktion letztendlich zu überwinden trachten, sondern – in vollem Bewusstsein ihrer Existenz – eine hinreichend genaue und den praktischen Aufwand vertretbare Annäherung an den betrachteten Weltausschnitt ermöglichen.

¹² Siehe Allemang, Dean; Hendler, James: *Semantic Web for the Working Ontologist. Effective Modelling in RDFS and OWL*. 2. Aufl. Amsterdam u.a. 2011, S. 6f.

¹³ Siehe dazu die grundlegende Untersuchung von Davis, Randall; Shrobe, Howard; Szolovits, Peter: *What is a Knowledge Representation?* In: *AI Magazine* 14(1) (1993) S. 17–33, hier S. 18–19. <http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>.

4.3 Explizität

Das Explizitmachen, das Explizitsein von Begriffen und Konzepten spielt eine entscheidende, nicht deutlich genug zu betonende Rolle. Sämtliche Bedeutungskontexte, die stets implizit Mitzudenken und Einzuordnen bereits die „hermeneutische Kompetenz“¹⁴ menschlicher Akteure vor Herausforderungen stellen kann, müssen einer maschinellen Verarbeitung zugänglich gemacht werden. In gewissem Sinne wird eine solche Kompetenz bei der Modellierung durch die Annahme einer offenen Welt (*open world assumption*) simuliert. Dass z.B. eine Klasse eine bestimmte Eigenschaft nicht hat, bedeutet hier eben nicht, dass diese Eigenschaft nicht existiert (wie unter der Annahme einer geschlossenen Welt, *closed world assumption*), sondern lediglich, dass bislang keine explizite Aussage über sie getroffen wurde. Was unter Annahme einer geschlossenen Welt nicht explizit modelliert bzw. aus einem Modell auch nicht inferenzierbar ist, existiert für die Maschine schlichtweg nicht.

Allgemein wird die Auflösung und Beseitigung von Ambiguitäten durch die eineindeutige semantische Zuschreibung innerhalb von Ontologien klar begrüßt. Im Zweifelsfall muss das Modell nur soweit explizit sein, um in dem von ihm betrachteten Ausschnitt, in seiner Fragestellung eineindeutig zu sein. Allerdings besteht dabei die Gefahr, dass durch die Modellierung und Deutlichmachung einer bestimmten Bedeutungsebene die gerade in den Geisteswissenschaften oft bewusst verwendeten begrifflichen Ambivalenzen oder Mehrdeutigkeiten¹⁵ auf einer anderen Ebene untergehen. Dieser Verlust der Unschärfe, der ‚Ambiguitätsdimensionen‘¹⁶ eines Begriffs, die erst seine multiperspektivische Anwendung ermöglicht, kann mitunter zu seiner ‚Verarmung‘, seiner ‚Simplifizierung‘ führen. Die Schwierigkeit besteht also nicht in der faktischen, reduzierenden Modellierung eines Begriffs, sondern in der expliziten Bewusstmachung und Abbildung der Vielschichtigkeit und Mannigfaltigkeit seiner Bedeutungen: Komplexität und Vielschichtigkeit bei gleichzeitiger Explizität.

4.4 Unsicherheiten, Argumente und Kontext

Unsicherheiten bilden ein zentrales Konzept und zuweilen auch Forschungsgegenstand der (digitalen) Geisteswissenschaften. Sie verhandeln zu großen Teilen Indizien, nicht Beweise, angefangen bei unklaren historischen Daten („Ende des 17. Jahrhunderts“) über Zuschreibungen von historischen Zeugnissen zu Personen, Gruppen oder Zeitabschnitten bis hin zu interpretativen Analysen von Musikstücken oder anderen Kunstwerken. In all diesen Fällen handelt es sich eben nicht um einfache ‚Rohdaten‘, sondern um

¹⁴ So der Berliner Philosoph Wilhelm Schmid über die „Fähigkeit, immer mehr in den Dingen zu vermuten, als wir momentan sehen“. Siehe Schwilk, Heimo: Interview mit Wilhelm Schmid. In: *Welt am Sonntag* (13.11.2005). <https://www.welt.de/print-wams/article134958/Ich-plaedierte-fuer-das-Glueck-der-Fuelle.html>.

¹⁵ Vgl. u.a. Deleuze/Guattari (Anm. 2) S. 33: „Problem der Schrift: nur mit ungenauen Ausdrücken kann man etwas genau bezeichnen. Nicht, weil man da hindurch müßte oder immer nur durch Annäherung vorankäme: die Ungenauigkeit ist keineswegs eine Annäherung, sie ist im Gegenteil der genaue Verlauf der Ereignisse.“

¹⁶ Bauer, Matthias; Knappe, Joachim; Koch, Peter; Winkler, Susanne: Dimensionen der Ambiguität. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 40 (2010) S. 7–75.

,kontextualisierte Daten'¹⁷. Auch Editionen, wollen sie mehr sein als nur das zugängliche Ablegen codierter Dateien, erzeugen kontextualisierte Daten, kontextualisiertes Wissen, das abgreifbar, verknüpfbar und weiterverwendbar aufbereitet werden muss.

Eine Möglichkeit, einen solchen Kontext formal abzubilden, wurde mit der Erweiterung *CRMinf*¹⁸ für das CIDOC Conceptual Reference Model (CRM; siehe unter Ontologiemodelle) eingeführt: Das dort spezifizierte „Argumentation Model“ erlaubt eine explizit semantische Modellierung von Unsicherheiten durch die Zuordnung eines Wahrheitswerts (*I6 Belief Value*: wahr/falsch/unbekannt) zu einer Annahme (*I2 Belief*) über einen Sachverhalt (*I4 Proposition Set*). Durch Ableitung weiterer Annahmen (*I5 Inference Making*) könnte ein kompletter Argumenationsstrang (*I1 Argumentation*) abgebildet werden.

