

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 327

LINKED OPEN VD 17 – VON METS/MODS
ZUM EUROPEANA DATA MODEL

ÜBERLEGUNGEN ZUM TECHNISCHEN MIGRATIONSPFAD
UND ZUM FUNKTIONALEN MEHRWERT
SEMANTISCHER NUTZUNGSSZENARIEN

VON
ESTHER CHEN

LINKED OPEN VD 17 – VON METS/MODS
ZUM EUROPEANA DATA MODEL

ÜBERLEGUNGEN ZUM TECHNISCHEN MIGRATIONSPFAD
UND ZUM FUNKTIONALEN MEHRWERT
SEMANTISCHER NUTZUNGSSZENARIEN

VON
ESTHER CHEN

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 327

Chen, Esther

Linked Open VD 17 – von METS/MODS zum Europeana Data Model :
Überlegungen zum technischen Migrationspfad und zum funktionalen Mehrwert
semantischer Nutzungsszenarien / von Esther Chen. - Berlin : Institut für
Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin,
2012. - 64 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und
Informationswissenschaft; 327)

ISSN 14 38-76 62

Abstract:

Die vorliegende Arbeit geht anhand eines repräsentativen Beispielcorpus, dem VD 17, der Frage nach, wie für existierende bibliographische Metadaten Mehrwerte mit graphenbasierten Linked Open Data-Modellen geschaffen werden können. Neben dem Mapping der im Format METS/MODS vorliegenden Daten auf das Europeana Data Model bietet die Arbeit eine detaillierte Einführung in den auf unterschiedlichen Standards wie RDF, OWL, OAI-ORE, DC und SKOS basierenden technischen Aufbau des EDM. Anhand des Beispiels einer Leichenpredigt aus dem 17. Jahrhundert werden schließlich die wesentlichen Vorteile einer Modellierung der Daten mit dem EDM herausgearbeitet: dabei ergibt sich neben dem Integrationspotential des EDM die durch RDF geschaffene Möglichkeit eines Datenzugriffs mithilfe des W3C-Standards SPARQL. Dem prosopographischen Forschungsinteresse an Leichenpredigten kommen die entstandenen Kontextualisierungsoptionen etwa durch eine Integration in Personennetze entgegen. Einen wesentlichen Vorteil stellt nicht zuletzt die Möglichkeit der explorativen Suche dar, die mit einer graphenbasierten Modellierung der Daten einhergeht. Für bibliothekarische Einrichtungen stellt die Partizipation an der Informationsvernetzung mit Linked Open Data schließlich eine Chance dar, sich mit ihren Beständen aktiv am Aufbau neuer Informationsarchitekturen zu beteiligen.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im postgradualen Fernstudiengang Bibliotheks- und Informationswissenschaft (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version: <http://edoc.hu-berlin.de/series/berliner-handreichungen/2012-327>



Dieses Werk steht unter einer Creative Commons [Namensnennung-NichtKommerziell-KeineBearbeitung 3.0](#) Deutschland-Lizenz.

0. EINLEITUNG	6
1. DAS EUROPEANA DATA MODEL	12
1.1 EUROPEANA – MEHR ALS EIN PORTAL	12
1.2 SEMANTIC WEB UND LINKED (OPEN) DATA	13
1.3 DER AUFBAU DES EUROPEANA DATA MODEL	16
2. VD 17 – DIE METADATEN IM METS/MODS FORMAT	22
2.1 METS - METADATA ENCODING AND TRANSMISSION STANDARD	23
2.2 MODS - METADATA OBJECT DESCRIPTION SCHEMA	30
2.3 MÖGLICHE SCHWIERIGKEITEN FÜR DIE DATENMODELLIERUNG	32
3. DIE LEICHENPREDIGT „CHRISTIANI VITA ET CORONA“: MODELLIERUNG MIT DEM EUROPEANA DATA MODEL	34
3.1 DIE OBJEKTPRÄSENTATION MIT DEM EUROPEANA DATA MODEL	34
3.2 DER ORE:PROXY: DESKRIPTIVE METADATEN	37
3.3 DIE EREIGNISZENTRIERTE OBJEKTBSCHREIBUNG	39
3.4 DIE BESCHREIBUNG KOMPLEXER OBJEKTE	40
3.5 „CHRISTIANI VITA ET CORONA“ IM PICA3-FORMAT	42
4. EIN AUSBLICK AUF NEUE SUCHOPTIONEN UND SEMANTISCHE NUTZUNGSSZENARIEN	47
4.1 DAS INTEGRATIONSPOTENTIAL DES EUROPEANA DATA MODEL	47
4.2 SPARQL – EINE RDF ANFRAGE SPRACHE	48
4.3 PERSONENNETZE MIT DEM RESOURCE DESCRIPTION FRAMEWORK	49
4.4 DIE EXPLORATIVE SUCHE	54
5. FAZIT	56
ABKÜRZUNGSVERZEICHNIS	58
BIBLIOGRAPHIE	59
ABBILDUNGSVERZEICHNIS	64
TABELLENVERZEICHNIS	64

0. Einleitung

Seit der Entwicklung des World Wide Web durch Tim Berners-Lee Anfang der 90er Jahre des letzten Jahrhunderts und dem darauf folgenden rasanten Wachstum des WWW befindet sich die Informationslandschaft in einer Phase weitgreifender Umbrüche, die vielfach mit den Folgen des Paradigmenwechsels von der Handschrift zum Buchdruck verglichen wurde. In der Tat haben wir es heute, im Zeitalter der digitalen Reproduzier- und Publizierbarkeit, mit einem Paradigmenwechsel zu tun, dessen Folgen nicht gänzlich absehbar sind. Da wäre zum einen der empfindliche Eingriff in tradierte Geschäftsmodelle von Verlagen und Buchhandel zu nennen. Der klassischen Wertschöpfungskette im Schreib- und Publikationsprozess stehen beispielsweise heute Open Access Modelle gegenüber. Mit den sich verändernden Publikationsprozessen geht die Erosion etwas so Wesentlichen wie des Dokumentbegriffs einher (Pédauque 2003). Im WWW publizierte Dokumente zeichnen sich oftmals durch ihre Dynamik aus. Hyperlinks sprengen die Grenzen der analogen Vorstellung eines linearen Textes. Davon bleiben auch jahrtausende alte Kulturtechniken wie Schreiben und Lesen nicht unberührt (Neffe 2009). Dabei deutet die noch häufig zu beobachtende Imitation analoger Techniken unter digitalen Bedingungen darauf hin, dass wir uns derzeit in einer Übergangsphase vom analogen zum digitalen Zeitalter befinden, die noch nicht vollständig vollzogen ist.¹

Das stellt nicht zuletzt Bibliotheken in ihrer informationsvermittelnden und –bewahrenden Rolle vor große Herausforderungen und führt zu Verschiebungen im Bereich der genuin bibliothekarischen Aufgaben: Bibliotheken sehen sich heute im Rahmen von Open Access Modellen mit der Rolle des Begleiters von wissenschaftlichen Publikationsprozessen konfrontiert und übernehmen damit Aufgaben, die noch vor kurzem rein verlegerische Tätigkeiten waren. Sie sind gezwungen, sich mit den technischen Möglichkeiten und Grenzen digitaler Langzeitarchivierung auseinanderzusetzen und sich die dafür notwendige technische Expertise anzueignen. Das schnelle Anwachsen der Menge digitaler Informationsobjekte² erfordert neue Werkzeuge um sie zu ordnen.³ Und schließlich sehen sich Bibliotheken dem mit

¹ Beispiele dafür sind etwa PDF, das Print-Formate imitiert oder der (künstlich) begrenzte Zugang zu digitalen Ressourcen im Rahmen von Lizenzierungen elektronischer Zeitschriften.

² Siehe dazu auch Gantz et al. 2008: Im Rahmen der Studie der International Data Corporation (IDC) wurde errechnet, dass im Jahr 2008 etwa 3,9 Trilliarden Bit an neuen digitalen Informationen erzeugt wurden. Davon ausgehend prognostiziert die Studie, dass sich das „digitale Universum“ alle 18 Monate verdoppeln wird.

