

Stefan Gradmann, Julia Iwanowa, Evelyn Dröge, Steffen Hennicke, Violeta Trkulja, Marlies Olensky, Christian Stein, Alexander Struck, Konstantin Baierer, Berlin und Leuven (Belgien)

## Modellierung und Ontologien im Wissensmanagement

Erfahrungen aus drei Projekten im Umfeld von Europeana und des DFG-Exzellenzclusters Bild Wissen Gestaltung an der Humboldt-Universität zu Berlin

Im Artikel werden laufende Arbeiten und Ergebnisse der Forschergruppe Wissensmanagement beschrieben. Diese entstanden vor allem durch die am Lehrstuhl Wissensmanagement angesiedelten Projekte *Europeana v2.0*, *Digitised Manuscripts to Europeana (DM2E)* sowie von Teilprojekten des vor kurzem gestarteten DFG-Exzellenzclusters *Bild Wissen Gestaltung*. Die Projekte befassen sich mit Spezialisierungen des *Europeana Data Model*, der Umwandlung von Metadaten in RDF und der automatisierten und nutzerbasierten semantischen Anreicherung dieser Daten auf Basis eigens entwickelter oder modifizierter Anwendungen sowie der Modellierung von Forschungsaktivitäten, welche derzeit auf die digitale Geisteswissenschaft zugeschnitten ist. Allen Projekten gemeinsam ist die konzeptionelle oder technische Modellierung von Informationsentitäten oder Nutzeraktivitäten, welche am Ende im Linked Data Web repräsentiert werden.

**Schlagwörter:** Wissensmanagement, Digitale Geisteswissenschaft, Ontologie, Linked (Open) Data, Modellierung, Datentransformation, Scholarly Domain Model, Europeana Data Model

**Modeling and Ontologies in Knowledge Management. Lessons learned in three research projects in the context of Europeana and the Excellence Cluster at Humboldt-Universität zu Berlin.**

The contribution reports on the work of the research unit Knowledge Management as well as on the results obtained so far as part of the projects affiliated with the chair Knowledge Management such as *Europeana v2.0*, *Digitised Manuscripts to Europeana (DM2E)* and of base projects as part of the recently started DFG excellence cluster *Knowledge Image Gestaltung*. The projects deal with specialising the *Europeana Data Model*, the transformation of metadata to RDF as well as the automated and user generated semantic enrichment of such data using newly developed or existing

and modified applications as well as the modeling of research activity currently focusing on the Digital Humanities domain. All projects share the technical and conceptual modeling of information entities and user activities with the ultimate goal to be represented on the Linked Data Web.

**Keywords:** knowledge management, Digital Humanities, ontologies, Linked (Open) Data, modeling, data transformation, Scholarly Domain Model, Europeana Data Model

**Modélisation et ontologies dans la gestion des connaissances. Expériences dans trois projets de recherche dans le contexte d'Europeana et du Cluster d'excellence à la « Humboldt-Universität zu Berlin »**

Cette contribution décrit les travaux du groupe de recherche « gestion des connaissances » ainsi que les résultats obtenus jusqu'alors dans des projets tels que *Europeana v2.0*, *Digitised Manuscripts to Europeana (DM2E)* ou encore les débuts des projets de base faisant partie du Cluster d'excellence (DFG) *Bild Wissen Gestaltung*. Ces projets produisent et étudient des spécialisations du *Europeana Data Model*, c'est à dire la transformation des métadonnées en RDF et l'enrichissement sémantique automatisé / par l'utilisateur des dites métadonnées à l'aide des applications développées ou encore existantes mais modifiées dans les projets ainsi que la modélisation des activités des recherche. Modélisation, qui pour le moment, est ciblée sur les SSH numériques. Ces projets partagent une approche de modélisation conceptuelle et technique des activités de recherche ou d'utilisateurs avec comme but final d'être représenté sur la toile de données avec lesquels ils sont liés.

**Mots-clés:** gestion des connaissances, SSH numériques, ontologies, données liées, modélisation, transformation des données, Scholarly Domain Model, Europeana Data Model

## 1 Modellierung und Verstehen

Zu den Bedingungen des Verstehens in semantischen Netzen (Gradmann, 2009) – so lautete der Untertitel der am 28. Oktober 2008 gehaltenen Antrittsvorlesung von Stefan Gradmann, mit der er programmatisch seine Lehrtätigkeit an der Humboldt-Universität zu Berlin begann und zugleich den Lehrstuhl Wissensmanagement positionierte, aus dem dann die gleichnamige Forschergruppe hervorging. Die drei Schlüsselwörter des Hauptsachtitels waren dabei *Signal*, *Information* und *Zeichen* – doch wäre diesen drei Begriffen aus der heutigen, erweiterten Sicht der Forschergruppe Wissensmanagement als viertes das Wort *Modellierung* hinzuzufügen.

Denn – so das Ergebnis der am 1. und 2. April 2011 am Institut für Bibliotheks- und Informationswissenschaft veranstalteten Tagung der Gesellschaft für Wissenschaftsforschung mit dem Titel *Digital Humanities – Wissenschaften vom Verstehen*<sup>1</sup> – ein für geisteswissenschaftliche Forschung, für Interpretation und Verstehen fundamentaler Akt ist eben das Modellieren von Welt, und zwar in der vollen Bandbreite des Begriffs wie in dem später vorgestellten Projekt *Bild Wissen Gestaltung* (BWG) und im ebenfalls unten dargestellten *Scholarly Domain Model*. Also im Sinne von Abbildung ebenso wie von Reduktion und Abstraktion: Ein Modell bildet einen Teil der Welt (oder, kantianisch gesprochen, eine Sicht derselben) immer ein Stück weit ab – jedoch niemals vollständig und unter Berücksichtigung all seiner Aspekte. Eine Modell-eisenbahn ist deutlich kleiner als das Original, versucht aber, möglichst viele Detailspekte in der Verkleinerung abzubilden, während etwa Atome ganz sicher nicht so aussehen, wie es das *Bohrsche Atommodell* nahelegt. Doch die Bandbreite des Begriffs ist noch größer: Modelle können präskriptiven Charakter haben als Anleitungen für die Konstruktion von *Welt*, wie dies etwa bei den meisten Architekturmodellen der Fall ist. Und schließlich sind viele Modelle nicht nur weniger als das Abgebildete, sondern paradoxerweise auch mehr, beinhalten einen mitunter massiven metaphorischen Überschuss, wie etwa im Falle des schon zitierten *Bohrschen Atommodells*, dessen Überzeugungskraft und Beharrungsvermögen trotz seiner Fehler (im Sinne der Abbildrelation) vielleicht vor allem deshalb so beeindruckend sind, weil es von der metaphorischen Aura des heliozentrischen Weltmodells zehrt, in dem ebenfalls kleinere Kugeln um einen größeren „Kern“ kreisen.

<sup>1</sup> Tagungsprogramm der „Digital Humanities: Wissenschaften vom Verstehen“. <http://www.wissenschaftsforschung.de/asstag.html#Tagung2011> [03.03.2013].

Der Akt der Modellierung von Weltausschnitten im Kontext des *Verstehens*, der Interpretation nun zieht sich als roter Faden durch die drei im Folgenden dargestellten Projekte der Forschergruppe Wissensmanagement. Sei es als Strategien der objekt- und ereigniszentrierten Modellierung (und dabei zentriert durch das *Europeana Data Model* (EDM))<sup>2</sup> wie in *Digitised Manuscripts to Europeana* (DM2E)<sup>3</sup> und *Europeana v2.0*<sup>4</sup>, sei es als der Versuch, Prozesse des Forschens und der Wissensgenerierung selbst zu modellieren wie (wiederum) in DM2E sowie den Projekten *Virtuelle und reale Architektur des Wissens*<sup>5</sup> und *Shaping Knowledge*<sup>6</sup> als Teil des Exzellenzclusters *Bild Wissen Gestaltung* (BWG).<sup>7</sup> Auch zahlreiche am Lehrstuhl Wissensmanagement entstandene und entstehende Dissertationsprojekte sind in diesem Zusammenhang zu nennen, die Titel der teilweise schon eingereichten, jedoch noch nicht veröffentlichten Dissertationen, lauten zum Beispiel *Signification and data modelling – On the common semiotic patterns underlying practises of data modelling in cultural heritage communities*, *Metadatenbasierte Kontextualisierung architektonischer 3D-Modelle* oder *Describing Data Patterns: A general deconstruction of metadata standards*.

Die Grafik in Abbildung 1 illustriert die Schnittstellen und Überlappungen der drei Forschungsvorhaben *Europeana v2.0*, *DM2E* und *Bild Wissen Gestaltung* und verdeutlicht die Zusammenhänge der einzelnen Arbeitspakete untereinander. Basis für die Entwicklungen in allen drei Projekten, und hier im Mittelpunkt der Abbildung dargestellt, sind das *Resource Description Framework* (RDF)<sup>8</sup> sowie das EDM<sup>9</sup>, das in *Europeana* entstandene Datenmodell für die Repräsentation von digitalisiertem und genuin digitalem Kulturgut im Kontext des *Linked Data Web*. RDF ist ein vom W3C<sup>10</sup> herausgegebener Standard zur Repräsentation strukturierter Daten im Netz.

<sup>2</sup> EDM-Dokumentation. <http://pro.europeana.eu/edm-documentation> [03.03.2013].

<sup>3</sup> Digitised Manuscripts to Europeana. <http://www.dm2e.eu> [03.03.2013].

<sup>4</sup> Europeana v2.0. <http://pro.europeana.eu/web/europeana-v2.0> [03.03.2013].

<sup>5</sup> Virtuelle Architektur des Wissens. <http://www.interdisciplinary-laboratory.hu-berlin.de/de/Virtuelle-%26-reale-Architektur-des-Wissens> [03.03.2013].

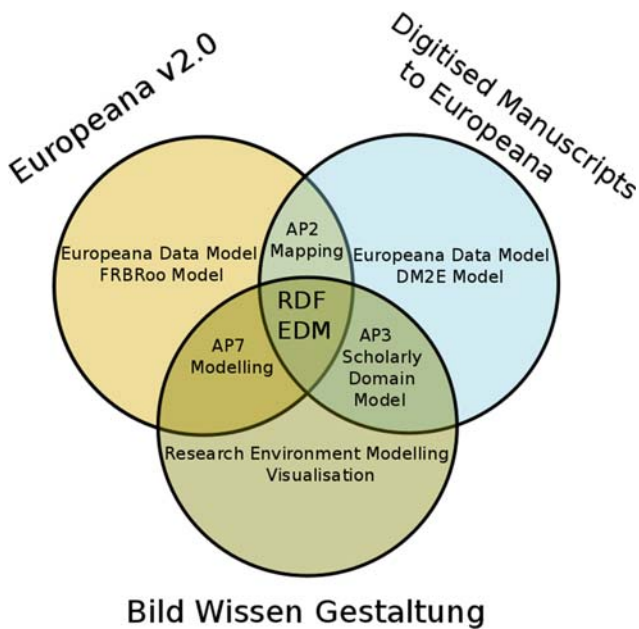
<sup>6</sup> Shaping Knowledge. <http://www.interdisciplinary-laboratory.hu-berlin.de/de/Shaping-Knowledge> [03.03.2013].

<sup>7</sup> Bild Wissen Gestaltung. <http://www.interdisciplinary-laboratory.hu-berlin.de/de> [03.03.2013].

<sup>8</sup> Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation 10 February 2004: <http://www.w3.org/TR/rdf-concepts> [03.03.2013].

<sup>9</sup> Weitere Details zum EDM s. Abschnitt 2.

<sup>10</sup> WWW Consortium: <http://www.w3.org> [03.03.2013].



**Bild Wissen Gestaltung**

**Abb. 1:** Interaktion der aktuellen Projekte der Forschergruppe Wissensmanagement: Europeana v2.0, Digitised Manuscripts to Europeana (DM2E) und Bild Wissen Gestaltung (BWG).

