

Stefan Münnich (Basel)

## Ontologien in der Praxis: Möglichkeiten und Herausforderungen für die Modellierung musikwissenschaftlicher / musikeditorischer Wissensstrukturen\*

Der im Rahmen der Jahrestagung der Gesellschaft für Musikforschung 2017 begründeten Fachgruppe „Digitale Musikwissenschaft“ ist mit ihrem ersten in dieser Funktion ausgerichtetem Symposium zum Thema „Wissenssystematiken im digitalen Zeitalter“ ein sehr feines Gespür für eine Problematik zu attestieren, der in der deutschsprachigen Musikwissenschaft bislang wenig bis kaum Beachtung geschenkt wurde, obwohl sie in anderen Bereichen und auf internationaler Ebene durchaus seit mehr als 15 Jahren diskutiert wird.<sup>1</sup> In Zeiten von „Fake News“, der Schnelllebigkeit von Medien und immer kürzerer Halbwertszeiten von Informationen stellen die (maschinell verarbeitbare) Systematisierung und Klassifizierung, die Aufbereitung, Verbreitung, Einordnung und Überprüfbarkeit von Wissen einige der drängendsten gesamtgesellschaftlichen Herausforderungen dieses noch jungen digitalen Zeitalters dar.<sup>2</sup> „Alternative Fakten“ können sich global in rasanter Geschwindigkeit verbreiten und ungeahnte Zustimmung erfahren aufgrund der Latenzzeit der notwendigen Gründlichkeit ihrer Gegenprüfung. Das Gerücht ist sprichwörtlich immer größer – und schneller – als die Wahrheit. Und während eine Halbwahrheit schnell ausgesprochen und verbreitet ist, bedarf es manchmal einiger Jahre, wenn nicht Jahrzehnte oder mehr, Daten zusammenzutragen und derart zu Informationen zu strukturieren und semantisch aufzubereiten, dass so etwas wie überprüfbares Wissen aus der Verknüpfung und Kontextualisierung dieser Informationen sich formieren kann.<sup>3</sup> Und selbst dann ist noch nichts über einen

\* Mein Dank gilt Hans Cools, Terhi Nurmikko-Fuller, John Pybus, Dominic Oldman und Barbara Wiermann für Anregungen und Diskussionen zur Thematik dieses Beitrags und ebenso Thomas Ahrend, auch für seine kritische, stets hilfreiche Durchsicht und Kommentare.

- 1 Vgl. Stefan Münnich, „Ontologien als semantische Zündstufe für die digitale Musikwissenschaft? Eine Bestandsaufnahme“, in: *Bibliothek – Forschung und Praxis* 42/2 (2018), S. 184–193. Online unter <<https://doi.org/10.1515/bfp-2018-0027>>. Grundlegend zum Thema Wissenssystematiken vgl. Theo Stammen und Wolfgang E. J. Weber (Hrsg.), *Wissenssicherung, Wissensordnung und Wissensverarbeitung. Das europäische Modell der Enzyklopädien* (= Colloquia Augustana 18), Berlin 2004; darin auch Markus Bandur, „Stichwort ‚Musik‘. Zur Problematik der Darstellung von Fachwissen in Enzyklopädien des 18. Jahrhunderts“, S. 233–244. Online unter <<http://nbn-resolving.de/urn:nbn:de:bsz:25-freidok-23939>>. Zum Zusammenhang von Digitalität und Wissensmodellierung vgl. Reinhard Bauer, *Die digitale Bibliothek von Babel. Über den Umgang mit Wissensressourcen im Web 2.0*, Boizenburg 2010; Peter Haber, *Digital Past. Geschichtswissenschaft im digitalen Zeitalter*, München 2011 sowie David Weinberger, *Das Ende der Schublade. Die Macht der neuen digitalen Unordnung*, München 2008.
- 2 Peter Stadler hat dies im Schwerpunktsymposium „Musikverlagswesen: gestern–heute–morgen“ auf der GfM-Tagung 2017 in seinem Vortrag zu Open-Access- und Online-Publikation sehr deutlich auch als eine „zentrale Gestaltungsaufgabe“ des Faches Musikwissenschaft bezeichnet.
- 3 Hinter dieser hilfreichen Trennung von Daten („Sachverhaltsbeschreibungen“), Informationen („interpretierte, zweckbezogene Daten“) und Wissen („begründete, miteinander in Beziehung gesetzte Informationen“) steht ein organisationstheoretisches Pyramidenmodell (vgl. Klaus Fuchs-Kittowski, „Wissens-Ko-Produktion. Verarbeitung, Verteilung und Entstehung von Informationen in kreativ-lernenden Organisationen“, in: *Stufen zur Informationsgesellschaft. Festschrift zum 65. Geburtstag von*

möglichen Wahrheitsgehalt oder eine wie auch immer geartete „Objektivität“ der in einem solchen Wissensnetzwerk kontextualisierten Informationen ausgesagt – nicht umsonst stellt der Zusammenhang von Wissen und Wahrheit eines der ältesten und kontroversesten Themen im menschlichen Nachdenken dar, und nicht umsonst ist die schon bei Platon zu findende Idee von Wissen als „wahrer, gerechtfertigter Meinung“ wiederholt in Frage gestellt worden (besonders prominent im sogenannten „Gettier-Problem“<sup>4</sup>). Und dass Wissen keineswegs objektiv, sondern auch immer im Wechselspiel mit Machtstrukturen, -verteilungen und -allianzen zu sehen ist, ist zwar schon seit Francis Bacon Gemeinplatz und wurde in der 2. Hälfte des 20. Jahrhunderts nochmals z. B. von Adorno, Foucault oder Bruno Latour expliziert. Im digitalen Kontext aber erhält diese Thematik neuerliche Brisanz, wenn z. B. algorithmische Lernprozesse, wie sie im Bereich von Machine Learning oder Künstlicher-Intelligenz-Forschung erprobt werden, nicht mehr nachvollziehbare Blackboxen erzeugen<sup>5</sup> oder auch die zum Großteil in Sprache und Begriffen verankerten menschlichen Vorurteile (Rassismus, Sexismus) sich in Algorithmen, Modellen und Künstlicher Intelligenz projiziert wiederfinden.<sup>6</sup>

Was notwendig ist, sind also Nachprüfbarkeit und Vertrauen, ist „belastbares Wissen“, wie es Laurenz Lütteken im Zusammenhang mit dem einjährigen Bestehen der *MGG Online* während der GfM-Tagung 2017 formuliert hat. Allerdings muss man sich immer wieder klarmachen, dass Wissen nicht per se belastbar *ist*, sondern dass es zu einem bestimmten Zeitpunkt im Kontext einer gemeinschaftlichen Verabredung einer bestimmten Community als belastbar gelten *kann*. Belastbarkeit ist kein einmalig erworbenes, permanentes Prä-

---

*Klaus Fuchs-Kittowski*, hrsg. von Christiane Floyd, Christian Fuchs und Wolfgang Hofkirchner, Frankfurt 2002, S. 59–125, hier S. 81). Wissen ist dabei stets ein nicht abzuschließender, vorläufiger Zustand und zugleich auf sich selbst bezogener Prozess. Es verändert sich mit jeder neu in das Netzwerk eintretenden Information. Das von Gilles Deleuze und Félix Guattari in den 1970er Jahren geprägte Bild des (wuchernden) Rhizoms verdeutlicht sehr prägnant diese permanenten Transformationsprozesse.

- 4 Nach dem US-amerikanischen Philosophen Edmund Gettier. Vgl. dessen kurzen, aber folgeschweren Aufsatz „Is Justified True Belief Knowledge?“, in: *Analysis* 23/6 (1963), S. 121–123. Online unter <<https://doi.org/10.1093/analys/23.6.121>>.
- 5 Führende Experten im Bereich der Artificial Intelligence-Forschung warnen davor, selbstlernende Algorithmen sich selbst zu überlassen und eine neue „digitale Alchemie“ zu begünstigen (vgl. auch Matthew Hutson, „Has Artificial Intelligence Become Alchemy?“, in: *Science* 360/6388 [4. Mai 2018], S. 478. Online unter <<https://doi.org/10.1126/science.360.6388.478>>). So sind z. B. die konkreten Lernschritte, die Googles DeepMind-System *AlphaGo* bei seinem Sieg über den 18fachen Go-Weltmeister Lee Sedol durchlief, aufgrund ihrer Komplexität größtenteils unbekannt. Gleichzeitig entstehen neue Forschungszweige wie „AI Neuroscience“, die versuchen derartige Verstehenslücken zu schließen (vgl. Paul Voosen, „The AI Detectives“, in: *Science* 357/6346 [7. Juli 2017], S. 22–27. Online unter <<https://doi.org/10.1126/science.357.6346.22>>). Vgl. auch Theo Röhle, *Der Google-Komplex. Über Macht im Zeitalter des Internets*, Bielefeld 2010.
- 6 Vgl. Aylin Caliskan / Joanna J. Bryson / Arvind Narayanan, „Semantics Derived Automatically from Language Corpora Contain Human-Like Biases“, in: *Science* 356/6334 (14. April 2017), S. 183–186. Online unter <<https://doi.org/10.1126/science.aal4230>>. Initiativen wie *Whose Knowledge?* (<<https://whoseknowledge.org/>>) zeigen zudem auf, dass der Großteil des heutzutage online verfügbaren Wissens von weißen Männern aus Europa oder Nordamerika ins Netz gestellt wird, obwohl etwa die Hälfte der Internet-Nutzer Frauen sind und drei Viertel der Internet-Nutzer aus Ländern des „Globalen Südens“ stammen. Daten: International Telecommunication Union (ITU) der Vereinten Nationen, *ICT Facts and Figures 2017*; <<https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx>>. Allgemein vgl. auch Nicole Zillien, *Digitale Ungleichheit. Neue Technologien und alte Ungleichheiten in der Informations- und Wissensgesellschaft*, Wiesbaden 2009.

dikat, sondern ein temporäres, historisches, immer wieder neu zu prüfendes. Erst Wissen, das sich dieser Prämissen bewusst ist, das dennoch dem Test der Nachprüfbarkeit wiederholt standhält und dadurch das Vertrauen der beteiligten Akteure gewinnt bzw. genießt, kann in diesem Sinne als belastbar gelten.