³ Siehe dazu Barbera et al. 2008: „The collection of resources available on-line is continuously expanding. It is now required to develop tools to access these resources in an intelligent way and search them as if they were part of a unique information space.“ (S.130) und Gradmann 2009A: „Angesichts dessen [der steigenden Anzahl der diskreten, digital prozessierbaren und semantisch unterscheidbaren Einheiten] benötigen wir neue Werkzeuge, um große Mengen digitaler Informationsobjekte einschließlich ihrer Konstituenten gedanklich zu organisieren und können dabei nicht mehr auf das althergebrachte bibliothekarische Katalogparadigma zurückgreifen. Wir müssen

„Web 2.0“ und Suchmaschinen vertrauten Nutzer gegenüber, dessen Suchverhalten durch interaktive Tools und ausgefeilte Suchmaschinentechnologie geprägt ist. Dieser (potentielle) Nutzer kann mit dem traditionellen OPAC nichts mehr anfangen – er konsultiert ihn häufig nicht einmal bei seiner Recherche, sondern beschränkt sich dabei auf Suchmaschinen.⁴ Bibliotheken haben – und auch das ist eine Folge der sich wandelnden Informationslandschaft unter digitalen Bedingungen – durch kommerzielle Informationsanbieter wie Google starke Konkurrenz in der Informationsvermittlung bekommen. Mit Angeboten wie Google Books und Google Scholar zielt Google auf bibliothekarische Kernbereiche (Lewandowski 2010). Um dem drohenden Bedeutungsverlust der eigenen Institution wirksam zu begegnen, bleibt den Bibliotheken künftig keine Alternative zu einer neuen Organisation ihrer Informationsressourcen. Karen Coyle weist darauf hin, dass Bibliotheken in der digitalen Abbildung ihrer Bestände trotz der rasanten technischen Entwicklungen der letzten 20 Jahre nach wie vor dem analogen Paradigma – dem Nachweis im Zettelkatalog – verhaftet sind (Coyle 2010). Stefan Gradmann schließt daran an, indem er bemerkt, dass ein Wandel im Denken von Informationsarchitekturen von einem Wandel im Begrifflichen begleitet werden müsse: statt mit „Dokumenten“ habe man es in WWW-genuinen Architekturen mit „Aggregationen“ zu tun, an die Stelle der „Suche“ trete die „Exploration“. Liege der Focus des analog geprägten Katalogisats auf der Trennung von Informationen in einzelne „records“, so stehe bei der anzustrebenden Organisation der Informationsressourcen ihre Verbindung und Kontextualisierung im Vordergrund (Gradmann 2010).

Bibliotheken, so lautet Coyles Forderung, müssten den Paradigmenwechsel vollziehen, ihre „Datensilos“ öffnen und in die offene Architektur des Web integrieren.

In diesem Zusammenhang spielt Linked (Open) Data⁵ eine bedeutende Rolle. Mit der Verwendung von W3C-Standards wie RDF(S) können Bibliotheken ihre Daten im Netz sichtbar und automatisch prozessierbar machen.⁶ Tim Berners-Lee, Begründer des World Wide Web Consortiums, sieht in Bibliotheken wichtige Akteure im Kontext von Linked Data. Von seinem Interesse an einer stärkeren Integration bibliothekarischer Daten ins WWW zeugt die im Mai

Mittel finden, bislang unvorstellbare Aggregationen strukturierter digitaler Informationen zu verstehen oder zumindest verstehbar zu machen.“ (S.2)

⁴ Eine Studie von OCLC ergab 2005, dass 84 Prozent der nach Informationen Suchenden ihre Recherche mit Suchmaschinen beginnen. Nur 1 Prozent beginnt die Suche in wissenschaftlichen Datenbanken oder Bibliothekskatalogen. (OCLC 2005)

⁵ Zur Unterscheidung zwischen Linked und Open Data siehe Pohl 2011: Linked Data bezieht sich vor allem auf die technischen, Open Data auf die rechtlich-politischen Voraussetzungen. Linked Data ist technisch auch funktionsfähig, ohne „open“ zu sein, allerdings ist die praktische Umsetzung dann wesentlich teurer, da sie aufwändige Authentifizierungsmechanismen erfordert. Für Kultureinrichtungen und die Verlinkung ihrer Daten ist das sicher kein denkbare Szenario. (Siehe dazu auch Gradmann 2010). Im Verlauf der Arbeit wird, soweit eine strikte Unterscheidung möglich ist, die Schreibweise Linked (Open) Data verwendet, solange es nicht explizit um Linked Open Data geht.

⁶ Siehe dazu auch Bermès et al. 2011: „The Web and Linked Data offers technical solutions for the integration of data stuffed away in „silos“.“

letzten Jahres bekanntgegebene Gründung der W3C Library Linked Data Incubator Group (LLD XG)⁷, die ihre Aufgabe wie folgt beschreibt:

The mission of the Library Linked Data incubator group is to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities—focusing on Linked Data—in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future.

The group will explore how existing building blocks of librarianship, such as metadata models, metadata schemas, standards and protocols for building interoperability and library systems and networked environments, encourage libraries to bring their content, and generally re-orient their approaches to data interoperability towards the Web, also reaching to other communities. It will also envision these communities as a potential major provider of authoritative datasets (persons, topics...) for the Linked Data Web.(...) (<http://www.w3.org/2005/Incubator/llid/>)

Antoine Isaac, Chair der W3C Library Linked Data Incubator Group, konstatiert mit Verweis auf die Linking Open Data Cloud (LOD) ein wachsendes Interesse an Linked Data in der Bibliothekslandschaft.⁸ Auf die Frage „Libraries and Linked Data – the perfect match?“ antwortet er mit dem Hinweis, dass Bibliotheken große Kompetenz und langjährige Erfahrung in der Produktion von Metadaten und Metadatenstandards haben – wichtige Bausteine für Linked Data (Isaac 2010B). Die Bedeutung bibliothekarischer Normdaten⁹ für das Linked Data Web nennt explizit die oben zitierte Charta der LLD XG.

Adrian Pohl weist darauf hin, dass Bibliotheken nicht nur bereits lange Daten verlinken, indem sie Verweise von Schlagwörtern und Personen auf Titeldaten dokumentieren¹⁰, sondern dass auch ihre Rolle als vertrauenswürdige, öffentliche Institutionen im Zusammenhang mit Linked Data von Bedeutung ist.¹¹

Welche Relevanz hat Linked (Open) Data nun aber für Bibliotheken und andere Kultureinrichtungen? Zum einen ermöglicht es eine domänenübergreifende Vernetzung von Daten aus dem Kulturbereich und darüber hinaus. Informationsressourcen können neu kontextualisiert und für die Nutzer mit Mehrwerten versehen werden. Dazu bedarf es bei einer

⁷ „Incubator Groups in the W3C are short-living groups (1 year) with light administration. The result is not a W3C Recommendation, but innovative ideas for specifications, guidelines, and applications that are not (or not yet) clear candidates as Web standards“ (Isaac 2010)

⁸ Nachdem an bibliothekarischen Daten in der LOD lange Zeit nur die Daten aus dem schwedischen Gesamtkatalog LIBRIS verfügbar waren, sind von deutscher Seite die Gemeinsamen Normdaten (GND) der Deutschen Nationalbibliothek und Daten der UB Mannheim hinzugekommen. Andere wichtige Akteure aus dem Bibliotheksbereich, die ihre Daten als Linked Open Data zur Verfügung stellen sind die Library of Congress mit den LoC Subject Headings (LCSH) und die British Library, die ihre gesamten bibliographischen Daten unter der Creative Commons Lizenz CC 0 in RDF zur Verfügung stellt.

⁹ Virtual International Authority Files (VIAF), Library of Congress Subject Headings (LCSH), Dewey Decimal Classification (DDC), Personennamendatei (PND) u.a.

¹⁰ Das geschieht jedoch derzeit vor allem intern – im „Silo“ -, auf der Grundlage proprietärer Formate und ohne die Beachtung wichtiger Standards wie stabilen URIs.

¹¹ Siehe Pohl 2011: „Warum sollen Bibliotheken ihre Daten als Linked Data publizieren? Im Prinzip verlinken Bibliothekare Daten spätestens seit der Etablierung des Kartenkatalogs, indem sie etwa Verweise von Schlagwörtern und Personen auf Titeldaten dokumentieren. Bibliotheken produzieren also schon lange verlinkte Daten. Außerdem sind sie vertrauenswürdige, öffentliche Institutionen, die tief in unserer Kultur verwurzelt sind und Vertrauenswürdigkeit ist eine wichtige Eigenschaft, die Linked Open Data haben sollte. Bibliotheken sind also geradezu prädestiniert dazu, das Linked-Data-Netz mit aufzubauen, zu pflegen und seine Vertrauenswürdigkeit und Stabilität zu sichern.“ Siehe dazu auch Gradmann 2009B

Modellierung der Daten in RDF keiner Datenkonvertierung, RDF-,Triple' aus unterschiedlichen Wissensdomänen können verhältnismäßig einfach zusammengeführt werden (Voß 2009). Das vermeidet Redundanzen, die heute zu wenig hinterfragt werden: Daten müssen nicht länger von jeder Institution einzeln erzeugt und vorgehalten werden.¹² Die frei verfügbaren Daten im Netz referenzieren aufeinander und können wie eine einzige verteilte Datenbank abgefragt werden. Die Modellierung der Daten in RDF und die Verwendung von Ontologien ermöglichen dabei eine ontologiebasierte Suche, die das Potential der Volltextsuche bei weitem übertrifft (Gradmann 2009B).