Das EDM verwendet diesen Standard und ist für weitere Spezialisierungen vorgesehen, welche in Abschnitt 3 näher erläutert werden. Die nachfolgend beschriebenen Forschungsprojekte erarbeiten Konzepte und entwickeln technische Lösungen für Prototypen, die solche strukturierte Daten im Sinne von strukturiertem Wissen definieren, sammeln, repräsentieren und vernetzen.

Die Ergebnisse aus dem Projekt Europeana v2.0 dienen als Grundlage für die Weiterentwicklung des EDM und sind Gegenstand der Aktivitäten im Arbeitspaket 2 *Interoperability Infrastructure* (AP2) des DM2E-Projekts. Im Rahmen von DM2E wird eine Spezialisierung des EDM für Handschriften und Manuskripte, das *DM2E-Modell*, entwickelt. Das Arbeitspaket 3 in DM2E *Digital Humanities Requirements and Related Engineering* (AP3) entwickelt eine digitale Arbeitsumgebung für Forschungen an handschriftlichen Manuskripten und befasst sich mit der Ausarbeitung des *Scholarly Domain Models*, einem formalisierten Modell geisteswissenschaftlicher Forschungsaktivitäten. Die hier entwickelten Modellierungsstrategien sind ein wesentlicher Bestandteil des Systementwurfs im Clusterprojekt *Bild Wissen Gestaltung* mit dem Ziel, eine komplexe experimentelle Beobachtungs- und Arbeitsumgebung für das interdisziplinäre Labor des Exzellenzclusters zu konzipieren und zu implementieren (Details hierzu unter Abschnitt 4). Die parallel stattfindenden Entwicklungen in allen Projekten ergänzen sich gegenseitig und das detaillierte Zusammenwirken der einzelnen Komponenten soll im Folgenden näher erläutert werden. Das Basisprojekt Eu-

ropeana v2.0 wird im Abschnitt 2 beschrieben und die Aktivitäten aus dem DM2E-Projekt werden im darauf folgenden Abschnitt 3 vorgestellt. Anschließend wird in Abschnitt 4 das BWG-Vorhaben dargestellt.

Im Sinne der besseren Lesbarkeit wird im Folgenden nur die maskuline Form bei Personenbezeichnungen verwendet. Die feminine Form ist dabei immer gleichberechtigt mitgedacht.

## 2 Europeana v2.0

Im Rahmen des *Europeana* Projektclusters gab es bisher eine Vielzahl von Projekten. Im Mittelpunkt standen bzw. stehen die Hauptprojekte *Europeana v1.0* sowie das derzeitige Projekt *Europeana v2.0*. Deren Aufgabe ist es, Infrastrukturkomponenten und Bausteine der Kernfunktionalitäten zu entwickeln sowie Europeana als Organisation aufzubauen. Im Projekt Europeana v2.0 geht es vor allem um die Verbesserung, Erweiterung und erleichterte Wiederverwendung von Inhalten sowie den Ausbau von Funktionalitäten und die damit verbundene Verbesserung der *User Experience*.

Erklärtes Ziel von Europeana war es immer, mehr als nur ein Objektkatalog des europäischen digitalen bzw. digitalisierten Kulturerbes zu sein (Concordia, Gradmann & Siebinga, 2010). Europeana sollte multilingualen und vor allem semantikbasierten Zugang zu den Objekten ermöglichen und damit das Entstehen und Verstehen von neuem Wissen unterstützen. Das Verstehen von Objekten außerhalb ihres Ursprungskontexts (z. B. kann ein Objekt aus der Museumswelt mit anderen Objekten aus Bibliotheken oder Archiven in Beziehung gesetzt werden) setzte allerdings eine gut strukturierte Modellierung von Objekten und Kontext voraus, die mit dem *Europeana Data Model* umgesetzt wurde. Ursprünglich wurden die Metadaten der Datenlieferanten von Europeana in die sogenannten *Europeana Semantic Elements (ESE)*<sup>11</sup>, die ihre Bezeichnung zu Unrecht trugen, übersetzt. Semantisch war an diesen Elementen nur, dass ein möglichst sinnvoller Abgleich der Metadatenelemente erfolgte, denn dem Modell lag das relativ flache Dublin-Core Modell<sup>12</sup> zugrunde. Die in den Originalmetadaten vorhandene Semantik ging dadurch allerdings verloren (so etwa die in vielen Datenquellen ursprünglich vorhandenen Verlinkungsinformationen) und es war klar, dass die ESE in Zu-

<sup>11</sup> ESE Spezifikation (2012). <http://pro.europeana.eu/documents/900548/dc80802e-6efb-4127-a98e-c27c95396d57> [03.03.2013].

<sup>12</sup> Dublin Metadata Element Set, Version 1.1 (2012). <http://dublincore.org/documents/dces/> [03.03.2013].

kunft nicht ausreichen würden. Die Kooperation zweier Arbeitspakete aus dem Europeana-Projektcluster (Arbeitspaket 3 aus *Europeana v1.0* und Arbeitspaket 1 aus *EuropeanaConnect*) entwickelte mit Hilfe von und in Abstimmung mit Vertretern der Bibliotheks-, Museums- und Archivlandschaft das EDM. Es erlaubt die Spezialisierung der Klassen und Eigenschaften im Sinne granularer Modelle sowie Verweise von Objekten innerhalb Europeanas aber auch die Verlinkung zu externen Informationsressourcen. Erst dadurch wurden „semantische Nutzungsszenarien“ überhaupt denkbar und der Weg für die Nutzung Europeanas innerhalb der Linked Open Data Cloud (Heath & Bizer, 2011) geebnet. Darüber hinaus ist es möglich, im EDM zwischen dem realen Objekt und seiner digitalen Repräsentation sowie zwischen dem Objekt und den beschreibenden Metadaten zu unterscheiden. Außerdem werden mehrere Sichten auf ein- und dasselbe Objekt unterstützt, die durchaus auch widersprüchliche Aussagen enthalten können. Eines der wichtigsten Prinzipien des EDM ist überdies die maximale Wiederverwendung existierender Standards (wie SKOS, OAI-ORE und DCTerms).

Das EDM ist keineswegs ein statisches Modell. Zusätzlich zur Möglichkeit der Spezialisierung der Klassen und Eigenschaften mittels anderer Metadatenstandards, wird an der Weiterentwicklung des Modells gearbeitet, um eine möglichst hohe Akzeptanz und Anwendung im Bereich des digitalen Kulturerbes zu erreichen. An dieser Weiterentwicklung gibt es auch außerhalb von Europeana Interesse: so hat sich etwa die Initiative *Digital Public Library of America* (DPLA) stark am EDM orientiert und der Gründer der vorbildhaften digitalen Bibliothek *Perseus*, Greg Crane, hat in Crane et al. (2012) großes Interesse an dieser Weiterentwicklung bekundet.

Im Rahmen des Projekts Europeana v2.0 wird im Arbeitspaket 7 *Innovation* (AP7) unter anderem das EDM weiterentwickelt. Weitere Aufgaben umfassen die (Weiter)Entwicklung innovativer Applikationen, die auf Europeana-Inhalten aufbauen, den Nutzer involvieren und ihn motivieren, Inhalte in Form von Metadaten, Beschreibungen, Kommentaren etc. beizutragen und diese Inhalte dann auch semantisch weiterzuverarbeiten sowie die (Weiter)Entwicklung multilingualer Interaktionsmodelle. Ausgangspunkt für die Verfeinerungsarbeit am EDM sind Empfehlungen, die am Ende des Projekts Europeana v1.0 formuliert wurden (Dekkers, Gradmann & Molendijk, 2011) und ein standardisiertes *Application Profile* fordern, das eine Abbildung des FRBRoo-Modells (*FRBR object-oriented*) in EDM ermöglicht. Das FRBRoo-Modell ist eine Initiative der Bibliotheks- und Museumswelt, um den Austausch von deren (bibliographischer) Information bzw. Metadaten zu vereinfachen, und geht davon aus, dass jedes Aggregat

eine selbständige Entität darstellt. Aus diesem Grund gründete AP7 die *Europeana Tech Taskforce EDM-FRBRoo Application Profile*<sup>13</sup>, kurz EFAP-TF genannt. Diese befasst sich mit der Harmonisierung der beiden Modelle und legt dabei Prinzipien für die Modellierung und Abbildung der Gruppe-1-Entitäten von FRBRoo (Werk, Expression, Manifestation und Exemplar) in EDM zugrunde. Diese Abbildung wird derzeit anhand von Beispieldaten illustriert.

Die EFAP-TF wird von Stefan Gradmann und Martin Doerr geleitet und umfasst 16 Mitglieder aus den Bereichen Bibliotheken, Museen, Archiven und Universitäten. Die Arbeit der EFAP-TF basiert auf einem Entwurf des CIDOC (*Comité international pour la documentation*), das ein harmonisiertes EDM-CRM-FRBRoo-Modell<sup>14</sup> vorschlägt. Die Gruppe bearbeitet vier Beispiele, für die sie ein Mapping zwischen EDM und FRBRoo erstellt. Bibliographische Datensätze rund um die Werke Don Quixote und Hamlet sind der Beitrag aus der Bibliothekswelt, ein weiteres Beispiel behandelt ein musikalisches Werk und das vierte Beispiel eine Posterkollektion. Alle Beispiele werden zunächst in EDM sowie in FRBRoo modelliert, und danach wird ein Mapping bzw. ein gemeinsames Modell erstellt. Darüber hinaus werden Forschungsfragen formuliert, die im Zusammenhang der jeweiligen Beispiele relevant sind, und über die festgestellt werden soll, wie weit die Modellierung der Daten in FRBRoo abstrahiert werden kann und dabei immer noch die Beantwortung der Fragen zulässt. Im Beispiel Don Quixote wäre hier die Überprüfung, welche Übersetzungen oder Adaptionen für die beiden Teile des Don Quixote existieren, zu nennen. Daraus lässt sich hernach das Set an notwendigen FRBRoo-Erweiterungen für EDM ableiten, das immer noch die Details der in den Objekten vorhandenen Informationen akkurat darstellt. Vorläufige Ergebnisse der Taskforce umfassen die Erkenntnis, dass die im FRBR-Umfeld viel diskutierte Unterscheidung zwischen Werk und Expression nur in Ausnahmefällen notwendig ist und dass dafür klare Richtlinien formuliert werden und die Feststellung, dass die Klasse Exemplar (*frbr:item*) mittels *edm:PhysicalThing* ausreichend repräsentiert werden kann. Der Endbericht der EFAP-TF wird voraussichtlich Ende April 2013 veröffentlicht.

Folgend den Empfehlungen aus Europeana v1.0 wird die domänenspezifische (Weiter)Entwicklung und Spezia-

<sup>13</sup> Task Force EDM FRBRoo. <http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+Force+EDM+FRBRoo> [03.03.2013].

<sup>14</sup> Doerr, M. (2011). Graphical representation of the harmonized EDM-CRM-FRBRoo-DC-ORE model. [http://www.cidoc-crm.org/docs/EDM-DC-ORE-CRM-FRBR\\_Integration\\_ORE\\_fix.ppt](http://www.cidoc-crm.org/docs/EDM-DC-ORE-CRM-FRBR_Integration_ORE_fix.ppt) [03.03.2013].

lisierung des EDM durch mehrere Europeana-Satelliten-Projekte fortgeführt. Das Modell wird u. a. im Rahmen des Projektes *Europeana Fashion*<sup>15</sup> für die Modedomäne angereichert und spezifiziert. Innerhalb des nachfolgend beschriebenen Projekts DM2E wird u. a. an einer Spezialisierung des EDM zur optimalen Beschreibung handschriftlicher Manuskripte, deren Digitalisaten, Faksimiles und Transkriptionen gearbeitet. In Abschnitt 3.1 wird die Herangehensweise dazu erläutert.