Bemerkenswerterweise spielen ebenfalls in der von Tim Berners-Lee und anderen Anfang der 2000er Jahre formulierten Vision eines *Web of Data* (auch *Semantic Web*<sup>7</sup>), also eines Web, in dem nicht mehr Dokumente, sondern Daten semantisch verknüpft und verflochten sind, die Verlässlichkeit von Daten und Informationen eine prominente Rolle.<sup>8</sup> Die angesprochene Systematisierung und Modellierung von Wissensstrukturen bildet dabei gleichsam das Herzstück, den zentralen Block, in dem mit Hilfe von sogenannten Ontologien Informationen formalisiert, strukturiert, kontextualisiert und semantisch vernetzt werden.

Um zu sehen, wie solche semantischen Geflechte – nichts anderes sind im informationstheoretischen Kontext „Ontologien“ – im Bereich musikwissenschaftlicher Fragestellungen zur Modellierung, Beschreibung und Transformation belastbaren Wissens eingesetzt werden können, sollen im Folgenden in einem ersten Teil grundlegende Begrifflichkeiten, Mechanismen sowie Herausforderungen betrachtet werden, bevor in einem zweiten Teil konkrete Anwendungsfälle zur Sprache kommen.

### *Ontologien als Wissensmodelle*

Nach der gängigsten und weitverbreitetsten Definition ist eine im informationstechnischen Sinn verstandene Ontologie „an explicit, formal specification of a shared conceptualisation“<sup>9</sup>, eine explizite und formale Spezifikation eines gemeinsam genutzten und verstandenen Konzepts. „Conceptualisation“ bezeichnet hierbei ein abstraktes Modell eines Weltausschnitts und „specification“ die Beschreibung dieses Modells. Mit „explicit“ wird eine eindeutige Definition der Bedeutung dieser Konzepte gefordert, mit „formal“ die Maschinenverständ-

- 
- 7 Tim Berners-Lee / James Hendler / Ora Lassila, „The Semantic Web“, in: *Scientific American* 284/5 (17. Mai 2001), S. 34–43. Online unter <<http://www.jstor.org/stable/26059207>>. Zuvor in Tim Berners-Lee, *What the Semantic Web Can Represent*. September 1998. Online unter <<https://www.w3.org/DesignIssues/RDFnot.html>>. Zur Frage der Verknüpfung von Semantic Web-Ansätzen mit speziell geisteswissenschaftlichem Fachwissen vgl. Dominic Oldman / Martin Doerr / Stefan Gradmann, „Zen and the Art of Linked Data: New Strategies for a Semantic Web of Humanist Knowledge“, in: *A New Companion to Digital Humanities*, hrsg. von Susan Schreibman, Ray Siemens und John Unsworth, Southern Gate 2016, S. 251–273. Online unter <<https://doi.org/10.1002/9781118680605.ch18>>.
- 8 Dass dies konkrete Auswirkungen auch auf viele Lebens- und Wissenschaftsbereiche hat, zeigt die Diskussion um eine „data-driven history“ (vgl. Markus Heidmeier, „Das Ende der Theorie: Data Driven History“, Blog-Eintrag vom 5. Oktober 2011. Online unter <<https://blog.zeit.de/open-data/2011/10/05/data-driven-history/>>); auch in den Editionswissenschaften lassen sich Tendenzen hin zu einer Edition von Daten beobachten, vgl. zuletzt Robert S. Kamzelak, „Von der Raupe zum Schmetterling oder Wie fliegen lernen – Editionsphilologie zwischen Infrastruktur und Semantic Web“, in: *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*, hrsg. von Roland S. Kamzelak und Timo Steyer, 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften 2). Online unter <[https://doi.org/10.17175/sb002\\_004](https://doi.org/10.17175/sb002_004)>.
- 9 Rudi Studer / Richard Benjamins / Dieter Fensel, „Knowledge Engineering: Principles and Methods“, in: *Data & Knowledge Engineering* 25/1–2 (1998), S. 161–197, hier S. 184. Vgl. auch die fast gleichlautende, aber allgemeinere Definition bei Thomas Gruber, „A Translation Approach to Portable Ontology Specifications“, in: *Knowledge Acquisition* 5/22 (1993), S. 199–220, hier S. 199. Online unter <<https://doi.org/10.1006/knac.1993.1008>>.

lichkeit bzw. Maschinenverarbeitbarkeit und „shared“ weist darauf hin, dass es sich um von einer möglichst großen Gemeinschaft getragene und verstandene, also konsensfähige Konzepte handeln sollte. Ontologien beschreiben also Konzepte und ihre Beziehungen zueinander, indem sie menschliches Wissen regelbasiert veranschaulichen, modellieren und es einer maschinellen Speicherung, Auswertung und Nachnutzung zugänglich machen.

Da diese Definition (bewusst) recht weit gefasst ist, versammeln sich unter ihr vielfältige Arten formaler Begriffssysteme und Wissenssystematiken. Verschiedene Klassifizierungsmodelle erlauben es, eine gewisse Systematik in diese Vielfalt zu bringen: Eine von Nicola Guarino vorgeschlagene Einteilung richtet sich nach der Spezifik des jeweiligen Anwendungsbereiches, d. h. generischere „top-level ontologies“ (mit Konzepten wie Zeit, Raum, Ereignis) stehen spezifischeren „application ontologies“ (z. B. Rollen, die während einer bestimmten Aktivität eingenommen werden) gegenüber sowie gegenstandsbezogene „domain ontologies“ (z. B. Musik, Medizin) den aufgabenbezogenen „task ontologies“ (z. B. Spielen eines Instruments, Aufnehmen von Musik; vgl. Abb. 1). In vielen Anwendungsfällen lässt sich eine solch klare Trennung allerdings nicht aufrechterhalten, da oft eine Verflechtung verschiedener Ebenen stattfindet.

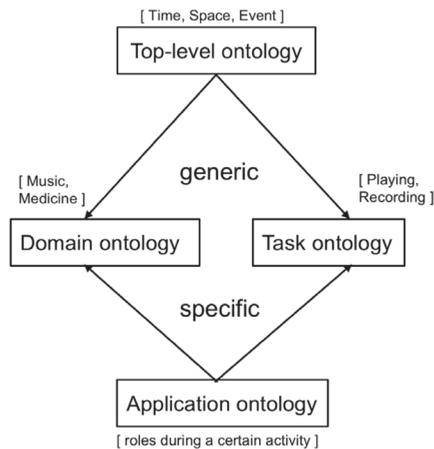


Abbildung 1: Klassifizierung von Ontologien (nach: Nicola Guarino, „Formal Ontology and Information Systems“, in: *Proceedings of FOIS* (1998). Online unter <<http://www.loa.istc.cnr.it/old/Papers/FOIS98.pdf>>)

Das Modell von Ora Lassila und Deborah McGuinness nimmt dagegen den Grad des semantischen Ausdrucks als Ordnungskriterium, so dass sich ein semantisches Spektrum ergibt von Verzeichnissen und Katalogen mit kontrollierten Vokabularen und Taxonomien (lightweight ontologies) über Thesauri hin zu semantisch komplexeren Formalisierungen, die die Anwendung logischer Operationen sogar höherer Ordnung erlauben (heavyweight ontologies).<sup>10</sup> Im engeren Sinne werden häufig erst oberhalb von Thesauri klassifizierte Formalisierungen, also auf formalen Beschreibungssprachen wie RDF (Resource Description

10 Vgl. Ora Lassila / Deborah L. McGuinness, „The Role of Frame-Based Representation on the Semantic Web“, in: *Knowledge Systems Laboratory Report KSL-01-02*, Stanford University, 2001; auch in: *Linköping Electronic Articles in Computer and Information Science 6/5* (2001). Online unter <<http://www.ep.liu.se/ea/cis/2001/005/>>.

Framework), RDF Schema oder OWL (Web Ontology Language) beruhende Modelle, als eigentliche „Ontologien“ bezeichnet.

Interessanter und wichtiger als die Frage der Klassifizierung ist allerdings die Frage, wie Ontologien helfen können, Wissen strukturiert zu formalisieren und zu modellieren. Dies geschieht durch den „Kunstgriff“, eine Trennung von abstrakter Terminologie (terminological knowledge), individuellen Zuweisungen (assertional knowledge) sowie logischen Ableitungen (inferred knowledge) vorzunehmen. In den resultierenden drei Säulen eines Wissensmodells (vgl. Tab. 1) wird so die Frage, mit welchen Begriffen man über einen betrachteten Weltausschnitt spricht (Terminologie), separiert von den Fragen, welche konkreten Behauptungen über diesen Weltausschnitt mittels der Terminologie explizit formulierbar sind (Zuweisungen), und welche Schlussfolgerungen implizit daraus abgeleitet werden können (Inferenz).