Es wird deutlich, dass für die geisteswissenschaftliche Nutzung intendierte Ontologien einen sehr hohen formalen Komplexitätsgrad aufweisen können, was sowohl bei ihrer Erstellung als auch späteren Nutzung zu einer sehr viel stärkeren Bindung zeitlicher Ressourcen führt.

5 Modellierungsschritte

Der Modellierungsprozess von Ontologien ist überhaupt ein zeit- und kostenintensiver Prozess, der in enger Rückkopplung von Spezialisten des betreffenden Fachgebiets (Domain-Experten) und Spezialisten im Bereich der Ontologie-Entwicklung durchgeführt werden sollte. Trotz diverser Ansätze und Methodologien hat sich noch kein Standard-Vorgehensmodell etabliert, da stets individuelle Ausgangspunkte sowie Zielsetzungen berücksichtigt werden müssen. In der praktischen Anwendung hat sich jedoch ein an Modelle des *software engineering* angelehnter, teilweise rekursiv auszuführender, Prozess aus sechs Schritten bewährt:¹⁹

- 1) Spezifikation & Analyse
- 2) Konzeptualisierung
- 3) Formalisierung

¹⁷ Auf diesen wesentlichen Aspekt geisteswissenschaftlicher Forschungsdaten wurde wiederholt aufmerksam gemacht, u.a. In Oldman, Dominic; Doerr, Martin, de Jong, Gerald, Norton, Barry, Wikman, Thomas: Realizing Lessons of the Last 20 Years: A Manifesto for Data Provisioning & Aggregation Services for the Digital Humanities (A Position Paper). In: D-Lib Magazine 20(7/8) (July/August 2014). DOI:10.1045/july2014-oldman.

¹⁸ Spezifikation unter <http://www.cidoc-crm.org/crminf/sites/default/files/CRMinf-0.7%28forSite%29.pdf>. Dominic Oldman vom British Museum gebührt herzlicher Dank für den Hinweis auf diese Modellierungsoption. Bislang ist mir jedoch kein Datensatz bekannt, der diese Vorgehensweise auf den gesamten Datenbestand anwendet.

¹⁹ Zuweilen werden der zweite und dritte Schritt als „Entwurfsphase“ zusammengefasst, wodurch sich – unter Auslassung der Dokumentation – ein nur vierphasiges Modell ergibt. Das sechsteilige Modell folgt Nurmikko-Fuller, Terhi, LD4DH Session 2.1: Ontologies. Unveröffentlichte Präsentation im Rahmen der Digital Humanities at Oxford Summer School (DHOxSS) 2017. Ausführlich werden Probleme und Methoden für die Modellierung diskutiert in Stuckenschmidt, Heinz: Erstellen von Ontologien. In: Ontologien. 2. Aufl. Berlin 2011 (Informatik im Fokus), S. 155–205. https://doi.org/10.1007/978-3-642-05404-4_5, und Karakol, Fatih: Konzeption einer ontologiebasierten Schnittstelle zur Integration von verteilt vorliegenden Informationsquellen, Hamburg: disserta 2016, S. 59–69.

- 4) Implementierung
- 5) Evaluation
- 6) Dokumentation

Im ersten Schritt werden Motivation und Ziele der Modellierung zusammen mit den Domain-Experten analysiert. Die für die Domain-Experten wichtigen Konzepte werden danach aufgenommen, und benötigte Klassen, Eigenschaften und deren Beziehungen zueinander festgehalten. Im darauffolgenden Formalisierungsschritt wird die Hierarchie dieser Konzepte definiert und in ein strukturiertes Modell gebracht (häufig noch per Hand auf Papier). Ebenfalls in diesen Formalisierungsprozess fällt die Festlegung, welche Konzepte aus anderen Ontologien übernommen und welche neu modelliert werden können. Das tatsächliche „Schreiben“ der Ontologie erfolgt in der Implementierungsphase. Dabei werden die formale Sprache, in der das Modell umgesetzt wird, und die verwendete Software (Editor) und ggf. weitere Tools (z.B. ein *reasoner*) bestimmt. Während der Evaluationsphase werden sowohl die formale (durch Validierungstools oder Abfragen an das Datenset) als auch die inhaltliche (durch Feedback der Domain-Experten) Korrektheit sichergestellt. Im Zweifelsfall müssen hier zuvor gefällte Entscheidungen überdacht oder verworfen werden und der Prozess wird ab dem entsprechenden Schritt wiederholt. Oft vernachlässigt wird eine den gesamten Prozess begleitende Dokumentation, die u.a. die getroffenen Design-Entscheidungen und Bedeutungen der verwendeten Konzepte erläutert und nachvollziehbar macht. Wichtig bleibt, dass sämtliche Einzelschritte des Workflows Auswirkungen auf die modellierte Wissensstruktur haben.

6 Handwerkszeug: Von Turtle, OWL und SPARQL

Die vom W3-Consortium empfohlenen Standard-Beschreibungssprachen im Zusammenhang mit Ontologien sind RDF, RDFS und OWL, als Abfragesprache dient SPARQL.²⁰ Elementar ist hierbei die Verwendung sogenannter IRIs (*Internationalized Resource Identifier*) zur Bezeichnung von Klassen, Eigenschaften, Relationen oder Instanzen. Diese global eineindeutigen Webadressen sind der ‚digitale Fingerabdruck‘ eines im Web repräsentierten Objekts, auch Ressource genannt, und ermöglichen so eine distinkte Identifizierung und Unterscheidung jeglicher Ressourcen.