### 3 Digitised Manuscripts to Europeana (DM2E)

Im Zeitalter von Social Web, Digital Science und Linked Data ist die Idee einer vernetzten kollaborativen Arbeitsumgebung für die Human- und Geisteswissenschaften keine Fiktion mehr. In zahlreichen Projekten aus dem Bereich der digitalen Geisteswissenschaften, wie zum Beispiel *ResearchSpace*<sup>16</sup> und *Isidore*<sup>17</sup>, werden innovative Arbeitsmodelle für die Wissenschaftsgemeinschaft erarbeitet, die durch die Entwicklung mit neuen Webtechnologien unterstützt werden. Die fortschreitende Entwicklung des WWW von einem Netz der Dokumente, bestehend aus Webseiten mit unterschiedlichen Inhalten wie Texten, Bildern und Videos, zu einem Netz der Daten, mit einzelnen maschinell dereferenzierbaren Einheiten als neuer Form von Webseiten oder als Teile solcher, bedingt, dass viele der bereits vorhandenen Dokumente und Inhalte für das Web der Daten entsprechend angepasst werden müssen. In Bibliotheken, Archiven und Museen sind es außerdem die oft über Jahre hinweg für die meist lokale Nutzung katalogisierten und gepflegten Sammlungen, die nun im Linked Data Web veröffentlicht werden (Hyvönen, 2012). Hier gilt es auf der einen Seite das wertvolle intellektuelle Wissen und die Indexierungskompetenz solcher Institutionen nicht nur zu bewahren, sondern ganz besonders für die breite Öffentlichkeit transparent zu machen, und auf der anderen Seite aber auch die ökonomischen Vorteile einer verteilten Arbeitsumgebung auf Basis dieser Ressourcen für Wissenschaftler attraktiv zu gestalten. In diesen Diskurs lassen sich die Arbeitsschwerpunkte des DM2E-Projektes einordnen.

<sup>15</sup> Europeana Fashion. <http://www.europeanafashion.eu> [03.03.2013].

<sup>16</sup> ResearchSpace. <http://www.researchspace.org> [03.03.2013].

<sup>17</sup> ISIDORE – Accès aux données et services numériques de SHS. <http://www.rechercheisidore.fr> [03.03.2013].

*Digitised Manuscripts to Europeana* ist ein EU-gefördertes Projekt im Umfeld von Europeana. Wissenschaftler und Praktiker erarbeiten gemeinschaftlich mit den im Projekt involvierten Datenlieferanten (wie der Österreichischen Nationalbibliothek, dem Wittgenstein Archiv der Universität Bergen, der European Association for Jewish Culture u. a.) und unter Berücksichtigung zusätzlicher Nutzeranforderungen prototypische und exemplarische Lösungsansätze zur Vernetzung von verteilten heterogenen Digitalisaten und deren Metadaten sowie deren Aufbereitung und Bereitstellung für (wissenschaftliche) Annotationen. Das Projekt hat zwei übergeordnete Ziele, die Gegenstand zweier Arbeitspakete in DM2E sind: Zum einen sollen die Metadaten möglichst vieler Datenlieferanten als Linked Data zur Verfügung gestellt und in Europeana integriert werden und zum anderen eine Arbeitsumgebung für die geisteswissenschaftliche Forschung erstellt werden.

Da viele Datenlieferanten ihre Daten in unterschiedlichen Formaten zur Verfügung stellen, wird als Teil des ersten Projektziels ein Werkzeug entwickelt, mit welchem Meta- und Objektdaten aus einer Vielzahl von Quellformaten in das EDM überführt werden können. Dieses Werkzeug besteht aus unterschiedlichen Transformations- und Kontextualisierungstools (wie etwa *MINT* aus der *Europeana*-Entwicklung oder *D2R* aus dem *LOD2*-Projektkontext), die für das Projekt angepasst und zusammengeführt werden. Auf diese Tools wird im folgenden Abschnitt genauer eingegangen. DM2E greift für die Erstellung des Werkzeuges auf Forschungsergebnisse aus *Europeana v1.0* und *Europeana v2.0* zurück. Weiterhin werden Open-Source-Lösungen von an DM2E beteiligten Institutionen einbezogen, wie beispielsweise das Kontextualisierungstool *Silk*. Durch das so entwickelte Werkzeug wird die Integration heterogener Metadaten in *Europeana* und deren Aufbereitung für das Linked Data Web vereinfacht. Die Entwicklung geht einher mit der Erstellung einer providerorientierten Spezialisierung des EDM für die Domäne der Manuskripte. Diese Spezialisierung wird vom Transformationstool bei der Datentransformation als Mappingschema in das Zieldatenmodell verwendet und ermöglicht es so, die gelieferten Daten ohne Bedeutungsverluste in RDF abzubilden. Anschließend werden die Daten kontextualisiert und durch Nutzer weiterbearbeitet.

Zu diesem Zweck wird im Rahmen des zweiten Projektziels eine Arbeitsumgebung eingerichtet, die die Nutzer bei der Erforschung handschriftlicher Manuskripte unterstützt. Bisher erlaubt diese Anwendung semantische Annotationen von Textinhalten und Digitalisaten in Form von RDF durch den Nutzer und damit letztlich die Erstel-

lung ergänzender Aussagen. Die Arbeitsumgebung ermöglicht im Kontext der Forschungen an handschriftlichen Manuskripten den Aufbau von Aussagensystemen als sozialen semantischen Graphen. Die Entwicklung der Arbeitsumgebung orientiert sich darüber hinaus am *Scholarly Domain Model*, welches ein formalisiertes Modell geisteswissenschaftlicher Forschungsaktivität darstellt. Das Modell wird unter anderem durch ein im Projekt angesiedeltes *Digital Humanities Advisory Board* (DHAB) konzipiert, welches aus prominenten Vertretern der digitalen Geisteswissenschaften besteht.

Über diese Ziele hinaus möchte DM2E auch zum Aufbau einer Community für die digitalen Geisteswissenschaften beitragen. Diese ist Zielgruppe für die im Projekt erstellte Arbeitsumgebung und wird in die Entwicklung mit einbezogen. Weiterhin hält das Projekt Kontakt zu verwandten internationalen Großprojekten wie DARIAH oder auch ResearchSpace, um sich über verwandtes Forschungsvorgehen austauschen zu können.

### 3.1 Modellierungs- und Transformationsprozesse als Teil einer interoperablen Infrastruktur

Die wachsende Popularität von Europeana<sup>18</sup> verstärkt die Notwendigkeit weiterer Forschung auf dem Gebiet der Dateninteroperabilität und -modellierung, um möglichst effektiv und zugleich semantisch korrekt mit der Heterogenität der in Europeana gesammelten Meta- und Objektdaten sowohl für den menschlichen Nutzer, als auch für deren maschinelle Weiterverarbeitung umgehen zu können. Teil des Projekts sind die Konzeption und (Weiter) Entwicklung von Anwendungen für die RDFisierung<sup>19</sup> und die Überführung von Daten in das spezialisierte EDM, deren dauerhafte Speicherung in Datenbanken, die für die Speicherung von RDF-Daten ausgelegt sind, (sog. Triplestores), sowie die Kontextualisierung dieser Metadaten mit Ressourcen im Linked Data Web, d. h. die Anreicherung der Objektrepräsentation mit Kontextlinks im Sinne des vierten Linked-Open-Data-Prinzips von Berners-Lee<sup>20</sup>. Um den Benutzer von der Last zu befreien, all

diese Werkzeuge lokal installieren, die Eingabe- und Konfigurationsdaten selbst verwalten und immer wiederkehrende Abläufe manuell abarbeiten zu müssen, wird die Web-Plattform *OmNom* entwickelt. *OmNom* ist zunächst eine Sammlung von semantischen Web Services, die als einfache Schnittstelle zu den Werkzeugen fungieren und das verteilte Speichern von Eingabe- und Konfigurationsdaten für diese Werkzeuge erlaubt. Darüber hinaus ermöglicht *OmNom*, verschiedene Werkzeuge mit verschiedenen Konfigurationen zu beliebig komplexen Prozessketten (sog. Workflows) zusammenzuschalten. Diese Workflows müssen nur einmal definiert und können dann auf beliebige Eingabedaten angewandt werden, sodass Nutzer der Plattform alle Schritte des Transformationsprozesses, von kleineren Datenkorrekturen über die Validierung der Daten und deren RDFisierung bis hin zur Veröffentlichung der Daten in einem Triplestore, iterativ und experimentell an einem Beispieldatensatz entwickeln und dann ohne Veränderungen auf ihren gesamten Datenbestand anwenden können. Aus Sicht des Benutzers ist *OmNom* eine dynamische, Javascript-basierte Web-Applikation, die es ihm ermöglicht, im Browser seine Daten zu verwalten, Werkzeuge anzusprechen und Workflows per Drag-and-Drop zu erstellen.

Im RDFisierungsprozess selbst kommen derzeit die Tools *Mint*, *D2R* oder manuell erstellte XSLT-Skripte zum Einsatz. Die Verwendung weiterer Anwendungen ist geplant und teilweise implementiert, aber noch nicht ausreichend getestet. *Mint* ist eine von der National Technical University Athens entwickelte Web-Anwendung und ermöglicht die Erstellung von Mappings per Drag-and-Drop von diversen Datenformaten wie XML zu EDM oder anderen Modellen, wie etwa der Spezialisierung des EDM von DM2E (Kollia, Tzouvaras et al., 2012). *D2R* dient ebenfalls der RDFisierung, indem es Umwandlungen von Inhalten relationaler Datenbanken zu RDF ermöglicht (Bizer & Cyganiak, 2006). Geplant ist darüber hinaus die Datenkontextualisierung mit *Silk*, das wie *D2R* an der Freien Universität Berlin entstanden ist. Mit *Silk* können die gemappten Daten zusätzlich über Ressourcen anderer Linked-Data-Vokabulare angereichert werden (Volz, Bizer et al., 2009). In das EDM integrierte Daten werden von *OmNom* in einem Triplestore veröffentlicht und sind von dem Moment an online referenzierbar. Zusätzlich werden Provenienzinformationen über den gesamten Transformationsprozess gespeichert, wie beispielsweise die verwendete Werkzeuge, der Name des *OmNom*-Nutzers oder die verwendete Konfiguration, um die Ergebnisse reproduzieren zu können und die Versionierung der Daten zu ermöglichen. Mithilfe einer angepassten Ver-

<sup>18</sup> Europeana Web Traffic Report – Q3 2012. <http://pro.europeana.eu/documents/858566/1415274/Europeana+Web+Traffic+Report+Q3+2012?version=1.0> [31.01.2013].

<sup>19</sup> RDFisierung bezeichnet die Transformation von Rohdaten (z. B. XML, MARC, CSV) in eine RDF-Struktur.

<sup>20</sup> Berners-Lee, T. (2006) Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html> [03.03.2013].