<b>Terminologisches Wissen</b>	<b>Assertionales (explizites) Wissen</b>	<b>Inferiertes (implizites) Wissen</b>
„Wie, also mit welchen Begriffen, spreche ich über den betrachteten Weltausschnitt?“	„Welche konkreten Behauptungen über den betrachteten Weltausschnitt kann ich mittels meiner Terminologie treffen?“	„Welche Schlußfolgerungen können aus dem Modell (automatisiert) abgeleitet werden?“
<ul style="list-style-type: none"> <li>- Klassen (classes, entities) = abstrakte Konzepte</li> <li>- Eigenschaften (attributes, properties) spezifizieren Klassen und verweisen auf Datenwerte (literals) oder andere Klassen (relations)</li> <li>- Grundannahmen (axioms)</li> <li>- Einschränkungen (constraints, restrictions)</li> </ul>	<ul style="list-style-type: none"> <li>- Instanzen/Individuen</li> </ul>	<ul style="list-style-type: none"> <li>- machine reasoning</li> </ul>

Tabelle 1: Strukturierung von Wissen mit Hilfe von Ontologien

Im Bereich des terminologischen Wissens werden Konzepte, also reale oder fiktive Subjekte und Dinge eines betrachteten Weltausschnitts (wie Person, Ort, Manuskript), als sogenannte „Klassen“ abstrakt gebündelt. Dazu werden gemeinsame Merkmale bestimmt, die wiederum als Attribute diesen Klassen zugeordnet sind und die verschiedenen Klassen untereinander unterscheidbar machen. (Klassen werden dabei häufig als Knoten in einem Graphen dargestellt, Attribute als deren Kanten.) Die Attribute/Kanten können dabei auf einfache Datenwerte (sog. literals) oder auch auf andere Klassen als Relation verweisen, wobei letzteres für eine Verknüpfung der Klassen untereinander sorgt, aus der sich letztendlich die Netzwerkstruktur, das „Rhizomartige“, das Geflecht des Modells speist. Im Bereich des terminologischen Wissens können zusätzlich formale Definitionen von Grundannahmen (axioms) oder Einschränkungen (constraints) vorgenommen werden, die zu einer logisch konsistenten Verwendung von Attributen und Eigenschaften beitragen. Durch die Zuweisung von modellierten Individuen, sogenannten Instanzen, zu den abstrakten Klassen lassen sich ganz spezifische Aussagen über ein konkretes Mitglied einer Klasse und über dessen konkrete Eigenschaften und Charakteristika treffen.

Semantisch komplexere Ontologien schöpfen ihr volles Potential aber eigentlich erst dann aus, wenn aus dem explizit modellierten (und bekannten) Wissen zusätzlich implizites (und möglicherweise neues) Wissen abgeleitet werden kann. Zu diesem Zweck kommen oftmals

sogenannte „machine reasoner“<sup>11</sup> zum Einsatz, also Programme, die auf Basis der dem Modell zugrundeliegenden Strukturen (vor allem der Grundannahmen und Einschränkungen) automatisierte logische Schlussfolgerungen ermöglichen. Bei der Überprüfung der Konsistenz von terminologischem und assertivem Wissen spielt dies u. a. eine wichtige Rolle.

### *Herausforderungen bei der Modellierung von Wissen*

Jede Modellierung repräsentiert nicht einfach nur (im Sinne Platons) Wissen in formalisierter Weise, sie transformiert, verändert es zugleich. Durch die Einschreibung in einen neuen (formalisierten) Kontext wird Wissen möglich, das vorher nicht verfügbar, nicht artikulierbar, nicht denkbar war.<sup>12</sup> Gleichzeitig sind die verwendeten terminologischen Begrifflichkeiten und deren Verknüpfungen sowie die Aussagen über konkrete Individuen stets subjektiven Zuweisungsakten unterworfen, wie breit der Konsens, auf dem sie beruhen, auch sein mag. Es gibt deswegen nicht den einen „richtigen“ Weg, Wissensstrukturen zu modellieren. Der Untersuchungsgegenstand, die darauf anzuwendenden Fragestellungen sowie die Perspektive, die der Untersuchende diesen gegenüber einnimmt, sind hier zu berücksichtigende Faktoren. Laut einer Untersuchung von Randall Davis, Howard Shrobe und Peter Szolovits sind jegliche Wissensrepräsentationen, und somit auch Ontologien, nur ein reduziertes, vereinfachtes und unvollkommenes Modell des betrachteten Weltausschnitts, ein „Surrogat“ der Realität. Sie sind stets von einer bestimmten, mehr oder weniger unvollkommenen Sicht- und Herangehensweise geprägt und können nicht allumfassend oder abschließend sein.<sup>13</sup> Mit dem Bewusstsein dieser unvermeidlichen Reduktion und unter Berücksichtigung des personellen, zeitlichen und finanziellen Aufwands, der für die Wissensmodellierung beträchtlich und nicht zu unterschätzen ist, kann das Ziel einer Modellierung daher auch immer nur eine hinreichend genaue, zwar kritisch reflektierte, aber dennoch den praktischen Aufwand vertretbare Annäherung an den betrachteten Ausschnitt der Realität sein, nach dem Motto: so kritisch wie möglich, aber auch so praktisch wie nötig.

„Realität“ ist in diesem Zusammenhang natürlich ein nicht unproblematischer Begriff: Zum einen konnten auf physikalischer Ebene Forscher der Australian National University 2015 nachweisen, dass eine Aussage über das Verhalten nicht nur von Photonen (Licht), sondern auch Atomen (Materie) erst nach ihrer Beobachtung, ihrer Messung möglich ist.<sup>14</sup>

11 Vgl. Uli Sattler / Robert Stevens / Phillip Lord, „How Does a Reasoner Work?“, in: *Ontogenesis* (12. August 2014). Online unter <<http://ontogenesis.knowledgeblog.org/1486>>.

12 Vgl. in diesem Zusammenhang Markus Bandur, „Musikalisches Wissen und seine Geschichte als lexikographische Herausforderung. Das Handwörterbuch der musikalischen Terminologie“, in: *Fontes Artis Musicae* 63/3 (2016), S. 179–191. Online unter <<https://muse.jhu.edu/article/625699>>: „[...] dass die begriffslose Materie der Musik sich geschichtlich nicht primär durch eine in den Sachen liegende Identität konstituiert [...], sondern durch eine vom vokabularen Kern herrührende Benennung [...] geprägt und anschließend durch Prozesse der Um- und Neudefinition innerhalb der Fachsprache erst gebildet wird“ (S. 180).

13 Vgl. Randall Davis / Howard Shrobe / Peter Szolovits, „What is a Knowledge Representation?“, in: *AI Magazine* 14/1 (1993), S. 17–33, hier S. 18f. Online unter <<http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>>. Vgl. auch Herbert Stachowiak, *Allgemeine Modelltheorie*, Köln 1973.

14 In dem Versuch wurde eines der ab 1978 von John Wheeler formulierten Gedankenexperimente („delayed-choice experiment“) erstmals mit stabilen Teilchen (Helium) erfolgreich umgesetzt. Vgl. A. G. Manning / R. I. Khakimov / R. G. Dall / A. G. Truscott, „Wheeler’s Delayed-Choice Gedanken Experiment With a Single Atom“, in: *Nature Physics* 11 (2015), S. 539–542. Online unter <<https://doi.org/10.1038/nphys3343>>.

Realität existiert (zumindest auf Quantenebene) nicht, solange sie nicht beobachtet wird; sie liegt also nicht mehr nur sprichwörtlich, sondern auch experimentell belegt, im Auge des Betrachters. Zum anderen verschwimmen in diesem Punkt interessanterweise die Begrifflichkeiten informationstechnischer Ontologien (im Plural) mit dem traditionellen Verständnis klassisch-philosophischer Ontologie (im Singular), die sich ja als Teilgebiet der Metaphysik gerade mit sprachlich vermittelten Strukturen des Wirklichen und Nicht-Wirklichen auseinandersetzt. Es sind hier, wie auch bei der Wissensmodellierung, Begriffe, die die Kommunikation über einen Weltausschnitt ermöglichen und dadurch eine Realität vermitteln.

Die Modellierung für ein digitales Medium zwingt uns, über die Bedeutung von vermeintlich ganz allgemeinverständlichen Begriffen und Konzepten neu nachzudenken bzw. deren bislang implizite Bedeutungsschichten explizit zu formulieren. Gerade dieses Explizit-Machen, das Explizit-Sein von Begriffen und Konzepten spielt eine entscheidende, nicht deutlich genug zu betonende Rolle. Als menschliche Akteure besitzen wir eine „hermeneutische Kompetenz“ (wie es der Berliner Philosoph Wilhelm Schmid ausgedrückt hat), also die „Fähigkeit, immer mehr in den Dingen zu vermuten, als wir momentan sehen“<sup>15</sup>. Einem Rechner, einer maschinellen Verarbeitung müssen all diese Bedeutungskontexte erst zugänglich gemacht werden (Verkürzungen wie „Es weiß ja jeder, was gemeint ist“, Metaphern oder Ironie versteht ein Rechner nur schwer bis gar nicht.) Auch hierbei ist mitunter keine Vollständigkeit zu erreichen. 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 im geisteswissenschaftlichen Kontext oft bewusst verwendeten begrifflichen Ambivalenzen oder Mehrdeutigkeiten auf einer anderen Ebene untergehen. Es droht also ein Verlust der Unschärfe eines Begriffs, seiner „Ambiguitätsdimensionen“<sup>16</sup>, die doch erst seine multiperspektivische Anwendung ermöglichen. „Container“-Begriffe wie das „musikalische Werk“, zunächst gerühmt, dann gescholten und später für tot erklärt, bündeln ja mithin abweichende Auffassungen der dahinterliegenden komplexen Konstrukte, erlauben aber gerade dadurch eine komplexitätsreduzierende, aufwandsökonomische Verständigung. Neben der stets reduzierenden Modellierung eines Begriffs besteht also eine weitere Herausforderung in der expliziten Abbildung der Mannigfaltigkeit seiner Bedeutungen: Komplexität und Vielschichtigkeit bei gleichzeitiger Explizität. Dieses Explizit-Machen hat drastische Auswirkungen auf das begriffliche Wissen selbst; es verändert sich unmittelbar, wird transformiert. Überspitzt gesagt, führt die Verwendung des Computers zwangsläufig zu einem anderen Werkbegriff.