Das *Resource Description Framework* (RDF) definiert eine simple, grundlegende Syntax für die Beschreibung von Ressourcen: <subject> <predicate> <object>. Jegliche Aussage wird in dieser dreiteiligen Form (Tripel) getätigt. Tripel können sowohl textuell in verschiedenen Serialisierungen (wie RDF/XML, Turtle, JSON-LD) als auch als gerichtete Graphen umgesetzt und dargestellt werden (vgl. Abb. 2).

Durch die Erweiterung *RDF Schema* (RDFS) werden begriffliche Basiskonzepte und -eigenschaften

²⁰ Zu diesen Themenkomplexen siehe die Standardwerke Allemang/Hendler (Anm. 12) und DuCharme, Bob: Learning SPARQL: Querying and Updating with SPARQL 1.1. 2. Aufl. O'Reilly 2013.

bereitgestellt, die elementare Aussagen über das zu modellierende Vokabular ermöglichen. Mit Hilfe von RDF-Syntax und RDFS-Vokabular lassen sich bereits *lightweight ontologies* wie Taxonomien modellieren: Konzepte können als Klassen und Unterklassen, Eigenschaften als Properties und Subproperties ausgewiesen und zueinander in Beziehung gesetzt werden.

Die *Web Ontology Language (OWL)*²¹ erweitert RDF und RDF Schema um komplexe sprachliche Konstrukte und ermöglicht logische Operationen unterschiedlicher Komplexität, abhängig vom verwendeten OWL-„Dialekt“ (*Lite, Description Logic DL, Full*). OWL stellt somit die wichtigste Beschreibungssprache zur Konzeption semantisch ausdrucksmächtiger Ontologien dar. Bedeutende OWL-Konzepte sind mögliche Kardinalitäts- und Werteinschränkungen (*restriction*) sowie mengenlogische Verknüpfungen von Klassen (UND/ODER-Beziehungen).

Die Menge aller Tripel wird als RDF Graph bezeichnet. Sie kann in spezialisierten Graphdatenbanken, sogenannten Triplestores²², gespeichert, verwaltet und verarbeitet werden. Dazu kommt häufig die standardisierte Anfragesprache SPARQL (*SPARQL Protocol and RDF Query Language*)²³ zum Einsatz, die komplexe Abfragen und Manipulationen von in RDF repräsentierten und strukturierten Daten ermöglicht. Von einigen, insgesamt aber noch zu wenigen Institutionen und Projekten werden bereits sogenannte SPARQL-Endpoints, also Zugriffspunkte auf ihre RDF-Datensätze, zur Verfügung gestellt.²⁴

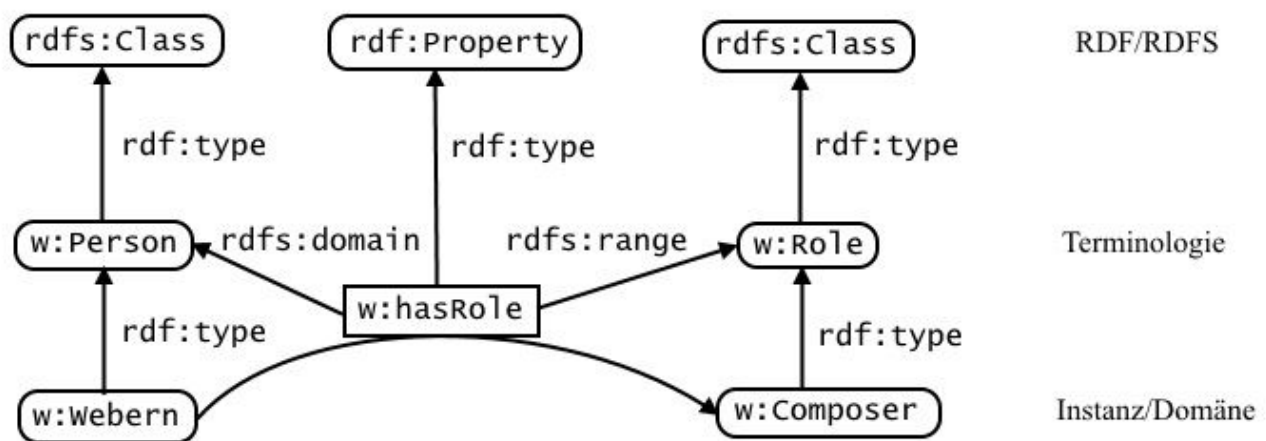


Abb. 2: Beispiel für einen RDF-Graph, der die Aussage „Webern hat die Rolle Komponist“ unter Annahme

²¹ Spezifikation der mittlerweile zweiten Version unter <https://www.w3.org/TR/owl2-overview/>.

²² Eine hilfreiche Übersicht (trotz letztmaliger Aktualisierung Anfang 2016) über existierende Triplestores und die ungefähre Größenordnung verarbeitbarer Datensätze findet sich auf der Seite <https://www.w3.org/wiki/index.php?title=LargeTripleStores&oldid=97382>.

²³ Es sei erwähnt, dass die vollständige SPARQL Spezifikation auch ein Übertragungsprotokoll für Suchanfragen und -ergebnisse sowie eine XML-Struktur für die Wiedergabe der Suchresultate umfasst, was hier nicht weiter betrachtet werden kann.