Abb. 2: Eine vollständige Prozesskette („Workflow“) in der Ingestion-Plattform *OmNom*.

sion des Linked-Data-Frontends *Pubby*<sup>21</sup> werden die Daten aus dem Triplestore schließlich als Linked Data ausgeliefert und in einer für den menschlichen Nutzer übersichtlichen HTML-Repräsentation dargestellt. In dieser Ansicht ist es möglich, sämtliche Metadaten zu inspizieren, auch solche, die durch die Kontextualisierung hinzugewonnen wurden. Darüber hinaus ist es möglich, innerhalb der Manuskripte zu navigieren, Faksimiles einzelner Seiten oder ganzer Werke zu betrachten und sich anhand von Querverweisen durch den gesamten DM2E-Datenbestand zu bewegen. Die HTML-Repräsentation ist aber bewusst einfach gehalten, da die tiefere Beschäftigung und insbesondere die Interaktion mit den Daten im Sinne des *Scholarly Domain Model* mit den Werkzeugen erfolgen soll, die in AP3 entwickelt werden.<sup>22</sup>

Eine vollständige *OmNom*-Prozesskette wird in Abbildung 2 dargestellt. In der linken Spalte stehen die verfü-

baren Komponenten, die in den Workflow in der mittleren Spalte eingefügt werden können. In der rechten Spalte werden die URIs<sup>23</sup> von Web-Ressourcen aufgelistet (bspw. URIs von Eingabedaten, XSLT-Skripten oder XSD-Schemas), die per Drag-and-Drop als Parameter der Komponenten verwendet werden können. Im dargestellten Beispiel wird eine XML/TEI-kodierte Handschrift von Ludwig Wittgenstein mithilfe einer Sammlung von XSLT-Skripten, die in einem ZIP-Archiv zusammengefasst sind, in das DM2E-Modell transformiert und somit zugleich RDFisiert und in einem Triplestore veröffentlicht.

Die Verbindung bereits vorhandener Tools durch *OmNom* bildet den Grundstein für eine funktionierende interoperable Infrastruktur, die aus Sicht der Forschergruppe in der semantischen Aufbereitung, der Transformation sowie der Kontextualisierung heterogener Ressourcen liegt. Das Arbeitspaket 2 *Interoperability Infrastructure* (AP2) entwickelt Lösungen für die automatisierte Aufbereitung und Transformationsszenarien für unterschiedliche Da-

<sup>21</sup> Bizer, C. & Cyganiak, R. Pubby. A Linked Data Frontend for SPARQL. <http://wifo5-03.informatik.uni-mannheim.de/pubby> [03.03.2013].

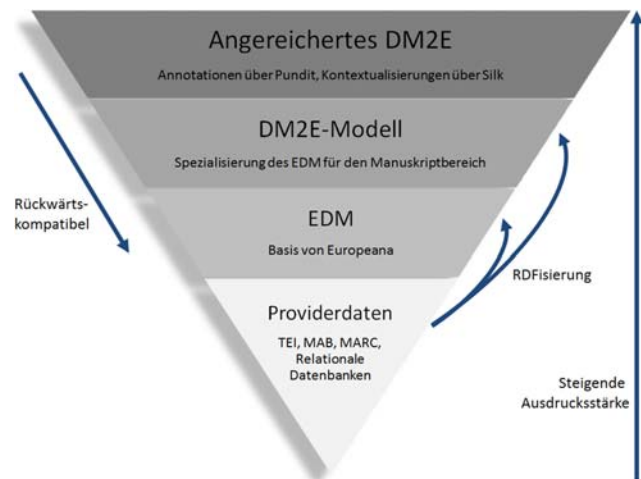
<sup>22</sup> Siehe hierfür Abschnitt 3.2.

<sup>23</sup> Naming and Addressing: URIs, URLs, .... <http://www.w3.org/Addressing> [03.03.2013].

tenformate. Repräsentationsformate von Datenlieferanten in DM2E sind MAB und MAB2, EAD, MARC, METS/MODS, TEI<sup>24</sup> und lieferantenspezifische Datenbankschemata. Die Metadaten der Digitalisate, die mit einem dieser sehr unterschiedlichen Formate ausgezeichnet sind, lassen sich in eine RDF-Struktur transformieren. Das ist notwendig, um Web-Ressourcen zu erhalten, die mit einer eindeutigen URI identifizierbar und im Web referenzierbar sind. In DM2E fällt der Schritt der Überführung nach RDF mit dem Mapping auf ein eigenes Datenmodell zusammen, welches das EDM spezialisiert. Die Spezialisierung des EDM ist notwendig, um alle Aspekte der Providermetadata in Linked Data überführen zu können. Das EDM ist ein bewusst generisch gehaltenes Modell, das die Breite der Beschreibungsmöglichkeiten für die Repräsentation von Objekten im Bereich des Kulturerbes abdeckt. Sind spezifischere Ausdrucksmöglichkeiten notwendig, wie etwa für die technische Spezifikation von Videoformaten oder für die Differenzierung unterschiedlicher Ereignistypen, so kann das EDM um diese ergänzt werden. In DM2E findet diese Erweiterung in die Tiefe auf Basis der gelieferten Daten statt, welche Manuskripte beschreiben. Das so entstandene DM2E-Modell ist eine OWL-Ontologie, deren aktueller Stand unter <http://data.dm2e.eu/dm2e> eingesehen werden kann.

Abbildung 3 gibt einen Überblick über die verschiedenen Datenformate, die im Projekt behandelt werden. An unterster Stelle der Grafik sind die Providerdaten aufgeführt. Stellvertretend für die unterschiedlichen Formate sind hier die im Bibliotheksbereich häufig verwendeten Metadatenformate MAB und MARC, der in den digitalen Geisteswissenschaften weit verbreitete Auszeichnungsstandard TEI sowie sich nicht an speziellen Formaten orientierende Inhalte relationaler Datenbanken aufgeführt. Mithilfe von Mint und D2R bzw. der übergreifenden Anwendung OmNom werden diese Daten in RDF umgewandelt. Die Umwandlung in RDF geschieht in DM2E dadurch, dass die Daten direkt in das DM2E-Modell überführt werden. Dieses ist spezifischer und dadurch ausdrucksstärker als das EDM, kann aber jederzeit wieder in das EDM zurückgeführt werden. Die Tripel im DM2E-Modell werden anschließend in AP3 als Grundlage für weitere Annotationen verwendet. Über diese könnten die ursprünglichen Providerdaten weiter angereichert werden. Darüber hinaus werden Daten in DM2E über Silk mit kontrollierten Vokabularen aus der LOD Cloud angereichert.

Die Spezialisierung des EDM erfolgt als iterativer Prozess und ist noch nicht abgeschlossen. Zunächst wurden



**Abb. 3:** Datenformate und Modelle in DM2E. Je höher das Format oder Modell in der umgekehrten Pyramide angesiedelt ist, desto reicher ist es in den Transformationsprozessen in DM2E. Die Ausdrucksstärke steigt durch die Anreicherung der Daten mittels modellspezifischer Klassen und Eigenschaften.

Metadaten in DM2E untersucht sowie darüber hinausgehende Anforderungen der Datenlieferanten gesammelt und beschrieben. Anschließend wurden diese Anforderungen als neue Eigenschaften oder Klassen im Modell angelegt. Der Aufbau des DM2E-Modells orientiert sich am EDM: Benennungen wurden konform zum EDM verwendet und Spezialisierungen fanden vor allem durch das Hinzufügen neuer Eigenschaften zu bereits vorhandenen Klassen im EDM statt. Die erste Version des Modells wurde im Oktober 2012 veröffentlicht und für die ersten Mappings von Daten aus dem Wittgenstein-Archiv der Universität Bergen<sup>25</sup> verwendet. In die zweite Version des Modells flossen außerdem Ergebnisse aus Mapping-Workshops mit ein, die mit allen im Projekt involvierten Datenlieferanten durchgeführt wurden. Klassenspezialisierungen fanden vor allem für die gelieferten Objekte selbst und für *Agents* statt. Zentrale Klassen im EDM sind `edm:ProvidedCHO`, also das von einer Kulturerbebeeinrichtung bereitgestellte digitale Objekt beziehungsweise *Cultural Heritage Object (CHO)*, und `ore:Aggregation`, einer Klasse, die Provenienzmetadaten über den Datenlieferanten und Rechte des Objektes umfasst. Im DM2E-Modell ist jede Instanz der Klasse `edm:ProvidedCHO` außerdem eine Instanz einer Subklasse von `edm:PhysicalThing`. Diese Subklassen umfassen beispielsweise Bücher, Seiten oder Fotos und geben zusätzlich Auskunft über die Objektart. Agents werden über `dm2e:Institution` in spezifische Institutionen und

<sup>25</sup> Digitalisate und deren TEI-Transformation aus dem Wittgenstein Archiv sind online über Wittgensteinsource zugänglich: <http://www.wittgensteinsource.org> [03.03.2013].

<sup>24</sup> Formate zur Beschreibung von Bibliotheksressourcen.



in die Klasse `dm2e:Person` unterteilt. Ergänzende Eigenschaften wurden vor allem für die Klasse `edm:ProvidedCHO` angelegt, wie beispielsweise das Incipit eines Manuskripts<sup>26</sup>, der Empfänger eines Briefes, der Signatur eines Buches und viele mehr. Über diese können detaillierte Aussagen über das Objekt selbst und den Ersteller sowie in die Erstellung miteinbezogene Personen oder Institutionen getroffen werden. Neu ist weiterhin die Verwendung sogenannter *Named Graphs* (Carroll, 2005). *Named Graphs* sind eine noch nicht standardisierte Technik, die auf RDF aufbaut. Durch die Verwendung von *Named Graphs* werden RDF-Graphen um eine weitere Eigenschaft ergänzt. Aus RDF-Tripeln werden so Quadrupel, die beispielsweise Versionierungsinformationen enthalten können. Dadurch werden RDF-Tripel nicht mehr allgemeingültig, sondern müssen im Kontext eines RDF-Graphen gesehen werden, welcher wiederum mit einer URI benannt ist und über diese URI selbst Teil von RDF-Tripeln werden kann. In *OmNom* wird das Ergebnis jedes Workflow-Durchlaufs in einem *Named Graph* gespeichert und dieser *Named Graph* mit allen verfügbaren Provenienzinformationen verknüpft. Dadurch ist die Herkunft und Entstehung der RDF-Daten jederzeit nachvollziehbar und gleichzeitig sind die Daten versioniert und gegen versehentliches Überschreiben geschützt.

DM2E-Spezialisierungen befinden sich in den ersten Versionen noch im eigenen Namensraum. Hier unterscheidet sich das Modell zurzeit vom darüber liegenden EDM. Im EDM sind Klassen und Eigenschaften externer Vokabulare unverändert in den hierarchischen Grundaufbau des Modells integriert worden. Um eine zu starke Vermischung unterschiedlicher semantischer Beschreibungen zum Teil heterogener Modellierungsansätze zu vermeiden, wurde in DM2E bisher auf die direkte Übernahme anderer Ontologieelemente verzichtet, wenn sich die Definition dieser nicht vollkommen mit der von DM2E deckt, und stattdessen auf diese, beispielsweise über OWL-Äquivalenzrelationen, verwiesen (Dröge, Iwanowa et al., 2013).

Die netzbasierte und verteilte Infrastruktur einer auf Semantic-Web-Technologien aufbauenden Plattform, wie sie in DM2E entwickelt wird, bietet aus unserer Sicht über die Nutzung im Projekt hinaus das Potential zur Lösung bereits bekannter Probleme und Herausforderungen im Bereich von Digitalen Bibliotheken und Linked Data, wie etwa des adäquaten Umgangs mit Versionierung und Provenienz mithilfe von *Named-Graph*-Ansätzen. So kön-

nen Ergebnisse aus DM2E nicht nur helfen, Inhalte in Europeana zu übertragen, sondern auch dazu beitragen mit der Deduplizierung identischer Datensätze umgehen zu können oder die Verlinkung und Kontextualisierung zu vereinfachen, was Datenbestände sichtbar und wertvoll für die Nachnutzung macht.

Darüber hinaus wird im Projekt auf Basis der verteilten semantischen Infrastruktur nach innovativen Lösungsansätzen für die digitalen Geisteswissenschaften geforscht. Diese sind der Schwerpunkt der Aktivitäten im nachfolgend beschriebenen Arbeitspaket 3.