Besonders deutlich zeigt sich hier die Wichtigkeit von ausführlichen Dokumentationen, die nachprüfbar machen, wer einen Begriff in welchem Zusammenhang wann und wie benutzt hat, und welche Bedeutung dem Begriff zu diesem Zeitpunkt zugeschrieben wurde. Die Modellierung eines Konzepts muss dessen explizite Anschlussfähigkeit ermöglichen. Weitere Problemfelder in diesem Bereich, gerade im geisteswissenschaftlichen Zusammenhang, betreffen die Modellierung von Argumentationssträngen und von echten Verneinun-

15 Heimo Schwilk, „Ich plädiere für das Glück der Fülle“. Der Berliner Philosoph Wilhelm Schmid über die Kunst des Lebens“, in: *Welt am Sonntag* (13.11.2005). Online unter <<https://www.welt.de/printwams/article134958/Ich-plaedierte-fuer-das-Glueck-der-Fuelle.html>>.

16 Vgl. Matthias Bauer / Joachim Knape / Peter Koch / Susanne Winkler: „Dimensionen der Ambiguität“, in: *Zeitschrift für Literaturwissenschaft und Linguistik* 40 (2010), S. 7–75. Online unter <<http://hdl.handle.net/10900/53217>>.

gen oder Nicht-Existenzen (wie z. B. bei verschollenen Quellen oder Manuskripten)<sup>17</sup> sowie die Heterogenität von Daten.<sup>18</sup>

Die im vorangegangenen Abschnitt diskutierten Aspekte (und es lassen sich noch einige weitere finden) bieten mögliche Ansatzpunkte für einen kritisch reflektierten Zugang zu Fragen der Modellierung von Wissensstrukturen, die meiner Kenntnis nach bislang noch wenig tatsächlich im Detail untersucht worden sind.

### *Technisches, Allzutechnisches?*

Das für die Standardisierung der im World Wide Web verwendeten Technologien verantwortliche W3-Consortium empfiehlt im Zusammenhang mit Ontologien RDF, RDFS und OWL als Beschreibungs- und SPARQL als Abfragesprache.<sup>19</sup> Grundlegend ist hierbei die Verwendung global eindeutiger (Web-)Adressen, sogenannter IRIs (Internationalized Resource Identifier), die zur Bezeichnung sämtlicher Entitäten, seien es Klassen, Attribute, Relationen oder Instanzen, herangezogen werden. Als „digitaler Fingerabdruck“ einer im Web repräsentierten Entität, oder auch: Ressource, dienen IRIs der distinkten Identifizierung und Unterscheidung jeglicher Ressourcen.

Die elementare Syntax für die Beschreibung von Ressourcen wird durch das Resource Description Framework (RDF) definiert. Jedes Statement muss in Form eines dreiteiligen Aussagesatzes (Tripel; engl.: triple) formuliert sein: <subject> <predicate> <object>, wobei zu einem Subjekt beliebig viele Aussagen getroffen werden können und Subjekte in weiteren Aussagen auch als Objekt fungieren können. Durch eine derartige Aneinanderreihung und Verkettung von Tripeln lassen sich aus dieser Minimalsyntax, die im Grunde nicht über das Formulierungsvermögen eines Kleinkindes hinausgeht, durchaus komplexe Wissenszusammenhänge formulieren. Tripel werden häufig als gerichtete Graphen dargestellt, können aber genauso textuell in verschiedenen Serialisierungen (wie RDF/XML, Turtle, JSON-LD) umgesetzt werden (vgl. Abb. 2). Die Menge aller Tripel wird auch als RDF Graph bezeichnet.

Begriffliche Basiskonzepte und -eigenschaften werden durch die Erweiterung RDF Schema (RDFS) bereitgestellt, wodurch ganz grundlegende Aussagen über das zu modellierende Vokabular möglich sind: Konzepte können als Klassen und Unterklassen, Eigenschaften als Properties und Subproperties ausgewiesen und zueinander in Beziehung gesetzt werden. Lightweight ontologies wie Taxonomien lassen sich auf dieser Ebene bereits modellieren.

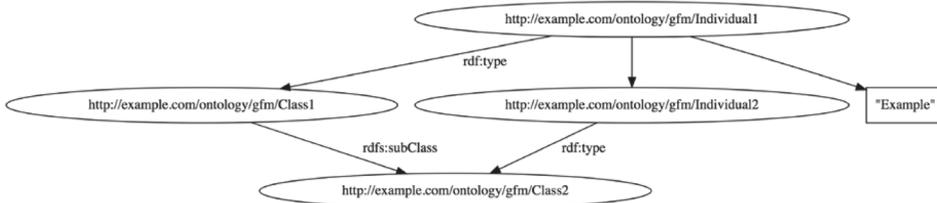
Sollen komplexe sprachliche Konstrukte oder logische Operationen unterschiedlichen Grades in der Modellierung eine Rolle spielen, reichen RDF und RDFS allein nicht aus. Erst die Web Ontology Language (OWL) ermöglicht die Konzeption semantisch höhergradiger Modelle, die, je nach „Dialekt“ (Lite, Description Logic DL, Full), z. B. mengenlo-

17 Vgl. Stefan Münnich, „Quellenverluste als methodologischer Unsicherheitsbereich für Editorik und Datenmodellierung am Beispiel von Anton Weberns George-Liedern op. 3 & 4“, Referat auf der Tagung Graphentechnologien 2018 („Modellierung des Zweifels“), 19.–20. Januar 2018, Akademie der Wissenschaften und der Literatur, Mainz (Veröffentlichung in Vorbereitung.)

18 Grundlegend dazu siehe Holger Wache, *Semantische Mediation für heterogene Informationsquellen*, Berlin 2003.

19 Vgl. hierzu die Standardwerke Dean Allemang / James Hendler, *Semantic Web for the Working Ontologist. Effective Modelling in RDFS and OWL*, Amsterdam u. a. 2011 und Bob DuCharme, *Learning SPARQL: Querying and Updating with SPARQL 1.1*, Beijing u. a. 2013. Beschreibungen, Dokumentation und Spezifikationen online zugänglich jeweils unter <<https://www.w3.org/2001/sw/wiki/{RDF|RDFS|OWL|SPARQL}>>>.

## a) Graph:



## b) Turtle:

```

@prefix ex: <http://example.com/ontology/gfm/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

ex:Individual1 rdf:type ex:Class1 .
ex:Individual2 rdf:type ex:Class2 .
ex:Individual1 ex:property1 ex:Individual2 .
ex:Individual1 ex:property2 "Example"@en .
ex:Class1 rdfs:subClass ex:Class2 .

```

## c) JSON-LD:

```

{
  "@graph": [
    {
      "@id": "http://example.com/ontology/gfm/Class1",
      "http://www.w3.org/2000/01/rdf-schema#subClass": {
        "@id": "http://example.com/ontology/gfm/Class2"
      }
    },
    {
      "@id": "http://example.com/ontology/gfm/Individual1",
      "@type": "http://example.com/ontology/gfm/Class1",
      "http://example.com/ontology/gfm/property1": {
        "@id": "http://example.com/ontology/gfm/Individual2"
      },
      "http://example.com/ontology/gfm/property2": {
        "@language": "en",
        "@value": "Example"
      }
    },
    {
      "@id": "http://example.com/ontology/gfm/Individual2",
      "@type": "http://example.com/ontology/gfm/Class2"
    }
  ]
}

```

Abbildung 2: Verschiedene Serialisierungen eines (fiktiven) RDF-Graphen mit zwei verknüpften Individuen als a) gerichteter Graph, b) in Turtle, und c) in JSON-LD

gische Verknüpfungen von Klassen (UND/ODER-Beziehungen) oder mögliche Kardinalitäts- und Werteschränkungen (restrictions) explizit machen können.

Spezialisierte Graphdatenbanken, sogenannte Triplestores, können Tripel speichern, verwalten und verarbeiten.<sup>20</sup> Sie können mit Hilfe der standardisierten Anfragesprache SPARQL (SPARQL Protocol and RDF Query Language) abgefragt und manipuliert werden. Vergleichbar der Datenbanksprache SQL für relationale Datenbanken, ist SPARQL jedoch auf die Verarbeitung von in RDF beschriebenen und strukturierten Daten spezialisiert. Im Unterschied zu SQL können zudem nicht nur lokale, sondern durch das in der SPARQL-Spezifikation enthaltene Übertragungsprotokoll auch externe Datensätze abgefragt werden. Mit sogenannten SPARQL-Endpoints können Institutionen und Projekte Zugriffspunkte auf ihre RDF-Datensätze zur Verfügung stellen.<sup>21</sup>

Im Folgenden soll ein Blick auf Anwendungsmöglichkeiten der hier beschriebenen Technologien für eine „semantische“ digitale Musikwissenschaft geworfen werden. Wo werden diese möglicherweise schon eingesetzt, welche Modelle / Pattern erscheinen aussichtsreich und welches Potential bietet sich hier?

### *Modelle für die Geisteswissenschaften*

Neben generischen Top-Level Ontologies wie FOAF (Friend-of-a-Friend), SKOS (Simple Knowledge Organization System) oder den Ontologien der DCMI (Dublin Core Metadata Initiative), die aufgrund der Allgemeinheit ihrer Konzepte extrem verbreitet sind, existieren verschiedene Modelle, die gerade im Zusammenhang mit geisteswissenschaftlichen Fragestellungen häufig zum Einsatz kommen:

Das CIDOC Conceptual Reference Model (CIDOC CRM)<sup>22</sup> hat im Bereich des kulturellen Erbes sicherlich den größten Bekanntheits- und Verbreitungsgrad. Als ISO-Norm standardisiert (ISO 21127:2006 bzw. 21127:2014), wird es von einer breiten und wachsenden Community, anfänglich vor allem der Gedächtnisinstitutionen wie Museen und Archiven, getragen und weiterentwickelt.<sup>23</sup>

20 Mittlerweile konnten Datensätze mit über 1 Billion ( $10^{12}$ ) Tripeln, was knapp 130 Aussagen zu jeder lebenden Person entspräche, geladen und abgefragt werden. Vgl. <<https://www.w3.org/wiki/index.php?title=LargeTripleStores&oldid=97382>>.