JISC (Joint Information Systems Committee) beantwortet in dem von ihm veröffentlichten Open Bibliographic Data Guide die Frage „Why do it?“ schlicht mit:

The core rationale is about discoverability and is gaining in credibility the more our resources are discovered from 'out there' (through such as Google) and not from 'in here' (through the local OPAC). (JISC)

Die genannten Gründe zielen vor allem auf eine bessere Funktionalität in der Suche nach Informationsressourcen und beziehen sich vorrangig auf das „Linked“ in Linked Open Data. Von politischer Seite nehmen die Forderung nach und die Förderung von Open Data zu, etwa durch Initiativen wie der Open Knowledge Foundation oder die Veröffentlichung von Daten der öffentlichen Hand wie Data.gov, Data.gov.uk und Mashup Australia. Auf der Berlin7 Konferenz in Paris 2009 wurden Metadaten in die Forderungen nach Open Access eingeschlossen (Bérard 2011, S.322f).

Wichtige Akteure im Bereich von Open Library Data sind Open Library, ein Projekt von Internet Archive, das auf offener Software und offenen Daten basiert und Biblios.net, das weltgrößte Repository frei verfügbarer bibliothekarischer Daten. Europeana ist auf dem Weg dahin, ein bedeutender Akteur für Linked Open Data aus dem Kulturbereich zu werden. Das derzeit in der Phase des Prototyping befindliche Europeana Data Model basiert auf dem Resource Description Framework (RDF) und ermöglicht eine Integration der von Kultureinrichtungen an Europeana gelieferten Daten in die LOD Cloud (Gradmann 2010).

Während die Open Bibliographic Data-Bewegung in der Bibliothekswelt nach und nach eine nennenswerte Anhängerschaft verzeichnen kann¹³, gibt es auch Stimmen dagegen: sie

¹² Ein gelungenes Beispiel für die Vermeidung solcher Redundanzen durch die Nutzung von LOD ist auf dem Portal „Museum Digital“ zu sehen, das zur Bereitstellung von Informationen über die ausgestellten Künstler auf die als LOD zur Verfügung gestellten Personennormdaten der DNB verlinkt (Svensson 2011)

¹³ Nach der DNB sind in Deutschland nun auch das hbz in Köln, die UB Mannheim, die UB Tübingen und die UB Konstanz dabei, ihre Daten als Linked Open Data zur Verfügung zu stellen bzw. haben das bereits getan. Siehe dazu auch Eckert 2011: „Nachdem die Universitätsbibliothek Mannheim im Juli 2010 als erste deutsche Bibliothek Katalogdaten als Linked Data veröffentlicht hat, der eigene Katalog also technisch in RDF aufbereitet wurde, sind die Katalogdaten ab sofort auch offen unter der CC0-Lizenz freigegeben und stehen damit jedem zur uneingeschränkten Nachnutzung zur Verfügung, und zwar sowohl im Ursprungsformat, als auch in der RDF-Aufbereitung.“ Adrian Pohl weist darauf hin, dass dies, so begrüßenswert diese Aktivitäten sind, erst der Anfang ist: mit der Publikation von Rohdaten allein ist es nicht getan. „Die Veröffentlichung der Daten unter einer freien Lizenz ist sicher ein notwendiger politisch-rechtlicher Schritt, der wichtige Signale setzt. Neben den rechtlichen und

gehören zu einem kommerziellen Anbieter bibliographischer Daten, wie OCLC oder Skyriver, deren Geschäftsinteressen in diametralem Gegensatz zu Open Data stehen.¹⁴ Zum anderen äußern sich Bibliothekare besorgt darüber, die Kontrolle über die Verwendung ihrer Daten mit deren Freigabe zu verlieren.¹⁵ In einer neuen und oftmals unerwarteten Nachnutzung oder Kontextualisierung der Daten liege aber das Potential für die Schaffung von Mehrwerten – so sieht es Tim Berners-Lee:

It is the unexpected re-use of information which is the value added by the web.
(Berners-Lee 2006)

In der vorliegenden Arbeit geht es um die Reorganisation von digitalen bibliothekarischen Informationsressourcen auf der Grundlage von W3C-Standards und ihre Integrierbarkeit in WWW-genuine Informationsarchitekturen.

Die Arbeit bewegt sich im Rahmen des Prototyping¹⁶ des Europeana Data Model (EDM), das mit dem nächsten Release des Europeana-Portals, voraussichtlich im Sommer dieses Jahres, die bislang angewendeten Europeana Semantic Elements (ESE) ablösen wird. Ein wesentliches Kapitel der Arbeit besteht in einem Mapping von Metadaten im bibliothekarischen METS/MODS-Format aus dem VD 17 Projekt der Staatsbibliothek zu Berlin auf das WWW-kompatible Europeana Data Model. Die Daten wurden von der Staatsbibliothek zur Verfügung gestellt und sind Teil des offiziellen Prototyping des EDM. Das Mapping erfolgte in enger Absprache mit Kolleginnen und Kollegen aus der Staatsbibliothek, teils im Rahmen eines Praktikums, teils in regelmäßig stattfindenden Konsultationen.

Steffen Hennicke, Mitarbeiter des Europeana Projekts an der Humboldt-Universität zu Berlin, erstellte, analog zu dem Mapping von METS/MODS, ein Mapping der zu den Datensätzen gehörigen Normdaten im PICA Format auf das EDM. Die Normdaten mussten dafür vom Gemeinsamen Bibliotheksverbund geharvestet werden, da die internen Links der PICA-

politischen Aspekten von Open Data ist es aber ebenso wichtig, zum einen die Entstehung einer Open-Data-Community und –Praxis zu fördern und zum anderen eine technische Open-Data-Infrastruktur zu entwickeln, welche diese Open-Data-Praxis unterstützt.“ (Pohl et al. 2010, S.206)

¹⁴ Zu welchen Kontroversen dieser Interessenskonflikt führt, zeigten die massiven Proteste, mit denen die Bibliotheks-Community 2008 auf den Entwurf der neuen OCLC-Richtlinien in den USA reagierte: man sah darin den Versuch seitens OCLC eine Monopolstellung zu etablieren, indem den Mitgliedern Restriktionen hinsichtlich des Datenaustauschs untereinander auferlegt wurden. OCLC sah sich gezwungen, den Entwurf zurückzuziehen und einen 13-köpfigen Rat von Bibliothekaren mit dem Entwurf neuer Richtlinien zur Nutzung und zum Austausch von Datensätzen zu beauftragen. In diesen Richtlinien ist explizit festgelegt, dass die WorldCat Daten nicht gemeinfrei im Sinne von Open Bibliographic Data, sondern dass sie vielmehr Besitz der Mitglieder von OCLC seien, die entsprechende Restriktionen für den Zugang zu den Daten bestimmen. (Siehe Bérard 2011)

¹⁵ Unsicherheiten bestehen auch nach wie vor, was die Rechtslage zu Open Bibliographic Data betrifft: sind bibliographische Datensätze urheberrechtlich geschützt? Datensätze fallen nicht unter den Schutz des Urheberrechts, da es sich um „rohe Information“ handelt, wie etwa Listen mit Namen oder Städten. Datenbanken, wie im Falle des WorldCat, können hingegen dann Urheberrechtsschutz geltend machen, wenn der Betreiber Investitionen finanzieller, technischer und personeller Art zum Aufbau und Betrieb der Datenbank nachweisen kann. (Siehe Bérard 2011)

¹⁶ Für weitere Metadaten-Standards, die bislang im Rahmen des Prototyping auf das EDM gemappt wurden, siehe Hennicke et al. 2011.

Datensätze bei der Umwandlung in METS/MODS nicht abgebildet werden können¹⁷. Mit Hilfe dieses zweiten Mappings können die Normdaten in die Suche integriert werden.