### 3.2 Modellierung geisteswissenschaftlicher Forschung

Das Arbeitspaket 3 *Digital Humanities Requirements and Related Engineering* (AP3) in DM2E beschäftigt sich mit den digitalen Geisteswissenschaften. Der erste Arbeitsschwerpunkt liegt auf der Entwicklung einer integrierten digitalen Arbeitsumgebung für Forschungen an handschriftlichen Manuskripten. Der Name dieser Arbeitsumgebung lautet *Pundit* (Grassi, Morbidoni et al., 2012).<sup>27</sup> Ihre Konzipierung wird durch ein *Digital Humanities Advisory Board* (DHAB) mit prominenten Vertretern aus dem Bereich der digitalen Geisteswissenschaften unterstützt. Das Advisory Board unterstützt darüber hinaus die Ausarbeitung eines *Scholarly Domain Model*, welches ein pragmatisches, formales Modell geisteswissenschaftlicher Forschungsaktivitäten darstellt. Ziel ist es dabei, aufgrund einer hinreichenden semantischen Formalisierung der in solcher Forschung typischerweise maßgeblichen Entitäten und der zwischen ihnen bestehenden Relationen zu einer Modellierungsstufe zu gelangen, die Aussagensysteme im Sinne von Metaforschung auf einer vergleichsweise elementaren Ebene ermöglicht. Ziel ist es nicht, eine geschlossene Theorie geisteswissenschaftlicher Forschungsprozesse zu formulieren: das wäre sicher ein vermessenem Unterfangen. Die Ausarbeitung dieses Modells stellt einen zweiten Arbeitsschwerpunkt in AP3 dar, auf welchen hier näher eingegangen werden soll.

Die Ausgangsüberlegung ist, dass die Entwicklung und Bereitstellung von Infrastruktur und Anwendungen für die digitalen Geisteswissenschaften stärker als bisher von einem besseren Verständnis der Aktivitäten angeleitet werden sollte, die Geisteswissenschaftler in ihrer Forschungsarbeit unternehmen. Vor diesem Hintergrund dient das *Scholarly Domain Model* zunächst als Bezugs-

<sup>26</sup> Gerade Manuskripte enthalten oft keinen Titel. An dessen Stelle wird bei der Katalogisierung der Anfangsteil des Textes, das Incipit, verwendet.

<sup>27</sup> Pundit. <http://thepundit.it> [03.03.2013].

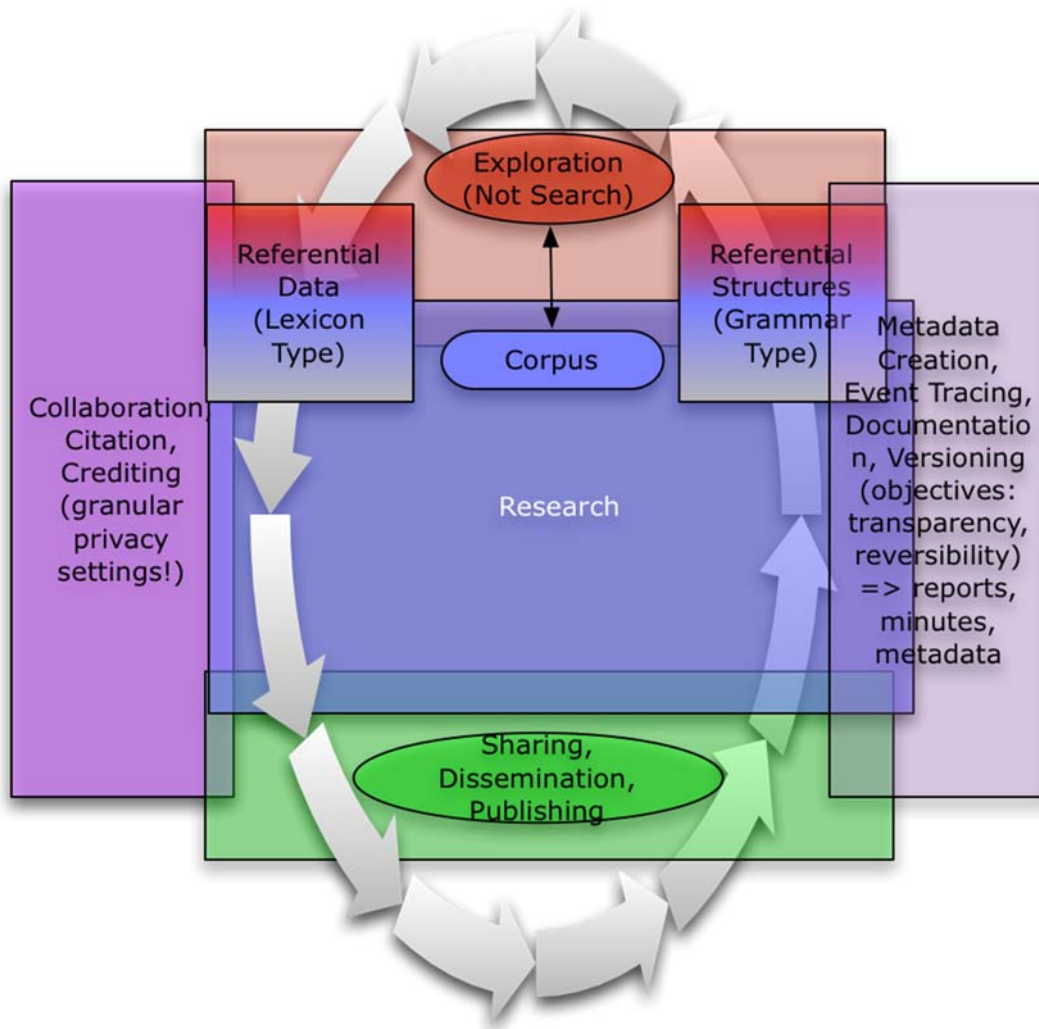


Abb. 4: Der engere Forschungsbereich in der Mitte überlappt sich mit den kontextualisierenden Bereichen, mit denen er Aktivitäten teilt.

rahmen und Diskussionsgrundlage für die Konzipierung und Ausgestaltung von digitalen geisteswissenschaftlichen Forschungsumgebungen im Allgemeinen und von *Pundit* im Besonderen.

Das Modell basiert auf den *Scholarly Primitives*, welche Unsworth (2000) als die grundlegenden Primärfunktionen von Forschungsaktivitäten vorschlug: *discovering*, *annotating*, *comparing*, *referring*, *sampling*, *illustrating* und *representing*. Ihnen sollen sich alle geisteswissenschaftlichen Forschungsaktivitäten zuordnen lassen, unabhängig von der jeweiligen Disziplin, dem historischen Zeitkontext und der theoretischen Orientierung. Die *Scholarly Primitives* sollen somit die Grundlage für die Entwicklung von Anwendungen für die digitalen Geisteswissenschaften bilden. Über eine definitive Auflistung oder Definition der Primitive herrscht zwar keine Einigkeit in der bisherigen Forschung, jedoch hat sich der grundsätzliche Ansatz als robust und fruchtbar herausgestellt, nicht

zuletzt dadurch, dass er immer wieder aufgegriffen worden ist.

Palmer et al. (2009) beispielsweise erweitern Unsworths eher statischen Ansatz zu einem Modell der *scholarly information activities*, welches den lebendigen Charakter von Forschung betont und die Primitive zu Grundsteinen von Forschungsaktivitäten macht. Auf Grundlage einer umfangreichen Literaturrecherche identifizierten Palmer et al. *searching*, *collecting*, *reading*, *writing* und *collaborating* als die zentralen fünf *scholarly information activities*, die wiederum aus mehreren Primitiven bestehen. Das *Scholarly Domain Model* übernimmt einen ähnlichen Ansatz, jedoch bilden die *Scholarly Primitives* hier die oberste Ebene und gliedern sich in mehrere granulare Forschungsaktivitäten. Das Bamboo Projekt (010) veranstaltete eine Serie von Workshops mit Praktikern aus dem Bereich der Digitalen Geisteswissenschaften zum Thema *scholarly practices*. Der *Scholarly Practice Report*

fasst die Ergebnisse der Workshops zusammen und stellt diese den *Scholarly Primitives* von Unsworth und den *scholarly information activities* von Palmer et al. gegenüber. Dieser Report bildet eine reichhaltige, empirische Quelle für unseren Modellentwurf. Weiterhin haben McCarty et al. (2002) mit den *methodological commons* einen Versuch unternommen, ein Gesamtbild von Forschungsaktivitäten im Bereich der Humanities zu skizzieren. Ihr Vorschlag stellt eine Art intellektuelle Karte von Forschungsaktivitäten dar, die in erster Linie als Diskussionsgrundlage für die Absteckung des Feldes der Digital Humanities dienen soll. Im Rahmen des Projekts DARIAH wurde dieser Ansatz von Anderson et al. (2010) mit den *Scholarly Primitives* von Unsworth zu einem konzeptuellen Rahmen für Anwendungsentwicklung in den Digitalen Geisteswissenschaften kombiniert. Die Primitive stellen aus ihrer Sicht unter anderem ein Mittel der Kommunikation dar, um traditionellen Forschern digitale Anwendungen näher zu bringen. Die Bedeutung der *Scholarly Primitives* und eines entsprechenden formalen Modells von geisteswissenschaftlicher Forschungsaktivität ist darüber hinaus auch von Blanke & Hedges (2011), Bernardou et al. (2010) und Kindling (2009) unterstrichen worden.

In DM2E greifen wir unter anderem diese Ansätze auf und erweitern sie zu einem integrierten Prozessmodell geisteswissenschaftlicher Forschungsaktivität, welches hierfür relevante Aspekte abbilden und formalisieren soll. Während Unsworth die *Scholarly Primitives* lose nebeneinander gruppiert, kontextualisiert das *Scholarly Domain Model* den engeren Forschungsbereich zunächst mit vier weiteren Bereichen (Abbildung 4): Ein *Input-* und *Output-Bereich* sowie flankierend ein *sozialer Kontext* und ein *Meta-Diskurs*. Jeder dieser Bereiche verweist auf wichtige Aktivitäten, die Einfluss auf die Forschungsaktivitäten im Zentrum des Modells ausüben. Die Bereiche überlappen sich insbesondere mit dem zentralen Forschungsbereich, da einige Aktivitäten aus den angrenzenden Bereich ebenso direkt zum Forschungsbereich in der Mitte zu zählen sind, wie beispielsweise *citation*, *collaboration* oder *documentation*.

Der *Input-Bereich* betont die Wichtigkeit der explorativen Erstellung eines Korpus, auf welchem geisteswissenschaftliche Forschung (einschließlich der Archäologie) ganz wesentlich basiert. Ein Korpus wird dabei als eine Aggregation von einem oder mehreren beliebigen Forschungsmaterialien, beispielsweise Texte, Bilder, Töne oder auch Tonscherben, verstanden, über die wissenschaftlich gearbeitet werden soll. Diese Materialien sind in der Regel keine empirisch erhobenen Daten, sondern werden mitunter zufällig durch Exploration entdeckt und in einem Korpus zusammengestellt. Das Durchstöbern ei-

ner Bibliothek oder eine archäologische Ausgrabung sind Beispiele für den explorativen Prozess der Zusammenstellung eines Korpus. Dieser Vorgang ist gleichzeitig bereits ein wichtiger Aspekt des Forschungsprozesses selbst. Weiterhin ist das Korpus immer vorläufig und kann im Verlauf der Forschung immer wieder modifiziert werden. Daher befindet sich das Korpus in dem überlappenden Bereich zwischen *Input-* und *Research-Bereich*. Dieser Aspekt ist wichtig, da das Korpus auch während des Forschungsprozesses durch neue Erkenntnisse oder Anforderungen verändert werden kann. Unterstützt und gespeist wird die Exploration durch *Referential Data* (damit sind zum Beispiel historische Wörterbücher oder Typologien von griechischen Vasen gemeint) und *Referential Structures* (damit ist auf jede Art von Grammatik oder allgemeiner Regelsysteme verwiesen). Im *Output-Bereich* am unteren Ende des Modells steht das Veröffentlichen und Verbreiten von Ergebnissen bzw. Teilergebnissen einer Forschung im Vordergrund. Hier besteht ebenso eine Überlappung zum zentralen Forschungsbereich, da Ergebnisse gerade in den textorientierten Geisteswissenschaften häufig eng mit dem Forschungsgegenstand, den angewandten Methoden und dem Diskurs selbst verbunden sind. Als Beispiel sei hier auf eine kritische Edition verwiesen. Im digitalen Umfeld stellt sich dabei die Herausforderung, derartige Ergebnisse, beispielsweise die verwendeten Forschungsmaterialien und den geführten Diskurs in der Form von RDF-Statements, als ein stabiles und referenzierbares *digitales Paket* zu veröffentlichen. Wichtig ist schließlich, dass es sich hier nicht um einen linearen Forschungsprozess handelt, sondern mindestens um einen Kreislauf, in welchem Ergebnisse eines Forschungsdurchlaufs immer wieder Input für einen weiteren oder neuen Forschungsschritt sein können.