21 In der Übersicht des W3-Consortiums über existierende SPARQL-Endpoints finden sich auch einige musikbezogene Datenbestände („BBC Programmes and Music“, „Jamendo“, „MusicBrainz“ und „DB-Tune“). Vgl. <<https://www.w3.org/wiki/SparqlEndpoints>>. Bis auf „BBC Programmes and Music“ waren diese bereits in der allerersten LOD Cloud-Darstellung aus dem Jahr 2007 (<<http://lod-cloud.net/versions/2007-05-01/lod-cloud.png>>) vertreten.

22 Dokumentation der aktuellen Spezifikation für CIDOC CRM (Version 6.2.3 vom 17. Mai 2018) online unter <<http://www.cidoc-crm.org/Version/version-6.2.3-0>>.

23 U. a. setzt das British Museum vollständig auf CIDOC CRM, um Daten zu seinen Sammlungsobjekten über einen „Semantic Endpoint“ zur Verfügung zu stellen; vgl. Dominic Oldman, „The British Museum, CIDOC CRM and the Shaping of Knowledge“, Blog-Eintrag 4. September 2012, <<http://www.oldman.me.uk/blog/the-british-museum-cidoc-crm-and-the-shaping-of-knowledge/>>. Mit ResearchSpace (<<http://researchspace.org>>), einer kollaborativen Semantic Web-Umgebung, wird zudem eine Plattform aufgebaut, die den Austausch von „knowledge based projects and applications“ vereinfachen soll. Als Anwendungsbeispiel vgl. zuletzt Florian Kräutli / Matteo Valleriani, „CorpusTracer: A CIDOC Database for Tracing Knowledge Networks“, in: *Digital Scholarship in the Humanities* 33/2 (2018), S. 336–346.

Mit FRBRoo<sup>24</sup>, einer objektorientierten Anpassung des bibliographischen Modells Functional Requirements for Bibliographic Records (FRBR) der International Federation of Library Associations and Institutions (IFLA), konnte zudem eine mit dem CIDOC CRM kompatible Erweiterung vorgelegt werden, die den konzeptuellen Austausch zwischen Museum und Bibliothek vereinfacht und vereinheitlicht. Besonders interessant auch für die Musikwissenschaft ist hier die Überführung der sogenannten Group 1 Entities aus FRBR (<work>, <expression>, <manifestation>, <item>) in ein doppelt dreistufiges Pattern (vgl. Abb. 3): Eines abstrakten Werkkonzepts (F14 Individual Work) individuelle Realisierung (F22 Self-contained Expression) wird durch einen Schaffensprozess (F28 Expression Creation) hervorgebracht (konzeptuelle Ebene). Zugleich entsteht dabei ein manifester Zeichenträger (F4 Manifestation Singleton), der den Inhalt der F22 Self-contained Expression aufnimmt (physische Ebene). Diese Aufspaltung des „Werk“-Begriffs ermöglicht sehr komplex modellierte und differenzierte Aussagen sowohl über den Entstehungs- als auch den Produktionsprozess z. B. einer Komposition.

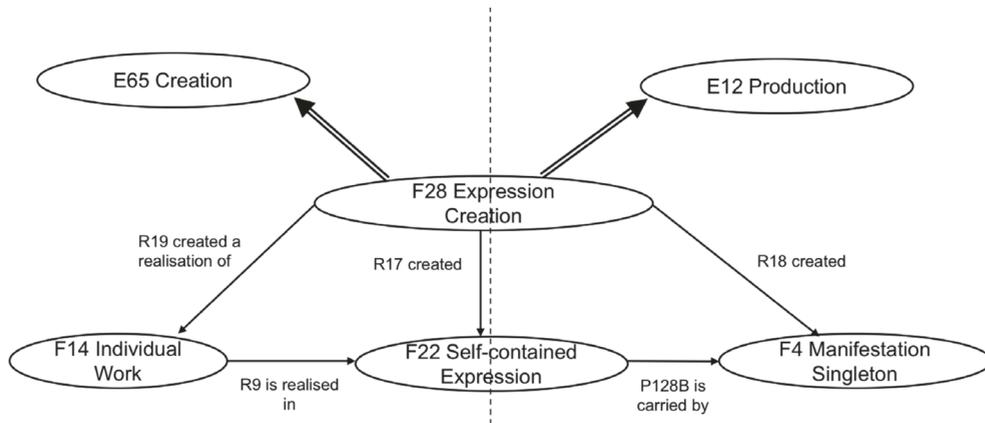


Abbildung 3: Doppelt dreistufiges „Werk“-Pattern in *FRBRoo*; links: Konzeptuelle Ebene, rechts: Physische Ebene (nach: Spezifikation *FRBRoo* [Anm. 24], S. 21)

Einen entscheidenden Ansatz zur Integration heterogener Datenquellen bietet das Europeana Data Model for Cultural Heritage (EDM).<sup>25</sup> Mit dem dort implementierten <Aggregation>-Konzept können aus verschiedenen Projekten stammende Webressourcen oder Daten zu ein und demselben Objekt auf den Datenprovider einzeln rückführbar gebündelt werden (vgl. Abb. 4). Einer der ersten vollständig mit dem EDM modellierten Datensätze entstammte dem MIMO-Projekt (Musical Instrument Museums Online), dessen über 40.000 Objekte in der Europeana „aggregiert“ wurden.<sup>26</sup> Mittlerweile lassen sich in der Europeana knapp 133 Millionen Webressourcen und ca. 55 Millionen Aggregationen und beschriebene Kulturerbe-Objekte nachweisen.

24 Dokumentation der aktuellen Spezifikation für FRBRoo (Version 2.4 vom November 2015) online unter <[https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo\\_v\\_2.4.pdf](https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf)>.

25 Dokumentation online unter <https://pro.europeana.eu/page/edm-documentation>. SPARQL-Endpoint der Europeana unter <<http://sparql.europeana.eu/>>.

26 Vgl. Valentine Charles, „MIMO and EDM“. Blog-Eintrag 18. November 2014, <<https://pro.europeana.eu/page/mimo-edm>>.

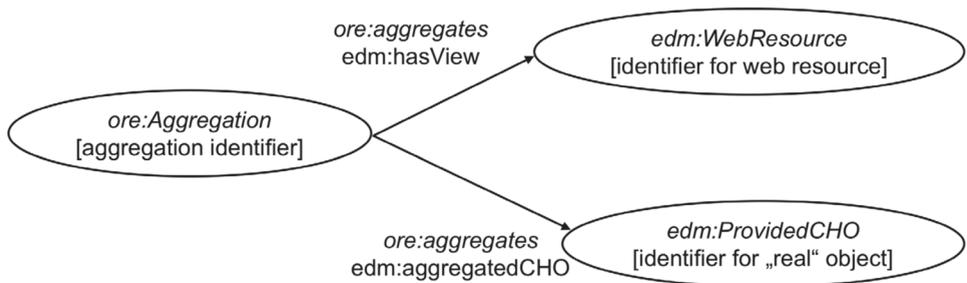


Abbildung 4: Visualisierung der drei Hauptklassen des Europeana Data Model for Cultural Heritage (EDM; CHO = Cultural Heritage Object) unter Einbindung des <Aggregation>-Konzepts aus dem ORE-Vokabular der Open Archives Initiative, <<http://www.openarchives.org/ore/1.0/toc>> (nach: Europeana Data Model Primer, 14.7.2013, S. 10. Online unter <[https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Primer\\_130714.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf)>).

Speziell auf musikalische Phänomene bezogen sind die ab 2007 am Centre for Digital Music der Queen Mary University of London entwickelte Music Ontology<sup>27</sup> sowie die auf ihr aufbauende, seit 2017 an der Universität Münster konzipierte MusicOWL – Music Score Ontology.<sup>28</sup> Beide beruhen auf OWL, der Web Ontology Language, betrachten jedoch unterschiedliche, einander ergänzende Bereiche: Der Music Ontology geht es vor allem um Aussagen zu Musikproduktionsprozessen (vgl. Abb. 5). Ihre Ableger und ergänzenden Modelle (timeline, event, keys, tonality, symbolic notation, chord, temperament, audiofeatures) erlauben die Modellierung und Repräsentation auch detaillierter musikalischer Informationen. Vor allem im Bereich von Musikempfehlungssystemen findet die Music Ontology breite Anwendung.

27 Spezifikation unter <<http://motools.sourceforge.net/doc/musicontology.html>>. Vgl. auch Yves Raimond / Samer Abdallah / Mark Sandler / Frederick Giasson, „The Music Ontology“, in: *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Wien 2007, S. 417–422. Online unter <[http://ismir2007.ismir.net/proceedings/ISMIR2007\\_p417\\_raimond.pdf](http://ismir2007.ismir.net/proceedings/ISMIR2007_p417_raimond.pdf)>. Auch wenn sie derzeit als Standardontologie für musikalische Sachverhalte gelten muss, ist die Entwicklung der Music Ontology seit 2014 eingestellt.

28 Spezifikation unter <<http://linkeddata.uni-muenster.de/ontology/musicscore#>>. Vgl. Jim Jones / Kleber Tertuliano / Diego de Siqueira Braga / Tomi Kauppinen, „MusicOWL. The Music Score Ontology“, in: *Proceedings of the International Conference on Web Intelligence (WI '17)*, Leipzig 2017, S. 1222–1229. Online unter <<https://doi.org/10.1145/3106426.3110325>>. Barbara Wiermann danke ich für den freundlichen Hinweis auf das Projekt.