Beide Mappings dienen als Vorlage für ein Konversionsskript, das Steffen Hennicke gemeinsam mit Mitarbeitern des Europeana Projekts der Vrije Universiteit Amsterdam erarbeitet. Auf der Grundlage des Skripts werden die etwa 2500 über die OAI-PMH Schnittstelle der Staatsbibliothek geharvesteten Datensätze aus dem VD 17 Projekt in RDF-,Triple' konvertiert, auf denen das EDM basiert. Anschließend können die Daten zunächst in der Umgebung des Europeana Thoughtlabs mit dem Prototypen einer semantischen Suchmaschine¹⁸ durchsucht werden, bevor die semantisch basierte Suche mit dem nächsten Release für das gesamte Europeana Portal ermöglicht wird.¹⁹

Wie können mit graphenbasierten Linked Open Data - Modellen Mehrwerte für existierende bibliographische Metadaten geschaffen werden? Dieser Frage wird im Rahmen der vorliegenden Arbeit anhand des Mappings der VD 17 Daten im METS/MODS Format auf das EDM nachgegangen.

Dafür sollen in einem ersten Schritt das Europeana Data Model und die darin enthaltenen Standards erläutert werden. Dabei stellt sich auch die Frage, wie sich das EDM zu Linked Open Data und dem Semantic Web verhält.

Das zweite Kapitel gibt einen genaueren Einblick in die VD 17 Daten der Staatsbibliothek. Wie kommen sie zustande und welche Inkonsistenzen sind vor dem Mapping festzustellen? Wie verhalten sich die XML-Formate METS und MODS zueinander?

Das dritte Kapitel erklärt das Mapping anhand einer graphischen Darstellung der Daten in RDF und nimmt dabei noch einmal Bezug auf die Designprinzipien des EDM.

Im vierten und letzten Kapitel schließlich soll ein Ausblick auf die zu erwartenden Mehrwerte für die Daten durch ihre Modellierung mit dem graphenbasierten EDM in einem Vergleichsszenario erfolgen: welche Möglichkeiten der Suche bieten die Digitalisierten Sammlungen der Staatsbibliothek Berlin heute und welche Suchoptionen werden künftig über das Europeana Portal möglich sein? In diesem Zusammenhang sind auch community-relevante Forschungsfragen zu berücksichtigen, wie sie sich speziell für die Drucke des 17. Jahrhunderts ergeben mögen.

¹⁷ Das war bei den bislang verwendeten MODS Versionen der Fall. Eine wesentliche Neuerung bei der kürzlich veröffentlichten Version 3.4 ist die Möglichkeit, die Links abzubilden. Darauf soll aber in dieser Arbeit nicht näher eingegangen werden.

¹⁸ Der Prototyp für semantische Funktionalitäten in Europeana operiert derzeit auf der Grundlage von Objektrepräsentationen aus drei Institutionen aus dem Museumsbereich (Louvre, Rijksmuseum und Rijksbureau voor Kunsthistorische Documentatie). Sie werden durch teils hausinterne, teils lizenzierte oder offene kontrollierte Vokabulare semantisch kontextualisiert. (Europeana Thoughtlab)

¹⁹ Leider wird die Datenkonversion in der zeitlichen Frist, die dieser Arbeit gesetzt ist, nicht zu einem Abschluss kommen, so dass neue Abfragemöglichkeiten der Daten lediglich prognostiziert werden können.

1. Das Europeana Data Model

1.1 *Europeana – mehr als ein Portal*

Europeana ist ein von der Europäischen Kommission gefördertes, domänenübergreifendes Internetportal, das derzeit Zugang zu über 15 Millionen digitalen Objekten in Text-, Video-, Ton- und anderen Formaten europäischer Bibliotheken, Archive und Museen bietet. Europeana macht als „Gedächtnis Europas“ erstmals kulturelle und wissenschaftliche Informationsressourcen aus unterschiedlichen Institutionen und Ländern Europas über ein Portal auffindbar. Dabei werden seitens der Europeana nur die Metadaten und eine Voransicht der Objekte vorgehalten und auf die digitale Präsentation des Inhaltsanbieters, der nach wie vor die Kontrolle über die Präsentation der Objekte auf seiner Internetpräsenz behält, wird verlinkt. Um die Objekte der verschiedenen Anbieter über ein zentrales Suchsystem auffindbar zu machen, ist eine Abbildung der gelieferten Metadaten auf das in Europeana verwendete Datenmodell erforderlich.

Noch erfolgt die Suche in Europeana mit den auf dem Dublin-Core Modell basierenden Europeana Semantic Elements (Europeana Semantic Elements 2010), die keine Verlinkungen auf externe Ressourcen erlauben und die Metadaten der Inhaltsanbieter nicht in ihrer Komplexität abbilden können. Das geplante Danube Release im Sommer 2011 umfasst die Implementierung des Europeana Data Model und wird eine semantisch basierte Suche ermöglichen.

Doch, so Concordia et al., sei Europeana mehr als ein Portal, das Zugang zu einer großen Zahl digitaler Objekte biete.²⁰ In ihrer Betrachtung legen sie den Fokus auf Europeana als Application Programming Interface (API). Europeana, so die Autoren, kontextualisiere die Repräsentationen der gelieferten Objekte untereinander und reiche sie semantisch an. Dabei bleibe es aber nicht: die entstandenen Mehrwerte können über die API auch an die Inhaltsanbieter zurückfließen. Der Datenfluss zwischen ihnen und Europeana erfolge bidirektional. Auf diese Weise biete Europeana den beteiligten Kulturinstitutionen einen Migrationspfad aus ihren Datensilos in eine WWW-genuine Informationsarchitektur:

²⁰ Siehe dazu auch Doerr et al. 2010: „Europeana is often presented in public as a portal giving access to millions of objects from all kinds of cultural heritage communities - and even though this way of presenting Europeana conveys some truth it doesn't capture some of the essential characteristics of what Europeana actually is designed to be in the end. Concordia et al. (2010) have tried to make clear that Europeana is not so much a portal characterised by sheer volume, but that the core agenda of our endeavour is to make rich data and functionality available on an API basis. This would allow all kinds of external communities to make use of our rich (and numerous) representations of European cultural treasures for their own needs - and the Europeana portal that is offered at <http://www.europeana.eu> should in the end be seen as one of the parties making use of this wealth of data and functionality by means of the API.“

Europeana thus offers cultural heritage institutions a migration path from their current collection silos into a layered, web service-based information architecture and is conceived as an environment facilitating – and requiring – the mentality shift cultural heritage institutions will have to operate in the future, anyway. (Concordia et al. 2010)

1.2 Semantic Web und Linked (Open) Data

Linked (Open) Data und Semantic Web-basierte Funktionalitäten waren von Beginn an ein zentraler Baustein im Aufbau der Europeana:

A central principle for building Europeana is that a network of semantic resources will be used as the primary level of user interaction. In a traditional librarian catalogue model all user access to information objects is mediated by descriptive metadata (...). Unlike in such librarian functional models users are expected to explore the Europeana data space using semantic nodes as primary elements for searching and browsing along paradigms indicated by the questions as to “Who?”, “Where?”, “When?” and “What?” (Dekkers et al. 2009, S.15)²¹

Bevor nun im Detail gezeigt wird, wie das EDM die technischen Voraussetzungen für die Umsetzung dieser Funktionen schafft, sei in groben Zügen erläutert, worum es sich bei dem Semantic Web handelt und in welchem Verhältnis es zu Linked (Open) Data steht.

Tim Berners-Lee beschrieb seine Vorstellung von einem Semantic Web in einem gemeinsam mit James Hendler und Ora Lassila verfassten und seitdem viel zitierten Aufsatz bereits 2001. Das World Wide Web, so die Verfasser, habe sich zu einem Web der primär für Menschen lesbaren Dokumente entwickelt und gegen den ursprünglich im WWW angelegten Ansatz eines Web der automatisch prozessierbaren Daten.

By augmenting Web pages with data targeted at computers and by adding documents solely for computers, we will transform the Web into a Semantic Web. (...)The Semantic Web is not a separate Web but an extension of a current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. (...) For the Semantic Web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. (Berners-Lee et al. 2001, S.36f)

Zentral für das Semantic Web ist also die Maschinenlesbarkeit der Daten. Sie wird dadurch erreicht, dass Informationen „wohl definierte Bedeutungen“ zugeordnet bekommen. Das geschieht auf der Grundlage von den W3C-definierten Standards Extensible Markup Language (XML) und Resource Description Framework (RDF) sowie dem Einsatz der stabilen Uniform Resource Identifier (URIs) und kontrollierter Vokabulare.