Der *soziale Kontext* verdeutlicht, dass Forschung nicht ausschließlich an Inhalten orientiert ist, sondern ebenso stark durch eine Vielzahl verschiedener sozialer Aspekte motiviert und beeinflusst wird. Als ein wichtiges Beispiel ist hier die immer wichtiger werdende Kollaboration mit anderen Forschern zu nennen. Für kollaborativ-orientierte, digitale Anwendungen rücken in diesem Zusammenhang flexible und granulare *privacy settings* in den Vordergrund, über die Forscher bestimmen können, welche Forschungsinhalte (Annotationen, Texte, Materialien etc.) sie mit wem teilen. Der *Meta-Diskurs* ist ein weiterer integraler Bestandteil von Forschung und meint die Selbstdokumentation des Forschers und des Forschungsprozesses in der Form von Protokollen und Berichten oder auch im weitesten Sinne eine Versionierung von Forschungsdaten, durch welche die Evolution eines Forschungsgegenstandes nachvollziehbar gemacht werden

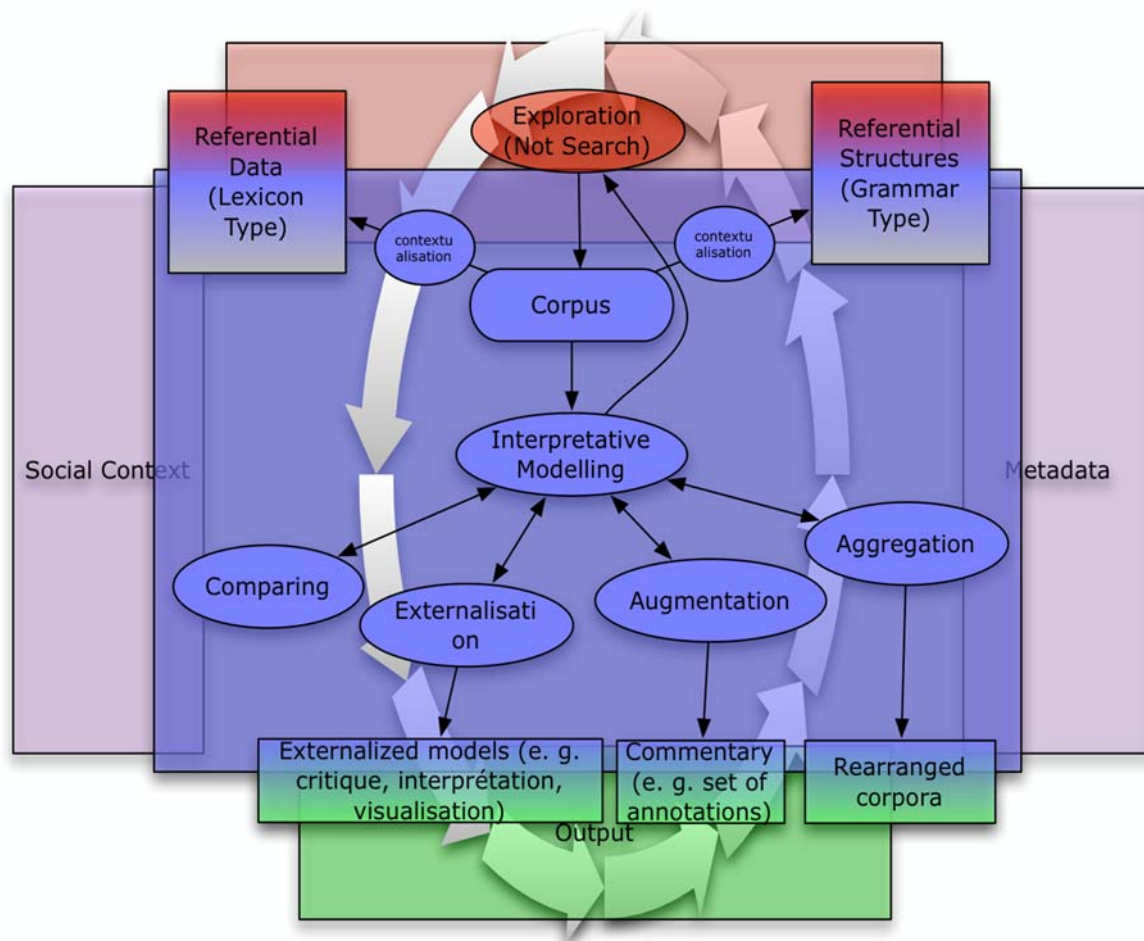


Abb. 5: Im Zentrum des *Scholarly Domain Models* stehen die Primitive geisteswissenschaftlicher Forschungsaktivität.

kann, beispielsweise im Fall von terminologischen Projekten. Das Ziel ist dabei die Sicherstellung von Transparenz und Nachprüfbarkeit von Forschungsergebnissen als wesentliches Merkmal seriöser Forschung.

Diese vier Bereiche rahmen den eigentlichen Forschungsprozess im Zentrum des Modells ein (Abbildung 5). Der hier gezeigte Forschungsprozess ist als generisches und disziplin-unabhängiges Modell gedacht. Im Forschungsprozess sind die beiden ersten und grundlegenden Schritte zum einen die Kontextualisierung des Korpus mit referentiellen Daten und Strukturen und zum anderen gleichzeitig das *Verstehen* des Korpus und seiner Bestandteile. Das *Verstehen*, welches hier vorläufig als *Interpretative Modelling* bezeichnet wird, konstituiert dabei die grundlegende Tätigkeit, in der alle anderen Primitive und Aktivitäten fußen. Die im *Scholarly Domain Model* vorgeschlagenen Primitive stellen dabei zunächst eine weitere Generalisierung der sieben *Scholarly Primitives* von Unsworth dar.

Die *Aggregation* meint das Re-Arrangement des Korpus, beispielsweise durch das Entnehmen einer Stichpro-

be oder dem Filtern nach einem bestimmten Kriterium. Ein typisches Ergebnis ist eine digitale Edition oder die Rekonstruktion einer Vase. Die *Augmentation* fügt dem Korpus neue Elemente hinzu. Beispiele hierfür sind eine Annotation oder ein Kommentar zu einem Text. Das Hinzufügen von derartigen Aussagen ist sicherlich eine der zentralen Aktivitäten im Zusammenhang mit textorientierten Wissenschaften. Die *Externalisation* macht Ergebnisse des *Interpretative Modelling*, also dem *Verstehen*, bzw. der anderen Primitive lesbar in Form einer verschriftlichten Interpretation, Kritik oder einer Visualisierung eines Modells. Das *Comparing* erscheint als eine weitere grundlegende Aktivität, obwohl bisher kein direktes Ergebnis assoziiert werden konnte. Der Vergleich erscheint in erster Linie als ein in anderen Aktivitäten partizipierendes Primitiv. In diesem Zusammenhang ist anzumerken, dass die Primitive weder in ihrer Anwendung noch in ihrer Definition vollkommen trennscharf sind. Primitive treten häufig in Kombination miteinander auf, wie beispielsweise das *Interpretative Modelling* jeder anderen Aktivität zugrunde liegt. Andere Primitive wie

Augmentation sind zudem rekursiv: Eine Annotation kann durch eine weitere Annotation annotiert werden. Diese Zusammenhänge sind aber bisher nicht im Modell ausgedrückt.

Das Modell befindet sich als ein erster Entwurf in der Entwicklung. Die Spezialisierung und Definition der einzelnen Primitive und ihr Verhältnis untereinander sind die kommenden Arbeitsschritte. Darüber hinaus können Teile des Modells mit Ontologien ausformuliert und damit in digitalen Anwendungen implementierbar gemacht werden. Damit ist eine weitere Funktion des Modells benannt, die über einen anleitenden Diskussionsrahmen für die Entwicklung und Integration von Tools für die digitalen Geisteswissenschaften hinausgeht. Beispielsweise könnte *Argumentation* als Spezialisierung von *Augmentation* mit einer Ontologie für disziplinspezifische Diskursfiguren ausgestaltet werden. Forscher, die diese Ontologie dann für einen wissenschaftlichen Diskurs nutzen, können den Verlauf des Diskurses später auf vielfältige Weise auswerten. Ähnliches ist für die anderen Primitive denkbar, genauso wie für das Verhältnis der Primitive untereinander. Dadurch ergeben sich neue Möglichkeiten für Funktionalitäten, die über die Emulation geisteswissenschaftlicher Forschungsaktivitäten hinausgehen. Anzumerken bleibt als weiterer wichtiger Arbeitsschritt, die Grenzen des Modells und seiner Formalisierung zu formulieren. Ziel ist es nicht, Kreativität in der Forschung zu ersetzen, sondern gerade innovative und kreative Prozesse durch das Potential digitaler Arbeitsumgebung zu unterstützen.

Das *Scholarly Domain Model* schafft im Rahmen von DM2E eine Verbindung zwischen Geisteswissenschaften und digitaler Anwendung. Das Modell wird darüber hinaus Anschluss in *Bild Wissen Gestaltung* finden. Dort wird es im Rahmen der beiden Teilprojekte *Virtuelle und reale Architektur des Wissens* und *Shaping Knowledge* im Kontext des *interdisziplinären Labors* weiterentwickelt werden.

## 4 Virtuelle und reale Architektur des Wissens und Shaping Knowledge

Der im Rahmen der Exzellenzinitiative neu entstandene Cluster *Bild Wissen Gestaltung* der Humboldt-Universität zu Berlin zeichnet sich neben seinem einzigartigen Programm vor allem durch Multidisziplinarität aus. Wissenschaftler aus 22 Disziplinen bringen hier ihre Kenntnisse, Methoden und Perspektiven ein, um gemeinsam etwas zu entwickeln, das mehr ist als die Summe seiner Teile.

Der Cluster selbst nennt sich ein *interdisziplinäres Labor*<sup>28</sup> und spannt einen Rahmen auf für fachübergreifende Forschung und die Erforschung der Interdisziplinarität selbst. Diese Metaforschung will verstehen, wie Prozesse über Fachgebiete hinweg funktionieren, wann geteilte Informationen ein neues Verständnis befruchten oder wann sie es vielleicht nur verzögern. Nicht zuletzt stellt sich dabei die Frage: Wie arbeiten wir eigentlich? Was heißt es, ein Wissenschaftler zu sein in der (digitalen) Praxis? Welche Bedingungen begünstigen Produktivität und welche Innovation? Wie gehen Menschen damit um, wenn sie mit ganz anderen Perspektiven und Methoden konfrontiert werden? Welche Quellen sind wie wichtig für eine Person oder ein Team? Und wie kann man Informationen und Wissen organisieren und abbilden? Diese und andere Fragestellungen versuchen die Basisprojekte *Virtuelle und reale Architektur des Wissens* und *Shaping Knowledge* auf vielfältige Art und Weise zu untersuchen und zu modellieren. Untersuchungsgegenstand sind dabei die digitalen Spuren bzw. expliziten Angaben der Mitarbeiter des Clusters selbst. Die Visualisierung ausgesuchter Relationen wird ein wichtiger Teil dieser Abbildung werden. Erste Vorarbeiten dazu sind bereits im Bereich von *OmNom* erfolgt. Zukünftig ist die Navigation in RDF-Graphen und die Manipulation von RDF-Strukturen in grafischen Oberflächen angedacht. Datenmodellierung, -navigation und -manipulation sollen transparent in nutzerfreundlichen Umgebungen ermöglicht werden.