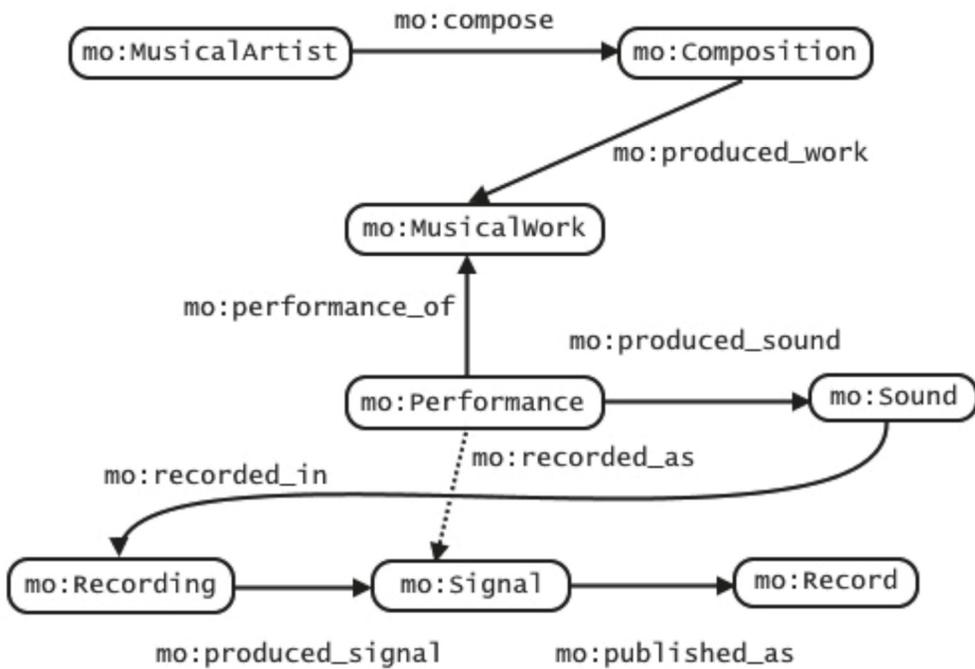


Abbildung 5: Ausschnitt (Musikproduktionsprozess) aus der Music Ontology (nach: Mark Sandler/Yves Raimond/Christopher Sutton, „Interlinking Music-Related Data on the Web“, in: *IEEE MultiMedia* 16/2 (2009), S. 54. Online unter <<https://doi.org/10.1109/MMUL.2009.29>>)

Die MusicOWL – Music Score Ontology dagegen setzt erst ab der Partiturebene an und beschreibt Elemente des Notensatzes, darunter Dynamik, Artikulation, Schlüsselung sowie System-, Stimmen- und Taktangaben, wobei sie die Music Ontology und ihre genannten Ergänzungsmodelle geschickt nachnutzt und erweitert. Bemerkenswerterweise können hierbei z. B. einzelne Noten oder Akkorde überlappungsfrei zugleich einem Takt und einer Stimme zugeordnet werden (vgl. Abb. 6).<sup>29</sup>

29 Die entweder vertikale (Takt) oder horizontale (Stimme / System) Auffassung musikalischer Notation (zumindest im Bereich der Common Western Notation) stellt in Bezug auf ihre Codierung eine Herausforderung dar, die durch die sich überlappenden Hierarchien in Baumstrukturen wie XML kaum aufgelöst werden kann. MusicXML forciert die alternative Verwendung einer von zwei Dokument-Strukturen (entweder <score-partwise> oder <score-timewise>), zwischen denen ein Konvertierungs-Stylesheet vermitteln kann. Das Schema der Music Encoding Initiative (MEI) erlaubt die parallele Verwendung von <score> (taktbezogen) und <parts> (stimmenbezogen), deren codierte Inhalte durch entsprechende Verweismöglichkeiten aufeinander bezogen werden können und so zumindest nicht gedoppelt werden müssen. Vgl. Johannes Kepler, *Musikedition im Zeichen neuer Medien. Historische Entwicklung und gegenwärtige Perspektiven musikalischer Gesamtausgaben* (= Schriften des Instituts für Dokumentologie und Editorik 5), Norderstedt 2011, S. 370f.

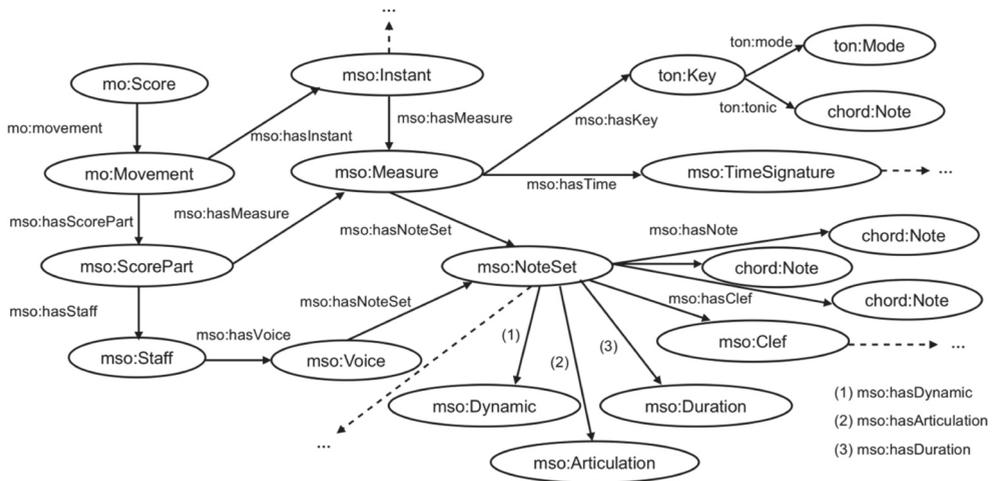


Abbildung 6: Ausschnitt aus der MusicOWL Ontology (mso) unter Einbindung von Music Ontology (mo), Chord Ontology (chord) und Tonality Ontology (ton) (nach: Jones/Tertuliano / de Siqueira Braga / Kauppinen, „MusicOWL. The Music Score Ontology“ [Anm. 28], S. 1224)

Es bleibt abzuwarten, welche Strahlkraft die Music Score Ontology erreichen kann, was sicher auch von der Frage abhängt, inwieweit sie bestehende Standards integrieren oder an diese anknüpfen kann. Momentan existiert nur eine Möglichkeit der Übertragung von MusicXML zu RDF, ein entsprechendes Modul zur Übertragung von im Format der MEI codierten Daten wäre im akademischen und philologischen Kontext begrüßenswert.<sup>30</sup>

Es ließen sich viele weitere Modelle aufführen, die im Zusammenhang geistes- und musikwissenschaftlicher Fragestellungen eine Rolle spielen können, wie z. B. die Provenance Ontology (PROV-O) oder Time Ontology. Allerdings ist die Menge bereits verfügbarer Ontologien kaum mehr überschaubar und zudem hängt es stark von der eigenen Perspektive und Fragestellung ab, welche Konzepte Eingang in eine Modellierung finden sollen. Übersichtsseiten wie Linked Open Vocabularies (<<http://lov.okfn.org/dataset/lov/>>) oder Open Metadata Registry (<<http://metadataregistry.org/>>) werden zum Auffinden existierender Modelle daher immer wichtiger. Eine entsprechende Suche ist sinnvoll, da der Rückgriff auf bereits bestehende Ontologien stets Vorrang gegenüber einer Eigenmodellierung haben sollte (diese ist natürlich im Fall nicht vorhandener oder in Bezug auf die eigene Fragestellung nicht adäquat anwendbarer Konzepte immer möglich). Durch das fundamentale Paradigma des World Wide Webs („Anybody can say Anything about Any topic“) ist Konsensbildung ein unentbehrlicher Mechanismus zur Einordnung der vorhandenen Daten und Informationen. Die An- und Wiederverwendung existierender Modelle trägt zu dieser Konsensbildung bei.

30 Vgl. MusicXML to RDF Converter im Beta-Stadium, <<https://github.com/jimjonesbr/musicowl>>. Der von Torsten Schrade und der Digitalen Akademie Mainz entwickelte Webservice XTriples (<<https://github.com/spatialhumanities/xtriples>>), der die Generierung von RDF-Statements aus beliebigen XML-Dateien erlaubt und somit eine Brücke zwischen X-Technologien und semantischen Technologien schlägt, ließe sich hier für eine weitergehende Anwendung auf MusicXML oder MEI-Format sinnvoll nachnutzen.

*Ontologien in der musikwissenschaftlichen Praxis*

Die Einführung und Anwendung neuer Technologien bedarf immer eines gewissen Mutes und Freiraums für Experimente, Fehlversuche und erfolgreiches Scheitern.<sup>31</sup> Umso erfreulicher ist es, dass einige musikwissenschaftliche Projekte jenen Mut aufbringen können, den mitunter hohen Modellierungsaufwand nicht scheuen und durchaus als Vorreiter einer „semantischen“ digitalen Musikwissenschaft gelten können:

DOREMUS (Doing Reusable Musical Data), eine sehr ambitionierte Kooperation von Radio France, französischer Nationalbibliothek BnF und der Philharmonie de Paris, hat die Beschreibung der Musikataloge in den betreffenden Institutionen zum Ziel.<sup>32</sup> Sehr eindrücklich ist das entwickelte Modell durch seine Verbindung und Nachnutzung von Music Ontology, CIDOC CRM, FRBRoo und Europeana Data Model (siehe oben). Dabei kommt auch das bereits angesprochene komplexe, doppelt dreistufige „Werk“-Pattern zum Tragen, jeweils zusätzlich verknüpft mit einem F15 Complex Work-Objekt (vgl. Abb. 7).