²¹ Siehe dazu auch Gradmann 2009A: „Dort [in der Europeana] realisieren wir eine digitale Bibliothek neuen Typus, in der digitalisierte und genuin digitale Kultur- und Informationsobjekte primär über semantisch basierte Zugriffswege zugänglich werden.“ (S.11)

In short, XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean (...). (...) Meaning is expressed by RDF, which encodes it in sets of triples, each triple being rather like the subject, verb and object of an elementary sentence. (...) The triples of RDF form webs of information about related things. Because RDF uses URIs to encode this information in a document, the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web.²²

Der Einsatz von Ontologien, die Objektklassen und die Beziehungen zwischen ihnen festlegen und Inferenzregeln, ermöglicht ein automatisches Generieren von Schlussfolgerungen aus den strukturierten Daten. Durch die Verwendung offener, vom W3C empfohlener Standards wird ein gemeinsamer Rahmen geschaffen, der es erlaubt, Daten automatisch zu prozessieren und über die Grenzen von einzelnen Anwendungen und Institutionen hinweg zu teilen, nachzunutzen und neu zu kontextualisieren.

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. (W3C Semantic Web)

In RDF modellierte, dezentral vorliegende Daten liegen im Web wie in einer verteilten Datenbank, die komplexe Abfragen und Schlussfolgerungen ermöglicht (Pohl 2011). Damit ergeben sich neue Suchansätze, die die Ergebnisse heute gängiger, stichwortbasierter Suchmaschinen bei weitem übertreffen: die semantische Vernetzung ermöglicht etwa die automatische Recherche nach bestimmten Themen (Gradmann et al. 2009). Semantic Web Technologien bieten neben einer verbesserten Informationsvernetzung und Informationsaggregation also auch einen wesentlichen Mehrwert für das Information Retrieval.²³

Wie kommt es angesichts dessen aber nun zu der Wahl des häufig kritisierten²⁴ Attributs „semantic“ für das Semantic Web? Zum einen, weil mit dem Einsatz des W3C Standards XML mit Tags wie <Autor> Informationen mit einer gewissen, maschinenlesbaren Bedeutung versehen werden – vor allem im Vergleich zu HTML. Zum anderen weil RDF und stabile URIs für eine genaue Definition dieser Bedeutung sorgen:

There are some semantics built into the middle third of all RDF triples, because the requirement that a predicate use a full URL means that I can't just say "title" there, leaving you to wonder whether I'm talking about a job title, the deed to a piece of property, or the title of a work; I have to say something like <http://purl.org/dc/elements/1.1/title> to make it clear that I mean the title of a work. In other words, I must make the semantics of the triple's predicate clear. (DuCharme 2008)

²² Berners-Lee et al. 2001, S.38ff

²³ Siehe dazu auch Blumauer et al. 2009: Informationsvernetzung basiert im Regelfall auf semantischer Ähnlichkeit. Diese Ähnlichkeit wird im Semantic Web nicht länger auf rein begrifflicher Ebene festgestellt, sondern auf der Grundlage von Wissensmodellen und Ontologien. (S.8)

²⁴ Siehe dazu Gradmann 2009A über den Begriff „Semantic Web“ : „ (...) es handelt sich auch von seiner Programmatik her eben wirklich nur um einen transaktions- oder logikzentrierten Ansatz, für den wahrscheinlich die Bezeichnung „Syntactic Web“ angemessener gewesen wäre.“ (S.9)

Darüber hinaus stellt aber auch der eben zitierte Text grundsätzlich die Frage:

Why do we describe technology for easier integration of machine-readable data on the web as "semantic"? (DuCharme 2008)

Zweifellos muss das Attribut "semantic" in diesem Zusammenhang sehr formal verstanden werden, insofern, als einem sprachlichen Zeichen ein genau abgegrenztes und unveränderliches („well-defined“) Denotat zugeordnet wird. Das hat mit Bedeutung im Sinne natürlicher Sprachen nichts zu tun, da es für natürliche Sprachen so konstitutive Elemente wie Konnotationen oder Polysemien ausschließt. Besonders wenig angebracht erscheint der Begriff „semantic“, wenn man davon ausgeht, dass mit jeder Bedeutung auch ein Verstehen einhergeht: von verstehenden Maschinen kann auch im Semantic Web noch keine Rede sein, aber - und das ist ein großer Gewinn – sie werden in die Lage versetzt, Informationen automatisch zu aggregieren und zu vernetzen, um daraus, in einem gewissen Rahmen, Wissen zu generieren.²⁵

The computer doesn't truly "understand" any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user. (Berners-Lee et al. 2001, S.40)

Tim Berners-Lee spricht heute selbst kaum mehr vom „Semantic Web“ sondern nutzt die Bezeichnungen Linked (Open) Data oder Giant Global Graph, die weniger die Semantik von Informationsressourcen als deren Vernetzung fokussieren. Das hängt auch damit zusammen, dass die Entwicklung des Semantic Web bis 2006 hauptsächlich in proprietären Anwendungen stattfand, vor allem in Firmennetzen, und im World Wide Web wenig präsent war. Um eine brauchbare Datengrundlage für das Semantic Web zu fördern, veröffentlichte Berners-Lee 2006 seine Design Issues zu Linked Data:

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

1. Use URIs as names for things

2. Use HTTP URIs so that people can look up those names.

3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)*

4. Include links to other URIs so that they can discover more things. (Berners-Lee 2006)

Adrian Pohl beschreibt den Zusammenhang zwischen Semantic Web und Linked Data treffend, indem er Linked Data als ersten Schritt und Herzstück des Semantic Web bezeichnet. Linked Data, so Pohl weiter,

(...) legt also das Augenmerk auf die ersten Schritte, auf das Veröffentlichen und Verlinken von Daten, während im Semantic Web Anwendungen, die auf maschinell durchführbaren logischen Schlussfolgerungen basieren, die entscheidende Rolle spielen. Es ist aber zu konstatieren, dass

²⁵ Andreas Blumauer und Tassilo Pellegrini schlagen in Anlehnung an John F. Sowa den Begriff des „Semiotic Web“ vor, wobei so wichtige Aspekte wie die Beziehungen der Zeichen zu Interpreteten und Kontexten in der Debatte um das Semantic Web mit zu berücksichtigen wären. (Blumauer et al. 2006, S.10)

im gegenwärtigen Sprachgebrauch ‚Semantic Web‘ und ‚Linked Data‘ häufig synonym verwendet werden. (Pohl 2011)

1.3 Der Aufbau des Europeana Data Model

Im Folgenden werden einzelne Elemente des Europeana Data Model, Namensräume, auf denen sie basieren, und die wichtigsten Designprinzipien vorgestellt. Das EDM besteht aus verschiedenen Klassen und Eigenschaften, die teils über den Namensraum Europeana (ens) eingeführt, teils aus anderen Namensräumen übernommen werden.²⁶ Die größtmögliche Nachnutzung bestehender Standards war bei der Entwicklung des EDM ein wichtiger Grundsatz.

Konstitutiv für den Informationsraum in Europeana ist ein Zwei-Schichtenmodell, das aus einer Schicht mit Objektrepräsentationen und dem darüber liegenden semantischen Netzwerk besteht:

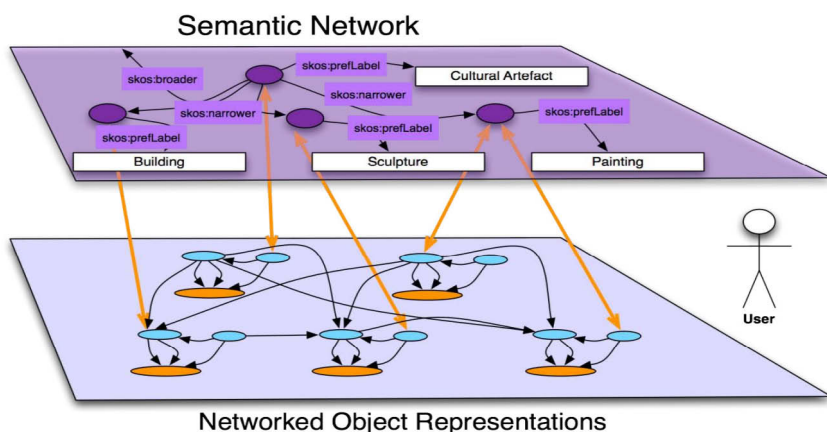


Abbildung 1.1 - Schichtenmodell der Objektrepräsentationen in Europeana (aus: Doerr et al. 2011)

Die untere Schicht bilden die vernetzten Objektrepräsentationen. Darüber liegt die Schicht der semantischen Kontextualisierung, die über SKOS mit verschiedenen Wissensorganisationssystemen (Thesauri und Vokabularen) operiert. Sie ist mit den Objektrepräsentationen verlinkt und reichert sie semantisch an.²⁷

²⁶ Siehe dazu auch Doerr et al. 2010, S.3: „The Semantic Web approach indeed enables the combination of various ontologies in the same descriptions. One can deploy different views on top of the same assertions, or build assertions that combine different vocabularies tailored to specific needs so as to match the requirements of a more general application. RDF data for a book catalogue may thus re-use an ontology for book description, to represent the core data of book records, and an ontology for persons to finely describe the authors of the books, as given in an authority file. EDM will re-use some of the reference ontologies already available (...).“

²⁷ Siehe dazu auch Doerr et al. 2010, S.2: „The idea furthermore is to offer rich semantic contextualisation for the object representations in Europeana in such a way as to enable complex semantic operations on these resources in a way that would not be supported by a traditional digital library interface. In order to enable such functionality the object representations in Europeana need to be systematically connected to Linked Open Data on the WWW or else

Eine wesentliche Anforderung an das EDM besteht darin, Objektrepräsentationen unterschiedlicher Inhaltsanbieter zu einem Netzwerk von ‚Resources‘ zu verbinden.