Eine der Herausforderungen dabei ist, dass sich Wissen als kontextualisierte Information unzureichend in statischen und streng hierarchischen Datenstrukturen abbilden lässt. Für interdisziplinäres Wissen gilt dies insbesondere, da für solche Übergangsbereiche häufig noch keine Modellierungsstrukturen etabliert sind. Multiple Kontexte für Termini verlangen multihierarchische Ordnungssysteme. Statische Datenstrukturen sind sinnvoll, um ein spezialisiertes, genau bekanntes und begrenztes Wissensgebiet zu modellieren. Sucht man jedoch nach einer Möglichkeit, all diese heterogenen Vokabulare zu verbinden, kann die Lösung nicht in einer simplen Addition dieser liegen. Vielmehr ist ein verbindendes Konstrukt in Form einer Top-level-Ontologie notwendig, das in der Lage ist, heterogene Vokabulare zu vereinen, ohne ihre Semantik zu reduzieren. Daher haben wir uns im Cluster entschieden, RDF als universelles Verbindungselement einzusetzen, und einen großen, performanten Triplestore für die im gesamten Cluster entstehenden Daten aufzubauen. Die in den vielen Basisprojekten anfal-

<sup>28</sup> Interdisziplinäres Labor. <http://www.interdisciplinary-laboratory.hu-berlin.de/de> [12.02.2013].

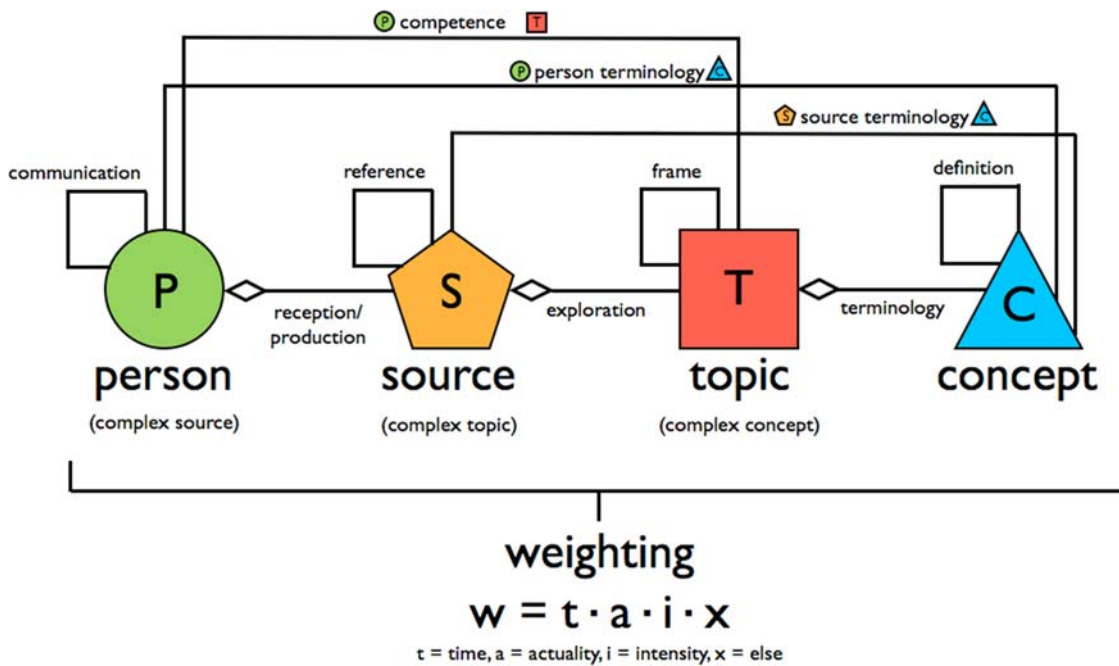


Abb. 6: Entwurf Entitätenmodell zur Abbildung von interdisziplinären Forschungsprozessen im BWG-Cluster.

lenden textlichen Inhalte indexieren wir mit *Solr*<sup>29</sup>, damit diese fachübergreifend recherchiert werden können.

Die RDF-Modellierung teilt sich dabei in zwei Bereiche auf: Auf der einen Seite die Modellierung aktuell untersuchter Wissensräume einzelner Projektteams und die Integration der darin vernetzten Daten in Linked (Open) Data. Auf der anderen Seite die RDF-Abbildung von Forschungsprozessen und Wissenschaftlerinteraktionen – eine Art semantisches Metaforschungsnetz. Derzeit wird eine mögliche Datengrundlage für eine prototypische RDF-Modellierung evaluiert. Hier kommen als potentielle Erschließungsprojekte eine Kostümsammlung, das Berliner Computerspielmuseum oder ein zoologisches Handbuch mit speziellen Funktion-Strukturzusammenhängen in Frage. Materialwissenschaftler, Informatiker, Architekten, Biologen, Gestalter, Interaktionsdesigner, Soziologen, Kunsthistoriker und Wissenschaftler vieler anderer Disziplinen sind hier involviert. Eine der nächsten Herausforderungen wird die Identifikation bzw. der eventuell notwendige Aufbau eines kontrollierten Vokabulars in RDF für das zu erschließende Gebiet sein.

Die Metaforschungsansätze basieren auf Arbeiten wie denen des Projektes Bamboo (2010). Im Cluster sollen auf atomare Aktivitäten herunter gebrochene Prozesse anhand der digitalen Spuren analysiert werden, damit sie später ontologisch formalisiert werden können. Diese

Prozesse entstehen bei der Nutzerinteraktion mit den uns im Cluster zur Verfügung stehenden Systemen wie zum Beispiel Logdateien. Lexikalische Statistik ist dabei für Logs aller Art angedacht, um ein umfassendes Bild zu erzeugen. Analoge Medien wie Papier oder aufgenommene Sprache werden dieser Analyse ebenfalls zugeführt, damit Metainformationen auf wissenschaftliche Inhalte abgebildet werden können.

In Zusammenarbeit mit den Datenschutzbeauftragten werden so die Grenzen interdisziplinärer Metaforschung ausgelotet. Gemeinsam mit anderen Projekten werden die digital erhobenen Daten durch Befragungen abgerundet und wo notwendig ergänzt.

Der Entwicklungsstand des diesbezüglich zurzeit diskutierten Datenmodells soll hier nun dargestellt werden.

Das Modell (Abbildung 6) geht von der kleinsten selbstständigen sprachlichen Einheit aus, dem Wort, das hier formalisiert als Lexem gefasst wird. Ein Lexem verbindet den Begriff (Bedeutung, gefasst als Definition) mit der diesen bezeichnenden Benennung (Laut- bzw. Buchstabenfolge) innerhalb einer Varietät (spezifische Fach-, Gruppen- oder Gemeinsprache).<sup>30</sup> Für einen Algorithmus, der einen beliebigen Text untersucht, ist ein Lexem zunächst einmal einfach nur ein von Leerzeichen umgebener String. Dieser kann dann herausgegriffen und untersucht werden. Konkret wird man versuchen, diesen

<sup>29</sup> Apache Solr. <http://lucene.apache.org/solr> [12.02.2013].

<sup>30</sup> Für eine Analyse von Lexem- und Terminusmodellierungen siehe Stein (2012).

String entweder auf seine grammatischen und syntaktischen Eigenschaften hin zu untersuchen, oder aber ihn zu identifizieren als die flektierte Repräsentation eines Lexems. Dazu benötigt man sogenannte Stemmer, die in der Lage sind, flektierte Formen auf Stamm- und (lemmatisierte) Zitationsformen zurückzuführen. Kommt in einem Text beispielsweise „Bücher“ vor, kann der Stemmer dieses syntaktische Wort automatisch auf „Buch“ zurückführen und somit einen Vergleich auf Treffer mit einer lexikalischen Datenbank, Terminologie oder Ontologie vornehmen. In einer solchen Datenbasis ist idealerweise vermerkt, in welcher fachsprachlichen oder gemeinsprachlichen Varietät das entsprechende Lexem gebräuchlich ist. Bedingt durch die Entwicklung einer Sprache ist diese Zuordnung weder eindeutig noch eineindeutig. So werden durch einen Prozess, der sich Transterminologisierung nennt, häufig Fachwörter aus einer Fachsprache in eine andere importiert und dort mit einer neuen Bedeutung versehen. Im Falle von „Virus“ ist das beispielsweise geschehen: heute handelt es sich dabei nicht nur um ein Lexem aus der Biologie sondern auch um eines aus der Informatik. Um in einem solchen Fall entscheiden zu können, welcher Virus in einem Text gemeint ist, geht der Algorithmus genauso vor wie der Mensch: Er betrachtet den Kontext, die Kookkurrenzen, die in der Nähe des Lexems im Text auftauchen und die fachsprachliche Zuordnung derselben. Auf diese Weise lässt sich in einem Text eine Reihe von Lexemen identifizieren. Diese wiederum können einen gewissen Bezug zueinander aufweisen, den man erhält, wenn man große Korpora untersucht: Wenn eine Gruppe von Lexemen häufiger zusammen auftritt, handelt es sich in diesem Text wahrscheinlich um ein gemeinsames Thema, das sie verbindet.

Textbasierte Clusteralgorithmen gestatten hier einen ersten Einblick in inhaltliche Ballungen. Daraus können Hinweise auf im Korpus inhärente Konzepte entnommen werden. Deren Formalisierung und Verknüpfung wird in RDF vorgenommen.

Themen haben eine deutlich höhere Abstraktionsstufe als Lexeme und sind eher in der Lage, die Inhalte eines Textes kurz und knapp zu umreißen. Für eine Analyse, wie sie im Cluster vorgesehen ist, ist es daher wichtig, die behandelten Themen zu identifizieren, um herausfinden zu können, wer sich womit beschäftigt und wann ein Thema transdisziplinär aufgegriffen wird. Das Thema wird in diesem Fall als eine Art komplexes Lexem behandelt, das inhaltlich viele andere umfasst und sie in Beziehung zueinander setzt. In der Datenmodellierung gehen wir also davon aus, dass eine Reihe von Lexemen repräsentativ für ein Thema ist. Entsprechend werden sie miteinander vernetzt.

Die auf ein oder mehrere Themen zurückführbaren Lexeme finden sich wiederum in Texten oder anderen Quellen. Das können ebenso Bildunterschriften, Tags, Annotationen oder per Spracherkennung erfasste Audioaufnahmen sein. Typische Quellen im Cluster sind neben Artikeln und Monographien auch Protokolle, Notizen, Exzerpte und Arbeitsdokumente. Wenn es gelingt, eine möglichst große Zahl davon zu analysieren, lässt sich ein dynamisches Profil der im Cluster behandelten Themen inklusive der Intensität und dem zeitlichen Verlauf erstellen. Typischerweise werden in Quellen unterschiedliche Themen angesprochen und miteinander in Beziehung gesetzt. Dieser Bezug rechtfertigt es, eine Quelle als ein komplexes Thema zu verstehen, eines das gleichzeitig mehrere Themen in Bezug zueinander setzt und abgesehen vom Review wesentlich konkretere Positionen dazu einnimmt. Dieses Verständnis muss allerdings dahingehend eingeschränkt werden, dass eine inhaltliche Analyse der Quellen auf Wertungen und Konnotationen nicht geplant ist. Identifizierte Themen können Teil einer Hierarchie sein, die wir hier jedoch nicht weitergehend untersuchen und stattdessen eine eher flache thematische Struktur annehmen. Für die ersten prototypischen Experimente wird auch das Bild (losgelöst vom Text) als Kommunikationsmedium nicht weiter betrachtet. Es ist zuerst zu prüfen, ob die bisher nur für Texterschließung vorhandenen Algorithmen der Fragestellung dienlich sind.