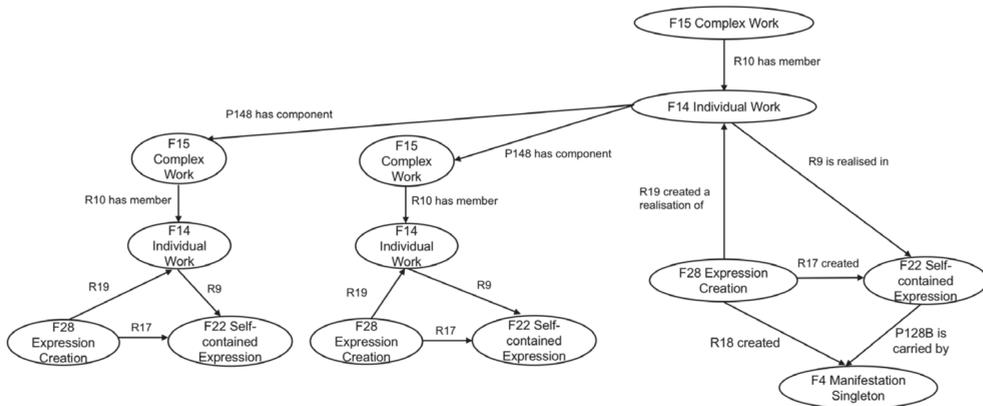


Abbildung 7: Modellierungskonzept von DOREMUS mit verschachtelten dreistufigen „Werk“-Patterns aus FRBRoo (nach: <<http://www.doremus.org/wp-content/uploads/2015/11/Capture-d%E2%80%99%C3%A9cran-2015-11-23-%C3%A0-23.25.03.png>>)

Derartige „Werkkomplexe“ ließen sich mit dem französischen Philosophen Gilles Deleuze als reine „Virtualitäten“ bezeichnen: Diese bündeln in einem abstrakten (oder bei Deleuze: virtuellen) Container sämtliche zu einer Komposition gehörenden, erst in ihren realisierten Manifestationen (bei Deleuze: Aktualisierungen) erkennbaren Entitäten, z. B. die Einzeltitel einer Liedersammlung, deren verschiedene Fassungen sowie Aufführungen und Aufnahmen (die sich alle im Detail drastisch unterscheiden können, dennoch als Teil dieses „virtuellen“ Werkkomplexes identifiziert und kommuniziert werden können).<sup>33</sup> Es zeigt sich,

31 Im Semantic-Web-Kontext bildet auch die Offenheit bzw. Offenlegung von (Forschungs-)Daten eine mögliche Hemmschwelle, vgl. Jens Mittelbach, „Modernes Datenmanagement: Linked Open Data und die offene Bibliothek“, in: *o-bib – das offene Bibliotheksjournal* 2/2 (2015), S. 61–73, hier S. 68. Online unter <<https://doi.org/10.5282/o-bib/2015H2S61-73>>.

32 Projektwebsite unter <<http://www.doremus.org>>. Daten, Modell und Dokumentation online unter <<http://data.doremus.org/>>.

33 Vgl. dazu Thomas Ahrend, „Werk und virtueller Text – Spuren von Performativität in den Kompositionen Anton Weberns“, Referat auf der 17. Internationalen Tagung der Arbeitsgemeinschaft für germanistische Edition: „Aufführung und Edition“, 14.–17. Februar 2018, Goethe-Universität Frankfurt

dass hier ein sehr komplexes und (selbst im Zusammenhang postmoderner philosophischer Konzepte) tragfähiges Wissensmodell erstellt wurde, das selbst wiederum ein breites Anwendungsspektrum bietet (so prüft die Anton Webern Gesamtausgabe momentan, inwieweit das DOREMUS-Modell zur Beschreibung musikalischer Editionen angepasst werden kann).

Mit MELD (Music Encoding and Linked Data) hat das britische Transforming Musicology-Projekt ein semantisches Framework aufgebaut, das die Live-Annotation von digitalen Notentexten (im Format der MEI) durch semantische Technologien erprobt.<sup>34</sup> Auch hier werden etliche existierende Modelle nachgenutzt, u. a. Music Ontology, FRBR, SKOS, PROV-O (Provenance), Web Annotation Ontology. Ziel ist die Anreicherung von MEI-Daten mit semantischen Web-Annotationen, die eine dynamische Echtzeit-Kommunikation, vermittelt durch den Notentext wohlgeordnet, zwischen den Teilnehmenden einer Aufführungssituation (Orchester- oder Bandmitglieder) gewährleisten sollen.<sup>35</sup>

Das in der Freizeit der beteiligten australischen Forscher realisierte Projekt JazzCats (Jazz Collection of Aggregated Triples) zieht Informationen zu Aufführungen, Aufnahmen und Personen aus drei verschiedenen jazzbezogenen Datensätzen (Body & Soul, WJazzD und LinkedJazz) zusammen und verknüpft diese durch Konzepte der Music Ontology.<sup>36</sup>

Im Fokus von IncipitSearch, einer Web-Anwendung, die im Rahmen der Gluck-Gesamtausgabe von der Digitalen Akademie Mainz entwickelt wird, steht die semantische Suche nach musikalischen Anfängen.<sup>37</sup> Dafür wird die in schema.org (einem suchmaschinenoptimierten Vokabular für strukturierte Daten) vorhandene Klasse MusicComposition um eine Unterklasse MusicIncipit erweitert, die Angaben zu Schlüssel, Taktart und -vorzeichnung sowie das eigentliche Incipit im Plaine & Easy-Format aufnehmen kann (weitere Formate wie MEI oder abc-Notation sind geplant). Momentan lassen sich über das Webportal Datensätze aus RISM, der italienischen Nationalbibliothek (SBN) sowie dem Werkverzeichnis der Gluck-Gesamtausgabe abfragen.

---

am Main. Zu Virtualität und Aktualität bei Deleuze siehe Gilles Deleuze, *Differenz und Wiederholung*, München 1992 (frz. Original: Paris 1968), S. 264–279.

34 Projekt-Repository unter <<https://github.com/oerc-music/meld>>. Vgl. auch David M. Weigl / Kevin R. Page, „A Framework for Distributed Semantic Annotation of Musical Score: ‚Take It to the Bridge!‘“, in: *Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017*, hrsg. von Xiao Hu, Sally Jo Cunningham, Doug Turnbull und Zhiyao Duan, 2017, S. 221–228. Online unter <[https://ismir2017.smcnus.org/wp-content/uploads/2017/10/190\\_Paper.pdf](https://ismir2017.smcnus.org/wp-content/uploads/2017/10/190_Paper.pdf)>.

35 Vgl. auch die Umsetzung von „Climb!“, einer „game-like performance for disklavier and electronics“ der finnischen Komponistin Maria Kallionpää, im Rahmen des MELD-Projekts, in Maria Kallionpää / Chris Greenhalgh / Adrian Hazzard / David M. Weigl / Kevin R. Page / Steve Benford, „Composing and Realising a Game-like Performance for Disklavier and Electronics“, in: *New Interfaces for Musical Expression (NIME'17)*, *Proceedings*, hrsg. von Cumhur Erkut, Kopenhagen 2017, S. 464–469. Online unter <<http://eprints.nottingham.ac.uk/id/eprint/44529>>.

36 Daniel Bangert / Terhi Nurmikko-Fuller / Alfie Abdul-Rahman, *JazzCats Project* (2016). Online (inkl. SPARQL-Endpoint) unter <<http://jazzcats.oerc.ox.ac.uk/>>.

37 Projektwebsite unter <<https://incipitsearch.adwmainz.net/>>. Vgl. auch Anna Neovesky / Frederic von Vlahovits, „IncipitSearch – Interlinking Musicological Repositories“, Referat auf der DH2018, 26.–29. Juni 2018, Mexico City (ausgezeichnet mit dem Paul Fortier Prize des Dachverbands ADHO [Alliance of Digital Humanities Organizations] für das beste „Young Scholar Paper“ auf der DH2018). Abstract online unter <<https://dh2018.adho.org/en/incipitsearch-interlinking-musicological-repositories/>>.

Als Proof-of-Concept haben die Autoren der Music Score Ontology einen Datensatz-Prototyp (Linked Music Score Dataset) mit aktuell 42 kompletten Partituren erstellt, die knapp über 200.000 Noten-Objekte enthalten (insgesamt etwas mehr als 3 Millionen Triples). Diese lassen sich über eine (an RISM oder auch IncipitSearch angelehnte) Portalsuche oder über einen frei zugänglichen Triplestore abfragen.<sup>38</sup>

Hingewiesen sei zudem auf die Linked-Data-Services der Deutschen Nationalbibliothek (DNB)<sup>39</sup> und des Répertoire International des Sources Musicales (RISM)<sup>40</sup> sowie das am Max-Planck-Institut für Empirische Ästhetik durchgeführte Projekt „Fachgeschichte der deutschsprachigen Musikwissenschaft von ca. 1810 bis ca. 1990“, das innerfachliche Beziehungsnetzwerke in semantisch strukturierter Form aufbereiten will.<sup>41</sup>

Ab einem gewissen Komplexitätsgrad und zunehmender Verknüpfungstiefe lassen sich Ontologien nur noch ungenügend mit Zettel und Bleistift niederschreiben und visualisieren (auch wenn dies am Anfang eines Modellierungsprozesses prinzipiell eine gute Wahl darstellt). Spezialisierte virtuelle Forschungsumgebungen können dabei helfen, Workflows zu erleichtern, kollaborative Arbeitsprozesse zu begünstigen und die Modellierung und den Einsatz von Ontologien durch die (Semi-)Automatisierung sich wiederholender Prozesse zu vereinfachen. Zwei prominente virtuelle Forschungsumgebungen nutzen ontologiebasierte Modelle: WissKI (<<http://www.wiss-ki.eu>>), eine in Erlangen/Nürnberg entwickelte wissenschaftliche Kommunikationsinfrastruktur, und Knora/SALSAH (<<http://www.knora.org>>; <<http://www.salsah.org>>), das am Digital Humanities Lab der Universität Basel realisiert wird. Beiden gemein ist, dass sie als generische Systeme eine Vielzahl von Projekten unterschiedlichster Fachrichtungen (Kunstwissenschaft, Literaturwissenschaft, Musikwissenschaft u. a.) unterstützen können.