²⁴ Eine Übersicht über bestehende SPARQL Endpoints mit einigen interessanten Zugriffsmöglichkeiten auf musikalische Datensätze („BBC Programmes and Music“, „Jamendo“, „MusicBrainz“ und „DBTune“) findet sich unter: <https://www.w3.org/wiki/SparqlEndpoints>.

des fiktiven Namensraumes w: <<http://example.org/webern#>> abbildet, und die daran beteiligten konzeptuellen Ebenen (Instanzen, Terminologie und RDF Schema).

7 Qual der Wahl: erprobte Ontologien und Modelle

Umso breiter der Konsens, auf dem die verwendeten Konzepte, Begrifflichkeiten und Beschreibungen beruhen, desto wahrscheinlicher ist auch eine breite An- und Wiederverwendung des Modells. Generell empfiehlt es sich, nur im Fall nicht vorhandener oder in Bezug auf die eigene Fragestellung nicht adäquat anwendbarer Konzepte diese selbst zu modellieren, anderenfalls, wo immer möglich, auf bereits bestehende Ontologien zurückzugreifen. Angesichts einer schwer zu überblickenden und in einem Printmedium kaum noch sinnvoll darstellbaren Menge²⁵ können Übersichtsseiten wie *LODstats* (<http://stats.lod2.eu/vocabularies>), *Linked Open Vocabularies* (<http://lov.okfn.org/dataset/lov/>) oder *Open Metadata Registry* (<http://metadataregistry.org/>) beim Auffinden existierender Modelle hilfreich sein.

Einige der bekanntesten und auch am häufigsten verknüpften Modelle sind Friend-of-a-Friend (FOAF: <http://xmlns.com/foaf/spec/>), die Ontologien der Dublin Core Metadata Initiative (DCMI: <http://dublincore.org/specifications/>) sowie das *Simple Knowledge Organization System* (SKOS: <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>). Diese liefern mit ihren eher generischen *Top-level*-Konzepten Anbindungsmöglichkeiten zwischen verschiedenen Projekten.

Im Bereich des kulturellen Erbes hat sich das von einer breiten und weiterhin wachsenden Community getragene, äußerst ausdrucks mächtige *CIDOC Conceptual Reference Model* (CIDOC CRM: <http://www.cidoc-crm.org/>) etabliert, das sogar als ISO-Norm (ISO 21127:2006 bzw. 21127:2014) standardisiert ist. Speziell ist hier der ereignisbasierte Ansatz, der den Prozess bzw. das Ereignis, z.B. die Herstellung eines Objekts, anstatt das Objekt selbst in den Vordergrund rückt. Um den konzeptuellen Austausch zwischen Museum und Bibliothek zu vereinfachen, wurde an dieses mittlerweile auch ein bibliographisches Modell der IFLA, die *Functional Requirements for Bibliographic Records* (FRBR), durch die objektorientierte Erweiterung *FRBRoo* angepasst.²⁶ Mit BIBFRAME, einer RDF-basierten Initiative der Library of Congress, wurden die Voraussetzungen für eine Transformation der von den Bibliotheken bislang nach MARC-Standards aufgenommen Daten auf ontologiebasierte Strukturen gelegt.²⁷ Das *Europeana Data Model for Cultural Heritage* (EDM: <http://pro.europeana.eu/page/edm-documentation>) bietet mit seinem <aggregation>-Konzept einen essentiellen Ansatz zur Integration heterogener Datenquellen, z.B. verschiedene

²⁵ Das „Linking Open Data Cloud Diagram“ (Stand August 2017) enthält aktuell 1163 Datensätze. Siehe <http://lod-cloud.net/>.

²⁶ Spezifikation unter https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf. Vgl. zu Motivation und Vorteilen der Angleichung u.a. Becker, Hans-Georg: FRBR, Serials und CIDOC CRM – Modellierung von fortlaufenden Sammelwerken unter Verwendung von FRBRoo. In: Danowski/Pohl (Anm. 3) S. 64–96. <https://doi.org/10.1515/9783110278736.64>.

²⁷ Übersicht und Spezifikation unter <https://www.loc.gov/bibframe/>. Das BIBFRAME-Modell beruht insbesondere auf den Grundklassen <work>, <instance> und <item>.

Webressourcen oder Daten aus verschiedenen Projekten zu demselben Objekt (siehe Abb. 3).²⁸ U.a. wurden aus einem der ersten vollständig mit dem EDM modellierten Datensätze über 40.000 Objekte von MIMO (Musical Instrument Museums Online) in Europeana „aggregiert“.²⁹