Noch mehr lässt sich mit diesen gesammelten Informationen anfangen, wenn man die Personen hinzunimmt, die mit den entsprechenden Quellen, Themen und Lexemen arbeiten. Das bedeutet nicht, dass eine Auflösung auf Realnamen geplant wäre, jedoch, dass man den gleichen Akteur im Cluster über eine pseudonyme ID identifizierbar hält. So lässt sich nicht nur ein clusterweites Abbild erstellen, sondern auch herausfinden, wie Informationen wandern: Tauchen bei einer Person auf einmal neue, ihr bisher fachfremde Themen auf, kann man von einer erfolgreichen transdisziplinären Übernahme sprechen. Diesen Prozess unmittelbar, sozusagen live, zu beobachten, wäre ein extrem spannendes Ereignis und eröffnet viele Möglichkeiten zur qualitativen Nachuntersuchung bezüglich der Bedingungen von Interdisziplinarität. Aber noch mehr ließe sich so die Methodenheterogenität von Disziplinen und Individuen, die Nachwirkkraft einzelner Quellen bzw. Treffen oder die relativen Barrieren bzw. Übergänge, die zwischen Disziplinen verdeckt oder offen bestehen quantitativ demonstrieren. Neben der reinen Analyse wollen wir mit diesen Daten aber auch konkrete Vorschläge unterbreiten können wie beispielsweise: Diese Quelle behandelt ähnliche Themen, jene ist im Cluster sehr populär.

Die für das oben skizzierte Szenario erforderlichen Informationen können teilweise automatisiert erfasst werden, das Datenmodell erlaubt aber auch das manuelle Einstellen solcher Relationen durch Beobachtung, Befragungen oder Selbstauskünfte. Dann ließe sich auch zeigen, wo Kommunikation zwischen Personen stattgefunden hat und u. U. auf Basis welcher Quellen bzw. über welches Thema. Die von verschiedenen Personen hauptsächlich studierten Themen können so quantifiziert werden. Ebenso ließe sich eine Art persönlicher Terminologie erstellen, die den spezialisierten und individualisierten Fachwortschatz einer Person repräsentiert.

Viele Informationen davon sind selbstverständlich nur für den betreffenden Forscher bestimmt und können nicht veröffentlicht werden. Im Rahmen des steigenden Interesses an einer Quantifizierung des eigenen Lebens und der eigenen Arbeit liegt genau darin jedoch noch mal eine signifikante Chance, um die Akzeptanz einer solchen Analyse zu befördern. Die QuantifiedSelf-Bewegung<sup>31</sup> bietet hier umfangreiche Softwarewerkzeuge, die im Cluster benutzt und erweitert werden sollen. Der Forscher ist damit nicht nur selbst Gegenstand der Metaforschung, sondern erhält auch unmittelbar die auf seine Person bezogenen Daten und Auswertungen, um das eigene Arbeitsverhalten besser überblicken zu können. So arbeiten wir auch hier darauf hin, dem Cluster als Untersuchungsgegenstand Erkenntnisse, die aus dem Arbeitsverhalten gezogen werden, zurückzugeben.

## 5 Fazit

Die Ausführungen zu den unterschiedlichen Projektaktivitäten der Forschergruppe sollten erkennbar gemacht haben, was diese im Kern zusammenhält, und inwiefern die Arbeit der Gruppe über die Jahre die von Gradmann 2008 gestellte Frage nach den *Bedingungen des Verstehens in semantischen Netzen* konsequent in den Fokus genommen und vertieft hat. Und wenn auch die Frage selbst vielleicht nie abschließend beantwortet werden kann, haben die letzten Jahre unserer Arbeit doch Elemente einer Antwort erbracht, an der wir weiter arbeiten wollen: unser Thema ist und bleibt die Organisation und Repräsentation von *Wissen* im Sinne von kontextualisierter Information in Modellen, welche die komplexen Verflechtungen von Wissensordnungen in Form von Prozessen und Strukturen verstehbar machen.

<sup>31</sup> QuantifiedSelf-Bewegung. <http://quantifiedself.com> [14.02.2013].

## Literatur

- Anderson, S.; Blanke, T.; Dunn, S.: Methodological commons: arts and humanities e-Science fundamentals. In: *Philosophical Transactions of the Royal Society A* 368 (2010), Nr. 1925, S. 3779–3796.
- Benardou, A.; Constantopoulos, P.; Dallas, C.; Gavrilis, D.: A Conceptual Model for Scholarly Research Activity. In: *iConference 2010 Proceedings*. 2010, S. 26–32.
- Blanke, T.; Hedges, M.: Scholarly primitives: Building institutional infrastructure for humanities e-Science. In: *Future Generation Computer Systems* 29 (2011), Nr. 2, S. 654–661.
- Bizer, C.; Cyganiak, R.: *D2R Server – Publishing Relational Databases on the Semantic Web*. (Poster at the 5th International Semantic Web Conference). Athens, USA, 2006.
- Carroll, J.; Bizer, C.; Hayes, P.; Stickler, P.: Named Graphs. In: *Journal of Web Semantics* 3 (2005), S. 247–267.
- Concordia, C.; Gradmann, S.; Siebinga, S.: Not just another portal, not just another digital library: A portrait of Europeana as an application program interface. In: *IFLA Journal* 36 (2010), Nr. 1, S. 61–69.
- Crane, G.; Almas, B.; Babeu, A.; Cerrato, L.; Harrington, M.; Bammann, D.; Diakoff, H.: Student Researchers, Citizen Scholars and the Trillion Word Library. In: *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2012)*. Washington D.C., USA., S. 213–222.
- Dekkers, M.; Gradmann, S.; Molendijk, J.: *Europeana v1.0: D3.4 Final Technical & Logical Architecture and future recommendations*. 2011. <http://pro.europeana.eu/documents/10602/370691/D3.4+final.pdf> [03.03.2013].
- Dröge, E.; Iwanowa, J.; Trkulja, V.; Hennicke, S.; Gradmann, S.: Wege zur Integration von Ontologien am Beispiel einer Spezifizierung des Europeana Data Model. To appear in: *Proceedings of the 13th International Symposium of Information Science*. 2013.
- Gradmann, S.: Signal. Information. Zeichen. Zu den Bedingungen des Verstehens in semantischen Netzen. In: *LIBREAS. Library Ideas* 14 (2009). <http://libreas.eu/ausgabe14/006gra.htm> [03.03.2013].
- Grassi, M.; Morbidoni, C.; Nucci, M.; Fonda, S.; Ledda, G.: Pundit: Semantically Structured Annotations for Web Contents and Digital Libraries. In: A. Mitschick; F. Loizides; L. Predoiu; A. Nürnberger; S. Ross (Hrsg.): *Proceedings of the 2nd International Workshop on Semantic Digital Archives*, 2012, S. 49–60.
- Heath, T.; Bizer, C.: *Linked Data: Evolving the Web into a Global Data Space*. Palo Alto, USA: Morgan & Claypool, 2011.
- Hyvönen, E.: *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Palo Alto, USA : Morgan & Claypool, 2012.
- Kindling, M.: *Möglichkeiten der Strukturmodellierung: Eine exemplarische Zusammenführung funktionaler Anforderungen an die Bereitstellung digitaler Forschungsdaten für ausgewählte geisteswissenschaftliche Disziplinen*. Berlin, Humboldt-Univ., Institut für Bibliotheks- und Informationswissenschaft, Magisterarbeit, 2009.
- Kollia, I.; Tzouvaras, V.; Drosopoulos, N.; Stamou, G.: A Systemic Approach for Effective Semantic Access to Cultural Content. In: *Semantic Web Journal* 3 (2012), Nr. 1, S. 65–83.
- McCarty, W.; Short, H.: *Mapping the Field*. Report. (ALLC meeting). Pisa, April 2002. <http://www.allc.org/node/188> [03.03.2013].
- Palmer, C.; Tefreau, L.; Pirmann, C.: *Scholarly Information Practices in the Online Environment: Themes from the Literature and*



*Implications for Library Service Development*. Report commissioned by OCLC Research, 2009. <http://www.oclc.org/resources/research/publications/library/2009/2009-02.pdf> [03.03.2013].

Project Bamboo Scholarly Practice Report. 2010.

<http://www.projectbamboo.org/wp-content/uploads/Project-Bamboo-Scholarly-Practices-Report.pdf> [03.03.2013].

Stein, C.: Intelligente Glossare: Vernetztes Terminologiemanagement für komplexe Fachtexte. In: *MDÜ – Fachzeitschrift für Dolmetscher und Übersetzer* 3 (2012), S. 46–50.

Unsworth, J.: Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this? (*Symposium on Humanities Computing: formal methods, experimental practice*). 2000. <http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html> [03.03.2013].

Volz, J.; Bizer, C.; Gaedke, M.; Kobilarov, G.: Silk – A Link Discovery Framework for the Web of Data. In: C. Bizer; T. Heath; T. Berners-Lee; K. Idehen (Hrsg.): *Linked Data on the Web* (LDOW2009). CEUR workshop proceedings. 2009. <http://ceur-ws.org/Vol-538/> [03.03.2013].

## Danksagung

Die Autorinnen und Autoren sind Mitarbeiterinnen und Mitarbeiter in Forschungsprojekten, die von der Europäischen Kommission sowie der Deutschen Forschungsgemeinschaft gefördert werden. Dank gilt daher allen Kolleginnen und Kollegen sowie der Europäischen Kommission und der Deutschen Forschungsgemeinschaft für die Förderung der Projekte.

### Forscherguppe Wissensmanagement

Die Autorinnen und Autoren sind Mitarbeiterinnen und Mitarbeiter der Forschergruppe Wissensmanagement und in verschiedenen Projekten am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin und der KU Leuven involviert. Europeana v2.0 und Digitised Manuscripts to Europeana (DM2E) sind EU-geförderte Projekte im Rahmen von Europeana, der Europäischen Digitalen Bibliothek. Teilprojekte des Exzellenzclusters Bild Wissen Gestaltung, mit Beteiligung von Mitgliedern der Forschergruppe Wissensmanagement, bauen auf deren Ergebnissen auf und verwenden sie weiter, um beobachtete Arbeitsprozesse von Wissenschaftlerinnen und Wissenschaftlern in RDF zu repräsentieren.



(v. l.) Evelyn Dröge, M. A. – [evelyn.droege@ibi.hu-berlin.de](mailto:evelyn.droege@ibi.hu-berlin.de); Marlies Olensky, M. A. – [marlies.olensky@ibi.hu-berlin.de](mailto:marlies.olensky@ibi.hu-berlin.de);  
Julia Iwanowa, M. A. – [julia.iwanowa@ibi.hu-berlin.de](mailto:julia.iwanowa@ibi.hu-berlin.de); Konstantin Baierer, B.A. – [konstantin@baierer.de](mailto:konstantin@baierer.de);  
Dr. Violeta Trkulja – [violeta.trkulja@ibi.hu-berlin.de](mailto:violeta.trkulja@ibi.hu-berlin.de); Alexander Struck, M. A. – [alexander.struck@ibi.hu-berlin.de](mailto:alexander.struck@ibi.hu-berlin.de);  
Prof. Dr. Stefan Gradmann – [stefan.gradmann@kuleuven.be](mailto:stefan.gradmann@kuleuven.be); Steffen Hennicke, M. A. – [steffen.hennicke@ibi.hu-berlin.de](mailto:steffen.hennicke@ibi.hu-berlin.de);  
Dr. Christian Stein – [christian.stein@hu-berlin.de](mailto:christian.stein@hu-berlin.de) (nicht im Bild)