For example, a digital object from Provider A may be contextually enriched by metadata from Provider B. It may also be enriched by the addition of data from the authority files held by Provider C, and a web-based thesaurus offered by Publisher D. (Isaac 2010A, S.4)

Dazu bedient sich das EDM der vom W3C empfohlenen Standards RDF und RDF Schema zur Datenmodellierung und Datenrepräsentation (Gradmann 2010).

RDF steht für Resource Description Framework und ist eine graphenbasierte Beschreibungssprache für Web-‚Resources‘. Grundelemente dieser Sprache sind die zu beschreibenden Entitäten oder ‚Resources‘, ihre Eigenschaften oder ‚Properties‘ und die Eigenschaftswerte oder ‚Values‘. In RDF-‚Triples‘ können damit über die ‚Resources‘ Aussagen wie in folgendem Beispiel getroffen werden:

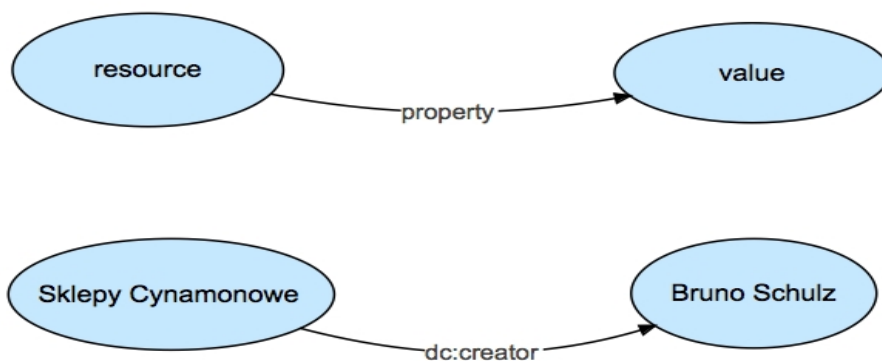


Abbildung 1.2 - RDF-‚Triple‘

RDF Schema (RDFS) erfüllt die Funktion einer Grammatik für RDF-Statements und bildet die Basis für den Aufbau von Ontologien mit Hierarchien von Klassen (rdfs:class) und Unterklassen (rdfs:subclass), Eigenschaften (rdfs:property) und Untereigenschaften (rdfs:subproperty) und Inferenzregeln.²⁸

RDF Schema Ontologien, die im EDM verwendet werden, sind SKOS, OAI ORE und DCMI Metadata Terms.

Um die einzelnen Elemente in RDF-‚Triples‘ eindeutig referenzierbar zu machen, werden sie mit URIs verbunden. URI steht für Uniform Resource Identifier und ist im WWW ein Bezeichner für

to semantic contextualisation resources held within the Europeana data space such as thesauri and structured vocabularies migrated to the SKOS standard.”

²⁸ Siehe dazu auch Brickley et al. 2004: „This document does not specify a vocabulary of descriptive properties such as "author". Instead it specifies mechanisms that may be used to name and describe properties and the classes of resource they describe.“

Web-Seiten oder andere Dateien. Die Idee der URIs kann aber generell als Mechanismus zur Erzeugung eindeutiger Bezeichner genutzt werden (Hitzler et al. 2008, S.26). So kann in unten stehendem Beispiel in einem maschinenlesbaren Format ausgedrückt werden, dass es sich bei dem Titel um einen Werktitel, nicht um einen akademischen Titel handelt, indem folgender URI verwendet wird: <http://purl.org/dc/elements/1.1/title>. Er stammt aus dem Namensraum des Dublin Core Metadata Element Set und definiert Titel mit *A name given to a resource*.

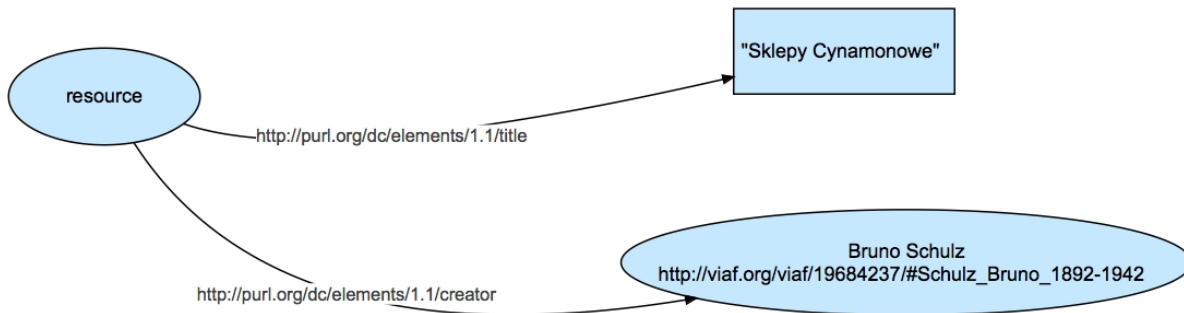


Abbildung 1.3 - Der Gebrauch von ‚Literals‘ und URIs im Zusammenhang mit dem RDF

Das Objekt ist das einzige Element in einem ‚Triple‘, das nicht zwingend mit einem URI versehen sein muss, sondern als einfache Zeichenfolge oder ‚Literal‘ stehen kann - in der Graphik durch die eckige Umrandung gekennzeichnet.

Ein Ziel für die Modellierung der Daten in Europeana ist es jedoch, aus möglichst vielen ‚Literals‘ ‚Resources‘ - und sie so anschlussfähig für neue Vernetzungen zu machen.

It [Europeana] will also create URIs for all objects, so as to implement a linked data publication strategy that relies on Europeana's own (HTTP) services. (Isaac 2010A, S.18)

Für die semantische Anreicherung einer Objektrepräsentation kann so zum Beispiel der Name einer Person mit dem entsprechenden URI mit Normdaten (etwa den Gemeinsamen Normdaten oder, wie in obigem Beispiel, den Virtual International Authority Files) verbunden werden, um mehr Informationen über die Person verfügbar zu machen.

Durch die Verlinkung mit Thesauri können in der Suche automatisch Empfehlungen generiert werden, die auf semantischen Beziehungen aufbauen.

Klassen, die im EDM zur Repräsentation von kontextualisierenden Entitäten dienen, sind `ens:Agent` für Personen und Organisationen, `ens:Place` für Orte, `ens:TimeSpan` für Zeiträume oder Daten und das aus SKOS übernommene `skos:Concept` für Entitäten aus Wissensorganisationssystemen.

SKOS (Simple Knowledge Organization System) ist ein weiterer im EDM angewandeter W3C-Standard. Er definiert ein Modell, um Elemente aus Wissensorganisationssystemen wie

Thesauri und Klassifikationen in RDF auszudrücken.²⁹ SKOS modelliert Thesauri und Klassifikationen in der semantischen Datenschicht von Europeana und ermöglicht Verbindungen zwischen ihnen. Wichtige, im EDM verwendete Elemente, sind die ‚Properties‘ `skos:prefLabel` und `skos:altLabel`, die für ‚preferred‘ und ‚alternative Label‘ stehen und die Möglichkeit bieten, verschiedene Bezeichnungen für ein `skos:Concept` parallel zu verwenden. `Skos:narrower`, `skos:broader` und `skos:related` drücken semantische Beziehungen zwischen Konzepten aus und mit `skos:exactMatch` können bedeutungsgleiche Konzepte aus verschiedenen Thesauri miteinander verbunden werden.