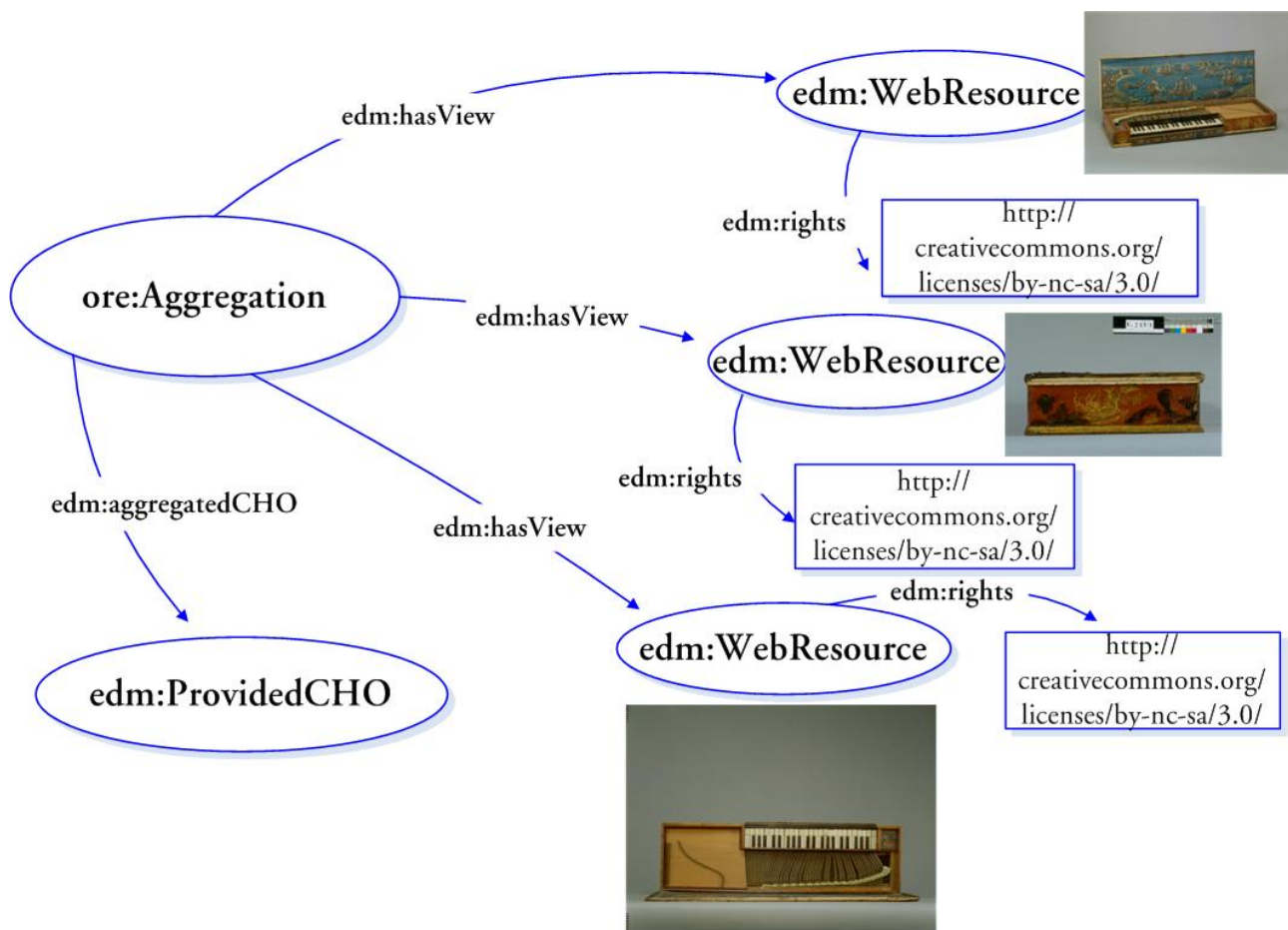


Abb. 3: Veranschaulichung des Aggregations-Modells des EDM am Beispiel eines MIMO-Datensatzes („Clavicorde lié dit de Lépante | Anonyme“). Quelle: <http://pro.europeana.eu/page/mimo-edm> (CC BY-SA 4.0).

Interessanterweise standen bereits in der allerersten LOD Cloud-Darstellung aus dem Jahr 2007 (<http://lod-cloud.net/versions/2007-05-01/lod-cloud.png>) drei von damals insgesamt nur zwölf Datenpools im Zusammenhang mit Musik (*MusicBrainz*, *DBtune* und *Jamendo*). Im gleichen Jahr wurde durch eine Forschungsgruppe am *Centre for Digital Music* der Queen Mary University of London ein auf OWL basierendes, speziell auf musikalische Sachverhalte zugeschnittenes Modell, die *Music Ontology* (<http://musicontology.com>; vgl. Abb. 4), konzipiert und zur Diskussion gestellt.³⁰ Inzwischen existiert ein

²⁸ Das Problem von Datenintegration und Heterogenitätskonflikten wird umfassend behandelt in Wache, Holger: *Semantische Mediation für heterogene Informationsquellen*. Dissertation, Berlin 2003.

<https://web.fhnw.ch/personenseiten/holger.wache/Papers/phd-03.pdf>. Vgl. auch Stuckenschmidt (Anm. 19) S. 211–232.

²⁹ Siehe <http://pro.europeana.eu/page/mimo-edm>. Die Klasse <Aggregation> entstammt dabei dem ORE-Vokabular der Open Archives Initiative (<http://www.openarchives.org/ore/1.0/terms>).

³⁰ Raimond, Yves; Abdallah, Samer; Sandler, Mark; Giasson, Frederick: *The Music Ontology*. In: *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Wien 2007, S. 417–422.

ganzes Ökosystem von erweiternden und ergänzenden Modellen (*timeline, event, keys, tonality, symbolic notation, chord, temperament, und audiofeatures*) um diese Ontologie. Sie muss mit ihren Ablegern als die derzeitige Standardontologie zur Modellierung und Repräsentation musikalischer Informationen gelten, die vor allem breite Anwendung bei der Konzeptualisierung von Musikempfehlungssystemen findet.³¹