Knora stellt ein Set von OWL-konformen Basis-Ontologien (Knora Ontologies) zur Verfügung, die vor allem Nutzungsrechte und Integrität von Transaktionen absichern, und auf die projektspezifische Ontologien aufgesetzt werden können. Zudem zählen eine Triplestore-Schnittstelle (Knora API Server), ein IIIF-kompatibler Medienserver und die graphische Benutzeroberfläche SALSAH zum „Ökosystem“ von Knora. Über SALSAH macht u. a. die Anton Webern Gesamtausgabe ihre Forschungsdatenbank offen zugänglich, die Bestandteile ihrer Online-Editionen werden über eine durch die Schweizer Nationale Infra-

38 WWU Music Score Portal im Beta-Stadium online unter <<http://linkeddata.uni-muenster.de/musicportal/>>. Triplestore online unter <<http://linkeddata.uni-muenster.de:7200/>>.

39 In Zukunft sollen hier auch Daten des Deutschen Musikarchivs zugänglich gemacht werden. Vgl. Jana Hentschke / Simon Zetzsche, „Linked Data und das Deutsche Musikarchiv“, Referat auf der Jahrestagung der AIBM, 4.–8. September 2017, Universitäts- und Landesbibliothek Münster. Präsentation online unter <<https://www.slideshare.net/JanaHentschke/linked-data-fr-das-deutsche-musikarchiv>>. Zum Linked-Data-Service der DNB siehe <[http://www.dnb.de/DE/Service/DigitaleDienste/LinkeddData/linkedd\\_data\\_node.html](http://www.dnb.de/DE/Service/DigitaleDienste/LinkeddData/linkedd_data_node.html)>. Des Weiteren bietet das Hochschulbibliothekszentrum des Landes NRW mit *lobid-gnd* (<<http://www.lobid.org/gnd>>) eine Linked-Open-Data-Schnittstelle für die Einträge der Gemeinsamen Normdatei (GND) an. Weiterführend zu Normdateien siehe den Beitrag von Barbara Wiermann in diesem Heft.

40 Zum Open-Data-Service von RISM siehe <<https://opac.rism.info/index.php?id=8>>. SPARQL-Endpoint unter <<https://opac.rism.info/index.php?id=16>>.

41 Vgl. Annette van Dyck-Hemming / Melanie Wald-Fuhrmann, „Vom Datum zum historischen Zusammenhang. Möglichkeiten und Grenzen einer fachgeschichtlichen Datenbank“, in: *Wissenskulturen der Musikwissenschaft. Generationen – Netzwerke – Denkstrukturen*, hrsg. von Sebastian Bolz, Moritz Kelber, Ina Knoth und Anna Langenbruch, Bielefeld 2016, S. 261–278.

struktur für Editionen (NIE-INE) implementierte Instanz von Knora/SALSAH organisiert und aufbereitet.<sup>42</sup>

WissKI beruht auf dem sogenannten Erlangen CRM / OWL (<<http://erlangen-crm.org/>>), einer eigenen auf OWL-DL zugeschnittenen Umsetzung von CIDOC CRM. Neben anderen Projekten wird mit Hilfe der Forschungsumgebung die Korrespondenz rund um die musikhistorische Instrumentensammlung von Wilhelm Rück und Söhnen in dem Projekt „Musikinstrumente sammeln – Das Beispiel Rück“ aufgearbeitet und die Forschungsdaten zum Ende der Projektlaufzeit zur Verfügung gestellt.<sup>43</sup>

### *Zurück in die Zukunft...*

Statt einer Beschränkung auf ein bestimmtes Projekt sollte in diesem Beitrag das größere Bild aufgezeigt werden, welche Möglichkeiten und Herausforderungen nicht nur auf eine digitale Musikwissenschaft, sondern auf alle von uns, im Zusammenhang mit der Modellierung von „belastbarem“ Wissen, unserem Fachwissen, jenseits von der reinen Korpus-Erstellung durch Codierung, in den folgenden Jahren unaufhaltsam auf uns zukommen werden. Die bereits 2009 von Johannes Kepper formulierte Überlegung, ob nicht „gerichtete Graphen [als welche die Grundstruktur von RDF ja betrachtet werden kann] gegenüber Baumstrukturen (und damit XML) die geeignetere Datenstruktur zur Codierung von (Noten-)Texten darstellen“<sup>44</sup>, wird nicht in diesem Artikel, sondern nur im Diskurs des musikwissenschaftlichen Fachwissens verhandelt werden können bzw. genauso davon abhängen, inwieweit es gelingt, Anwendungsbeispiele und Nachnutzungsmöglichkeiten zu schaffen, die allgemeine Verbreitung und Effektivitätsvorteile für das Fach generieren. Möglicherweise müssen sich aber auch für diese neuen Ansätze erst Wissenstrukturen transformieren, neue Allianzen entwickeln und Mechanismen etablieren, vor allem aber Mut und Experimentierfreude zugelassen und belohnt werden.

Das Rück-Projekt ist ein gutes Beispiel dafür, dass neben der Forschung und den Bibliotheken auch die Museen in solche Denk- und Wissensmodelle einbezogen werden müssen, und natürlich auch die Verlage sowie überhaupt sämtliche Gedächtnisinstitutionen und auch politischen Instanzen. Denn die Modellierung von „belastbarem Wissen“ mittels Ontologien ist per Definition eine kollektive Herausforderung, es bedarf kollaborativer, gemeinschaftlicher Anstrengungen, um diese zu bewältigen. Und nur wenn dieses Wissen in Sekundenschelle, also elektronisch, zugreif- und offen abrufbar wird, lässt sich „alternativen“ Darstellungen und Tendenzen möglicherweise zwar nicht zuvorkommen, aber zumindest schneller und offensiver begegnen. Wissen muss daher nicht nur „belastbar“, sondern auch

42 Projektwebsite unter <<http://www.anton-webern.ch>> bzw. Editionsprototyp unter <[edition.anton-webern.ch](http://edition.anton-webern.ch)>. Informationen zu NIE-INE unter <<http://www.nie-ine.ch>> sowie <<https://github.com/nie-ine>> .

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

44 Johannes Kepper, „XML-basierte Codierung musikwissenschaftlicher Daten – Zu den Voraussetzungen einer digitalen Musikedition“, in: *it – Information Technology. Methoden und innovative Anwendungen der Informatik und Informationstechnik* 51/4 (2009), S. 216–221, hier S. 220. Online unter <<https://doi.org/10.1524/itit.2009.0544>>.

unbedingt zugänglich sein. Wenn es sich hinter Bezahlschranken versteckt, werden immer alternative Darstellungen, die ohne Beschränkung erreichbar sind, subventioniert.<sup>45</sup>

Denn es geht ja auch in der Vision des Semantic Web letztendlich nicht primär darum, Maschinen das Denken beizubringen, sondern dem Menschen Zugang zu einem Wissensuniversum zu ermöglichen, das längst sämtliche individuellen Aufnahme- und Verarbeitungskapazitäten übersteigt und nur noch mit maschineller Unterstützung gefiltert werden kann. Es bleibt der Mensch im Mittelpunkt aller Bemühungen. Oder in den Worten von John Unsworth, einem digitalen Humanisten der ersten Stunde: „In some form, the semantic web is our future, and it will require formal representations of the human record. [...] we will require the rigor of computational methods in the discipline of the humanities not in spite of, but because of, the way that human understanding and human creativity violate containment, exceed representation, and muddle distinctions.“<sup>46</sup>

In diesem Sinne sei es erlaubt, mit einem kleinen Gedankenexperiment zu enden: Auf einer Zugfahrt durch Griechenland sitzen drei Wissenschaftler gemeinsam in einem Eisenbahnabteil, als auf einer Wiese draußen ein blaues Schaf zu sehen ist. Da sagt der erste Wissenschaftler: „Oh! Alle griechischen Schafe sind blau.“ Der Zweite wendet jedoch ein: „Nein! In Griechenland gibt es wenigstens eine Wiese mit blauen Schafen!“ Meldet sich der Dritte zu Wort: „Nein, nein! In Griechenland gibt es mindestens eine Wiese mit mindestens einem Schaf, das auf mindestens einer Seite blau ist...“

Egal, welche Perspektive und Beschreibungstiefe bei diesem Versuch, sich über einen gemeinsam wahrgenommenen Weltausschnitt begrifflich zu verständigen, das Wissen über diesen Weltausschnitt explizit zu artikulieren und miteinander zu teilen, der eigenen Position am nächsten kommt ... das Schaf bleibt blau.

### Abstract

The processing, dissemination, classification and verifiability of “resilient” knowledge represents some of the most pressing challenges facing society as a whole in this still young digital age. Information-theoretical ontologies that are integrated into the vision of the so-called semantic web can be of great use here. On the basis of examples of applications as well as of the explanation of basic concepts, mechanisms and challenges for the systematization and modelling of knowledge structures the current possibilities for a „semantic“ digital musicology are shown.

45 Sehr deutlich zeigt sich dies u. a. an der Online-Plattform IMSLP (International Music Score Library Project), die Digitalisate von gemeinfreien Notentexten verfügbar hält, was zu einem zumindest fragwürdigen Revival editorisch eigentlich überholter Notenausgaben führt. Vgl. Johannes Kepper und Laurent Pugin, „Was ist eine Digitale Edition? Versuch einer Positionsbestimmung zum Stand der Musikphilologie im Jahr 2017“, in: *Musiktheorie – Zeitschrift für Musikwissenschaft* 32/4 (2017), S. 347–363, hier S. 347.

46 John M. Unsworth, „What is Humanities Computing and What is Not?“, in: *Jahrbuch für Computerphilologie* 4 (2002). Online unter <<http://computerphilologie.uni-muenchen.de/jg02/unsworth.html>>.