Die aus dem Namensraum DCMI Metadata Terms (Dublin Core Metadata Terms) übernommenen Elemente bilden im EDM die Basis für semantisch interoperable deskriptive Objekt-Metadaten. Sie stellen eine Spezialisierung der ursprünglich 15, generischen Dublin Core Elements (Dublin Core) dar. Die Nutzung von Dublin Core Metadata Terms und Elements im EDM stellt zum einen die Kompatibilität des Modells mit den in Europeana Semantic Elements abgebildeten Daten her, was angesichts der Menge von über 15 Millionen Objekten, deren Metadaten Europeana derzeit vorhält, von großer Bedeutung ist. Zum anderen handelt es sich bei den DC Terms um ein im Netz weit verbreitetes Metadaten Format, das für den domänenspezifischen Bedarf spezialisiert werden kann. Das Europeana Data Model ist ein domänenübergreifendes Modell. Die in Europeana mit Objekten vertretenen Domänen (Bibliotheken, Museen und Archive) nutzen unterschiedliche granulare Modelle³⁰, deren Abbildung das EDM leisten muss. Dieser Anforderung kommt der generische Ansatz der Dublin Core Elemente entgegen. Die spezifischen Metadaten der Inhaltsanbieter bestehen neben den generischen des EDM weiter und werden über `rdfs:subPropertyOf` und `rdfs:subClassOf` mit ihnen verbunden. So erhält das Europeana Data Model in der Funktion einer Top-Level Ontologie die Semantik der Originaldaten ohne den Verlust der Interoperabilität.³¹

Zu den wichtigsten Designprinzipien im Europeana Data Model gehört die Unterscheidung zwischen dem Objekt (Gemälde, Buch, Film etc.) und einer oder mehreren digitalen Repräsentationen. Um diese Unterscheidung in der Modellierung zu ermöglichen, wurden im Europeana Namespace die Klassen `ens:PhysicalThing`³² und `ens:WebResource` eingeführt.

²⁹ Siehe dazu Isaac et al. 2009: “The Simple Knowledge Organization System (SKOS) is an RDF vocabulary for representing semi-formal knowledge organization systems (KOSs), such as thesauri, taxonomies, classification schemes and subject heading lists.”

³⁰ Wichtige Standards, die abzubilden das EDM in der Lage sein muss, sind LIDO und CIDOC-CRM für den Museumsbereich, METS/MODS und MARC für Bibliotheken und EAD für Archive.

³¹ Siehe dazu auch Doerr et al. 2010, S.7: „The original richness of community standards like LIDO, CIDOC CRM, MARC or EAD can thus be preserved whereas access by implicit toplevel relationships will guarantee unprecedented recall and precision across data provided in various formats.“

³² Die Einführung dieser Klasse wurde von dem Modell CIDOC CRM motiviert, einem sehr granularen Modell zur Abbildung von Objekten aus dem Museumsbereich. Entsprechend geben die „Definition of the Europeana Data

Diese Art der Modellierung erlaubt die Unterscheidung in ein Werk³³, das im Fokus des Interesses der Nutzer steht und die digitalen Objektrepräsentationen im Informationsraum Europeana.

Ein weiteres wichtiges Designprinzip besteht in der Unterscheidung zwischen dem Objekt und den deskriptiven Metadaten. Da mehrere Inhaltsanbieter Datensätze für ein Objekt bereitstellen können, muss das EDM verschiedene, eventuell sogar einander widersprechende Aussagen über ein Objekt ermöglichen. Dafür nutzt das EDM die Klasse ore:Proxy aus dem Namensraum OAI ORE - Open Archives Initiative Object Reuse and Exchange (Isaac 2010 A S. 12 und Lagoze et al. 2008 B). Der Proxy-Mechanismus unterstützt verschiedene Sichten auf ein Objekt. Liefern Inhaltsanbieter verschiedene Datensätze zu einem Objekt, werden mehrere Proxys über ore:proxyFor mit dem ens:PhysicalThing verbunden. Für den (wahrscheinlichen) Fall, dass die Inhaltsanbieter für dasselbe Objekt zwei unterschiedliche URIs vergeben haben, werden die URIs über owl:sameAs miteinander verbunden (Isaac 2010 A, S.13).

Dem Ansatz von OAI ORE³⁴ folgend, betrachtet das EDM das Objekt, alle dazugehörigen Bestandteile eines Metadatensatzes und sämtliche digitalen Repräsentationen eines Inhaltsanbieters als Aggregation. Sie wird als ore:Aggregation präsentiert und im Informationsraum Europeana mit dem ens:PhysicalThing über die ‚Property‘ ens:aggregatedCHO und einer oder mehreren digitalen Repräsentationen über die ‚Property‘ ens:hasView oder ens:landingPage verbunden.

Im Zentrum einer Datenmodellierung mit dem EDM stehen demnach die vier Klassen ens:PhysicalThing, ens:WebResource, ore:Aggregation und ore:Proxy:

Model elements“ als äquivalente Klasse zum ‚PhysicalThing‘ das E18_Physical_Thing aus CIDOC CRM an. (Definition 2010, S.12)

³³ Diese Unterscheidung erinnert an diejenige zwischen „Work“ und „Manifestation“ aus den Functional Requirements for Bibliographic Records. Mit FRBR ist es u.a. möglich, verschiedene Auflagen eines Werks abzubilden. Derzeit ist FRBR nicht Teil des EDM, es ist aber eine Erweiterung des EDM um die FRBR geplant.

³⁴ Siehe dazu auch Gradmann et al. 2010, S.1: „(...) OAI ORE [ist] eine im Oktober 2008 in der Version 1.0 veröffentlichte Spezifikation der Open Archives Initiative, die Standards für die Beschreibung und den Austausch zu komplexen Einheiten aggregierter WWW-Entitäten („resources“) definiert. In diese Aggregationen können verteilte Quellen mit multiplen Medientypen wie Text, Bild, Daten und Video zu komplexen Entitäten kombiniert sein. Das Ziel von ORE ist es, den Inhalt und die Binnenstruktur solcher komplexer Aggregationen so zu modellieren, dass diese für Anwendungen verarbeitbar werden, welche ihre Erstellung, das Einstellen, den Austausch, die Visualisierung, Weiterverwendung und Langfristarchivierung ermöglichen.“

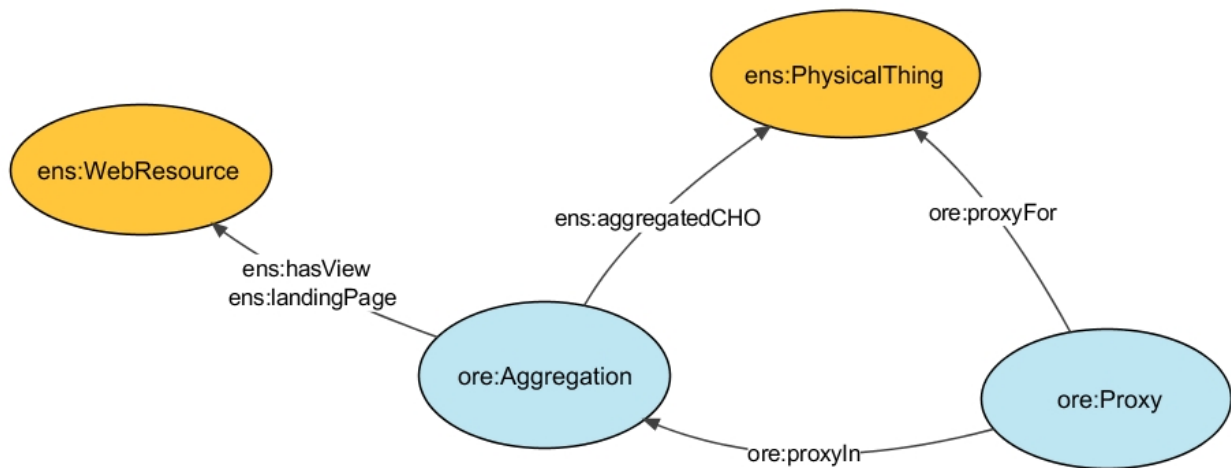


Abbildung 1.4 - Vier Klassen in der Datenmodellierung mit dem EDM

Eine wichtige Anforderung an das EDM besteht in der Abbildung von Hierarchien. Das gilt sowohl hinsichtlich der Modellierung von Kompositobjekten, etwa mehrbändigen Werken oder Kapiteln eines Werks, als auch der Abbildung von Derivation und Versionierung, wie im Falle künstlerischer Bearbeitungen oder verschiedener Auflagen eines Werks. Das EDM realisiert diese Anforderung mit der Nachnutzung und Einführung der ‚Properties‘ `dcterms:hasPart` und `dcterms:isPartOf` zwischen den repräsentierten Objekten, `ore:aggregates` zwischen den `ore:Aggregations` und `ens:isNextInSequence`, um die korrekte Reihenfolge abzubilden.