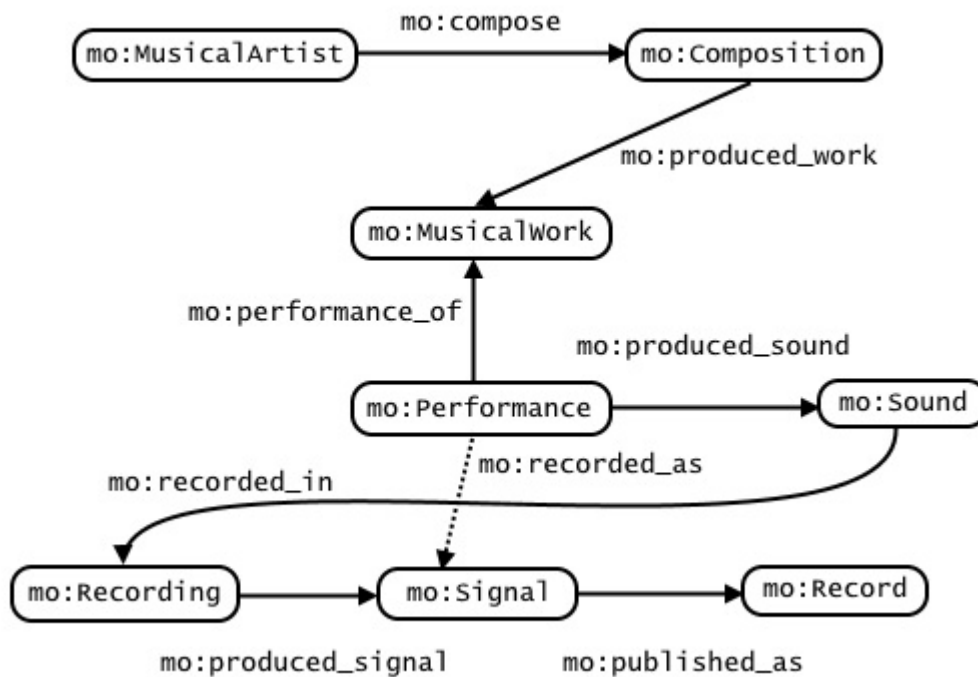


Abb. 4: Grundklassen der *Music Ontology* (mo) und ihre Beziehungen (Ausschnitt). Nach Sandler, Mark; Raimond, Yves; Sutton, Christopher: Interlinking Music-Related Data on the Web. In: IEEE MultiMedia 16(2) (2009) S. 54. doi:10.1109/MMUL.2009.29.

8 Musikwissenschaftliche Anwendungsbeispiele

Im Bereich der Musikbibliotheken gibt es bemerkenswerte Initiativen für semantische Pilotprojekte an der SLUB Dresden (*SLUBsemantics* und *D:SWARM*³²) und der BSB München, die u.a. einen SPARQL-Endpoint für die RISM-Daten initiiert hat und auch selbst hostet.³³ Die Deutsche Nationalbibliothek bietet ebenfalls seit

http://ismir2007.ismir.net/proceedings/ISMIR2007_p417_raimond.pdf. Vgl. auch die sehr viel ausführlichere Darstellung in Raimond, Yves: A Distributed Music Information System. Dissertation Queen Mary, University of London, November 2008. <http://raimond.me.uk/phd/thesis.pdf>. Die Entwicklung wurde zwischen 2011 und 2014 eingestellt. Vgl. das Github-Repositoryum: <https://github.com/motools>.

³¹ Vgl. Bailer, Werner; Boll, Susanne; Celma, Oscar, Hausenblas, Michael; Raimond, Yves: Use Case Scenarios (2.2 Music Use Case). In: Multimedia Semantics: Metadata, Analysis and Interaction, hg. von Raphael Troncy, Benoit Huet und Simon Schenk, Southern Gate 2011; sowie Chen, Y.-S.; Cheng, C.-H.; Chen, D.-R.; Lai, C.-H.: A Mood- and Situation-Based Model for Developing Intuitive Pop Music Recommendation Systems. In: Expert Systems 33(1) (2016) S. 77–91. doi: 10.1111/exsy.12132.

³² Siehe Mittelbach (Anm. 3).

³³ SPARQL-Ednpoint unter: <https://opac.rism.info/index.php?id=16&L=0>. Generell zum Open-Data-Konzept von RISM siehe: <http://www.rism.info/de/community/development/data-services.html>.

2010 einen Linked-Data-Service an, über den auch die Daten des Deutschen Musikarchivs in RDF bereitgestellt werden sollen.³⁴ Allerdings ist für eine großflächige Öffnung der deutschsprachigen Bibliothekslandschaft „abgesehen von [diesen] mutigen Vorstößen“ immer noch „zu wenig praktische Aktivität“ zu verzeichnen.³⁵

Ähnliches gilt für die deutschsprachige Forschungslandschaft. So nimmt es kaum Wunder, dass zwei große internationale, kollaborative Vorhaben auf diesem Gebiet weit voraus sind und zu den Vorreitern zählen: Zum einen das *Transforming-Musicology*-Projekt in Großbritannien, das in vielzähligen Teilprojekten „the enhancement of Semantic Web provisions for musical study [...] augmenting existing controlled vocabularies (known as ontologies) for musical concepts“³⁶ vorantreibt. Unter anderem wird mit *MELD* (Music Encoding and Linked Data) das Zusammengehen von MEI-Codierung und *Semantic-Web*-Technologien erprobt.³⁷

Zum anderen in Frankreich das *DOREMUS*-Projekt (*Doing Reusable Musical Data*), das „tools and methods to describe, publish, connect and contextualize music catalogues on the web of data“³⁸ entwickelt. Gerade das Modell von *DOREMUS* ist sehr eindrücklich durch seine Verbindung und Nachnutzung von *Music Ontology*, CIDOC CRM, FRBRoo und *Europeana Data Model*:

³⁴ Vgl. den Beitrag „Linked Data und das Deutsche Musikarchiv“ von Jana Hentschke und Simon Zetzsche auf der Jahrestagung der AIBM, 4.–8. September 2017. Nähere Informationen zum Linked Data Service der DNB unter http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkedata_node.html.

³⁵ Mittelbach (Anm. 3) S. 68.

³⁶ Projektbeschreibung von *Transforming Musicology* unter <http://www.transforming-musicology.org/about/>.

³⁷ Demonstrator unter: <http://www.semanticaudio.ac.uk/demonstrators/24-music-encoding-and-linked-data-meld/> bzw. <https://github.com/oerc-music/meld>.

³⁸ Projektbeschreibung von *DOREMUS* unter http://www.doremus.org/?page_id=18. Eine Übersicht der von *DOREMUS* erstellten und verwendeten kontrollierten musikbezogenen Vokabulare unter: <https://github.com/DOREMUS-ANR/knowledge-base/tree/master/vocabularies>.

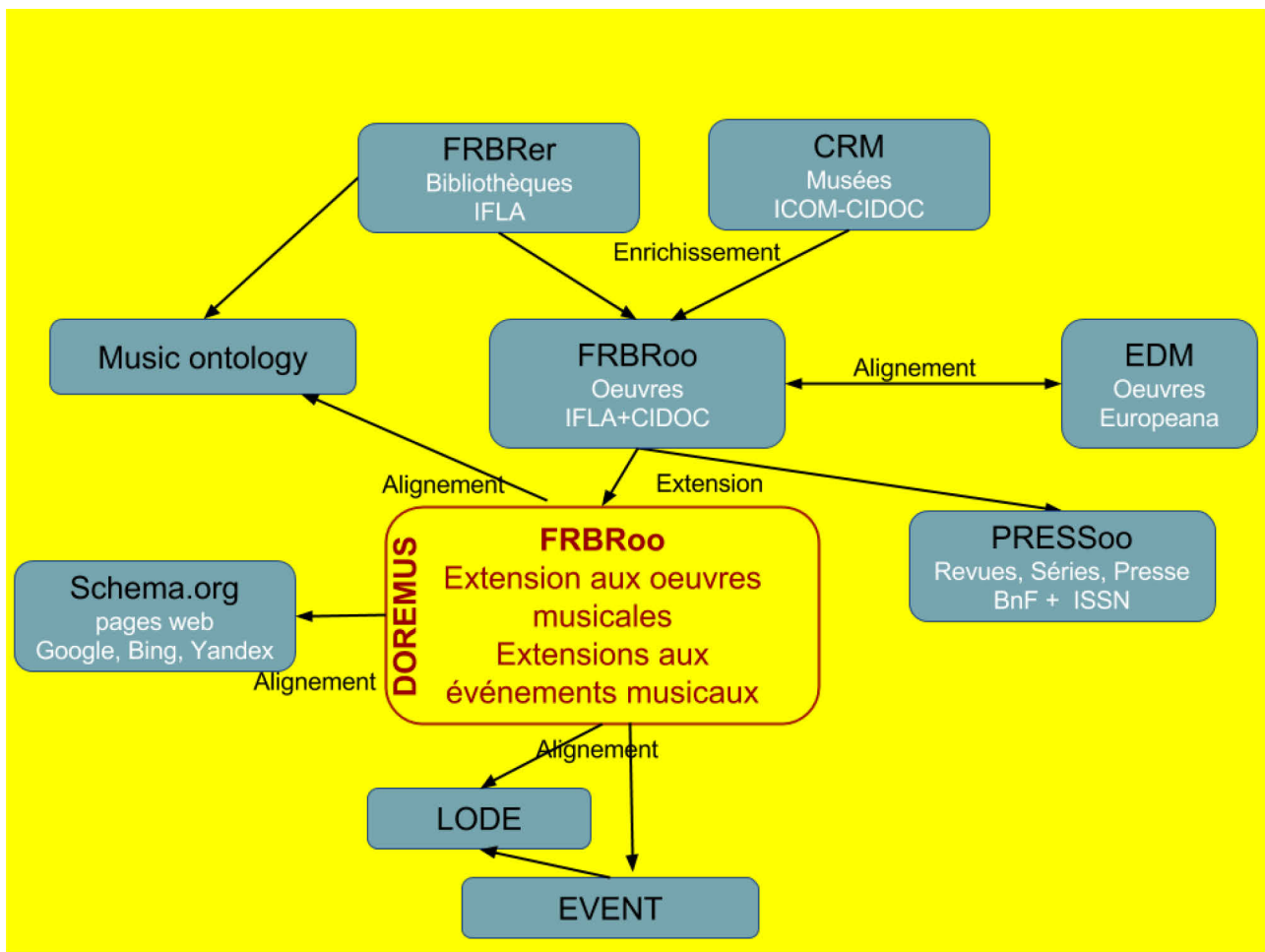


Abb. 5: Das Zusammenspiel verschiedener Ontologien im DOREMUS-Projekt. Quelle: http://www.doremus.org/?page_id=30.

Das finanzierungsfrei, aber mit großem Engagement durchgeführte Projekt *JazzCats* (Jazz Collection of Aggregated Triples)³⁹ verknüpft drei bestehende Datensätze, die Informationen zu jazzbezogenen Aufführungen, Aufnahmen und Personen enthalten (*Body & Soul*, *WJazzD* und *LinkedJazz*), über Konzepte der *Music Ontology* und macht diese über einen eigenen SPARQL-Endpoint zugänglich.

Im deutschsprachigen Raum wird für die am Max-Planck-Institut für empirische Ästhetik vorbereitete „Fachgeschichte der deutschsprachigen Musikwissenschaft von ca. 1810 bis ca. 1990“ die Verwendung ontologiebasierter Datenbankstrukturen diskutiert.⁴⁰

Zudem bauen zwei prominente Virtuelle Forschungsumgebungen auf Ontologiemodellen auf: die in Erlangen/Nürnberg entwickelte Wissenschaftliche Kommunikationsinfrastruktur *WissKI* (www.wiss-ki.eu) und das an der Universität Basel entwickelte *Knora/SALSAH* (www.knora.org/www.salsah.org).

WissKI nutzt eine eigene, auf OWL-DL zugeschnittene Umsetzung der CIDOC CRM Spezifikation, das

³⁹ Bangert, Daniel; Nurmikko-Fuller, Terhi; Abdul-Rahman, Alfie: *JazzCats project* (2016). <http://jazzcats.oerc.ox.ac.uk/>.

⁴⁰ Siehe van Dyck-Hemming, Annette; Wald-Fuhrmann, Melanie: Vom Datum zum historischen Zusammenhang. Möglichkeiten und Grenzen einer fachgeschichtlichen Datenbank. In: *Wissenskulturen der Musikwissenschaft. Generationen – Netzwerke – Denkstrukturen*, hg. von Sebastian Bolz, Moritz Kelber, Ina Knoth & Anna Langenbruch. Bielefeld 2016, S. 261-278.

sogenannte Erlangen CRM / OWL (<http://erlangen-crm.org/>). Das Projekt „Musikinstrumente sammeln – Das Beispiel Rück“ arbeitet die Korrespondenz rund um die musikhistorische Instrumentensammlung von Wilhelm Rück und Söhnen innerhalb von *WissKI* auf und stellt seine Forschungsdaten darüber zur Verfügung.⁴¹

Knora/SALSAH beruht auf einem Set von OWL konformen Ontologien (*Knora Ontologies*), einer Triplestore-Schnittstelle (*Knora API Server*), einem IIF-kompatiblen Medienserver und der graphischen Benutzeroberfläche SALSAH. Die Anton Webern Gesamtausgabe macht ihre gesamte Forschungsdatenbank über SALSAH zugänglich und wird die Bestandteile ihrer Online-Editionen über eine durch die Schweizer *Nationale Infrastruktur für Editionen* (NIE-INE) implementierte Instanz von Knora/SALSAH organisieren und aufbereiten.⁴²

9 Ausblick

Das Potenzial von Ontologien und semantischen Technologien steht außer Frage. Dennoch spielten sie in den Diskussionen um eine digitale Musikwissenschaft im deutschsprachigen Raum bislang keine bzw. eine eher verminderte Rolle. Die Gründe dafür sind vielfältig: Zum einen mag es an einer gewissen generellen Verspätungstendenz der Musikwissenschaft liegen, die ihr aufgrund der Komplexität ihres Untersuchungsgegenstandes aber auch gestattet sei. Mitverantwortlich sind aber sicher auch eine bislang unzureichende Infrastruktur und mangelnde generische Tools, die den Musikinstitutionen (und hier seien sämtliche Gedächtnisinstitutionen einschließlich Forschung, Editionen, Verlagen und Konzerthäusern gemeint) eine Hinwendung zu diesen relativ neuen Technologien erleichtern würden. Denn wenn es auch prinzipiell möglich ist, eine Ontologie auf Papier niederzuschreiben, werden mit steigender Komplexität, Ausdrucksmächtigkeit und Verknüpfungstiefe zunehmend unterstützende Systeme notwendig, die die Modellierung und Verwendung von Ontologien erleichtern, Workflows systematisieren, kollaboratives Arbeiten ermöglichen und sich wiederholende Prozesse (semi-)automatisieren. Spezialisierte virtuelle Forschungsumgebungen bieten dazu einen vorstellbaren Ansatz. Ein weiterer sind Forschungsverbünde, Akademien, Bibliotheksverbünde, nationale Infrastrukturen, unter welchem Namen auch immer ein Zusammenspiel von Projekten, Bibliotheken, Museen, Archiven und auch Verlagen gebündelt wird. Denn Ontologien sind *per definitionem* kollaborativ, die Herausforderungen nicht im Alleingang, nur gemeinschaftlich zu bewältigen.

Sollen also musikalisches Experten- und Fachwissen kontrolliert formalisiert und musikwissenschaftliche Wissensstrukturen explizit maschinenverarbeitbar gemacht werden, und sollen dabei kritisch-differenzierte Sichtweisen und wissenschaftliche Perspektiven in ein solches Wissens- und Datengeflecht einfließen, so ist

⁴¹ Siehe die Projektbeschreibung unter <http://www.gnm.de/forschung/forschungsprojekte/musikinstrumente-sammeln-das-beispiel-rueck/>.

⁴² Siehe www.anton-webern.ch bzw. einen vorläufigen Editionsprototypen unter edition.anton-webern.ch.

die 2014 von Jörg Wettlaufer aufgeworfene Frage, ob die Hinwendung zu Ontologien und semantischen Technologien der „nächste Schritt“⁴³ sei, durchaus zu bejahen. Der „Tipping Point“ aber, die Schwelle, ab der eine neue Technologie zündet und sich unumkehrbar durchsetzt,⁴⁴ steht in diesem Zusammenhang aus Sicht der (digitalen) Musikwissenschaft möglicherweise noch etwas weiter am Horizont als in anderen Bereichen, aber er ist immerhin sichtbar. In diesem Sinne, *caveat auditor*.



Stefan Münnich

Musikwissenschaftliches Seminar der Universität Basel

Petersgraben 27/29

CH-4051 Basel

Schweiz

stefan.muennich@unibas.ch

⁴³ Siehe Wettlaufer (Anm. 5).

⁴⁴ Universelle Mechanismen und Voraussetzungen für das Überschreiten eines solchen neuralgischen Punkts will Malcolm Gladwell erkannt haben in: *The Tipping Point. How Little Things Can Make a Big Difference*. New York 2000.