Mit seiner offenen Struktur unterstützt das EDM das Verlinken von Daten.³⁵ Dabei folgt es in seiner Architektur Berners-Lees Empfehlungen zu Linked Data (vgl. Kapitel 1.2) in allen Punkten. Die technischen Voraussetzungen für eine Integration des Europeana Informationsraums in die Linked Open Data Architektur sind mit dem EDM geschaffen und erste Schritte für die Umsetzung der Integration bereits unternommen:

Europeana's rich semantic data is also added to this Linked Open Data cloud. We use the Europeana Data Model to describe the rich resources from museums, libraries, archives and audiovisual archives, and then link this data to existing datasets such as DBpedia (a data version of Wikipedia) or Geonames (structured geographical data). Currently, the complete Amsterdam Museum's collection is available as Europeana Linked Open Data. As a result, there are now more than 5 Million RDF triplets (or "facts") describing over 70,000 cultural heritage objects related to the city of Amsterdam. Links are provided to the Dutch Art and Architecture Thesaurus (AATNed), Getty's Union List of Artists Names (ULAN), Geonames and DBpedia, enriching the Amsterdam dataset. (Europeana Thoughtlab)

³⁵ Siehe dazu auch Isaac 2010A: „(...) [T]he Linked Data approach emphasizes the re-use and linkage of richly described resources over the web.“

2. VD 17 – Die Metadaten im METS/MODS Format

Die etwa 2500 Datensätze der Staatsbibliothek zu Berlin, die im Rahmen des Prototyping auf das Europeana Data Model gemappt werden, beschreiben einen Teil der Digitalisierten Sammlungen aus dem Projekt VD 17. Intern wird in der Staatsbibliothek für diese Sammlung die Bezeichnung P 17 (Preußen 17 digital) verwendet, die für Drucke aus dem 17. Jahrhundert mit Druckorten im Gebiet des damaligen Preußen steht. Sie bildet einen Teil der von der Deutschen Forschungsgemeinschaft geförderten retrospektiven Nationalbibliographie VD 17, die im deutschen Sprachraum erschienene Drucke des 17. Jahrhunderts verzeichnet.

Die VD17-Datenbank, eine PICA-Datenbank, die von der Verbundzentrale des GBV gehostet wird, enthält mehr als 270.000 Titel mit rund 640.000 Exemplar-Nachweisen. Daneben sind derzeit über 29.000 Volldigitalisate nachgewiesen.

Im Rahmen der Aktionslinie "Kulturelle Überlieferung" stellt die DFG Fördergelder für die Digitalisierung der in nationalen Verzeichnissen nachgewiesenen Drucke des VD 16 und VD 17 zur Verfügung. In einem kooperativen Verfahren sollen die bereits vorhandenen Erschließungsdaten des VD 17 durch den digitalen Zugriff gezielt ergänzt werden. Auf der Grundlage eines gemeinsamen Masterplans der Träger- und Partnerbibliotheken wird angestrebt, bis zum Jahr 2020 für etwa 90 Prozent der nachgewiesenen Titel Volldigitalisate bereitzustellen und den Zugriff darauf über die VD17-Datenbank zu ermöglichen. (VD 17)

Die Förderrichtlinien für die Fördergelder schreiben unter anderem vor, die digitalisierten Drucke über den DFG-Viewer zugänglich zu machen.³⁶

Das Ziel dieser Vorgabe besteht darin, den Nutzern die Digitalisate auf einer einheitlichen Oberfläche zu präsentieren. Die Nutzung des METS/MODS-Formats für die Meta- und Strukturdaten der Digitalisate ist, neben der Bereitstellung verschiedener Auflösungen und Nutzung bestimmter Speicherformate, Teil der vorgeschriebenen Förderrichtlinien. Diesen Richtlinien folgend, liegen die auf das EDM abzubildenden Daten der Staatsbibliothek im METS/MODS-Format vor.

Im Folgenden werden die Formate METS und MODS exemplarisch anhand eines Datensatzes aus den Digitalisierten Sammlungen der Staatsbibliothek erläutert. Der Datensatz aus dem im VD 17 nachgewiesenen Bestand preußischer Drucke der Staatsbibliothek zu Berlin beschreibt eine Leichenpredigt des Erfurter Diakon und Pastoren Jeremias Alberti von 1656 mit dem Titel „Christiani Vita Et Corona“.³⁷ Der Grund für die Wahl dieses Beispiels liegt zum einen in dessen

³⁶ „Der DFG-Viewer ist ein Browser-Webdienst zur Anzeige von Digitalisaten aus dezentralen Bibliotheksrepositorien. Er verfügt über eine XML-Schnittstelle zum Austausch von Meta- und Strukturdaten im METS/MODS-Format.“ (DFG-Viewer)

³⁷ Der Datensatz (im Verlauf der Arbeit in Teilen abgebildet) ist vollständig unter <http://digital.staatsbibliothek-berlin.de/dms/mets-resolver/?PPN=PPN651724848> einzusehen.

Erschließungstiefe – es sind neben dem Titel des Gesamtwerks auch die Titel der Epicedien³⁸ in den Metadaten aufgeführt -, zum anderen darin, dass es sich hierbei um eine Leichenpredigt³⁹ handelt. Die Leichenpredigten bilden ein geschlossenes Corpus in den Digitalen Sammlungen der Staatsbibliothek und sind zu einem überwiegendem Teil Bestandteil von VD 17. Eine Besonderheit dieses Corpus liegt darin, dass neben dem eigentlichen Verfasser der Predigt und dem Gefeierten (Verstorbenen), sämtliche sogenannte Beiträger, das heißt die Verfasser der Leichabdankung, Epicedien, Programma Academicum und gegebenenfalls musikalischer Beigaben in den Metadaten erfasst, und teilweise mit Normdaten verlinkt werden. Zwar sind der Gefeierte und die Beiträger nur in den PICA- und nicht in den METS/MODS-Datensätzen enthalten, sie können aber durch das Mapping der vom GBV bereitgestellten PICA-Daten auf das EDM mit den gemappten METS/MODS-Daten verlinkt und in die Suche integriert werden. Anhand derart reicher Metadaten lässt sich das Potential des Europeana Data Model besonders gut veranschaulichen. Dazu ausführlicher in Kapitel 3.

2.1 METS - Metadata Encoding and Transmission Standard

METS steht für Metadata Encoding and Transmission Standard und ist ein, derzeit in der Version 1.9 vorliegendes, von der Library of Congress verwaltetes XML-Format zur Beschreibung von digitalen Sammlungen (METS). Es handelt sich dabei um ein Containerformat, das Metadaten unterschiedlichen Formats wie MODS, MAB (Maschinelles Austauschformat für Bibliotheken), MARC (Machine-Readable Catalogue) oder Dublin Core, aufnehmen kann. METS enthält Elemente, um komplexe Objekte zu gruppieren und sie mit deskriptiven und administrativen Metadaten zu verbinden. Damit lassen sich hierarchisch strukturierte Werke wie Bücher mit mehreren Kapiteln oder Buchreihen kodieren und beschreiben.

Ein METS-Dokument besteht aus sieben Hauptabschnitten. Im vorliegenden Beispiel wird mit dem zvdd/DFG-Viewer METS-Profil in der Version 2.0⁴⁰ ein bestimmtes METS-Profil

³⁸ Epicedien sind Trauergedichte von Verwandten oder Bekannten des Verstorbenen und ein fester Bestandteil von Leichenpredigten.

³⁹ Die Gattung der Leichenpredigt entwickelte sich in der Reformationszeit und folgt einem festen Aufbau: dem Titelblatt, das eine Widmung und eventuell bildliche Beigaben enthalten kann (etwa die Ahnentafel des Verstorbenen), folgt das Exordium, der Leichtext mit Textausschmückungen. Daran schließen die Abhandlung, die Auslegung der Bibelstelle, und die Personalia, die Schilderung des Lebenslaufs des Verstorbenen, an. Die darauf folgende Leichabdankung wurde meistens von einem Freund der Familie verfasst und nach der Rückkehr vom Grab verlesen. Die Epicedien bilden häufig den Schluss der gedruckten Leichenpredigt, es sei denn, der Verstorbene gehörte einer Universität an. In dem Fall schließen an die Epicedien Programma Academicum – ein vom Rektor der Universität verfasster, kurzer Bericht über den wissenschaftlichen Werdegang des Verstorbenen – und akademische Trauerrede an. Einige Leichenpredigten enthalten zudem Musikbeigaben: dabei handelte es sich meist um Auftragskompositionen, die später nicht selten Eingang in evangelische Gesangbücher gefunden haben.

⁴⁰ Siehe dazu zvdd/DFG-Viewer METS-Profil 2009, S.1: „Dieses METS Profil beschreibt das Datenformat für den DFG-Viewer und definiert darüber hinausgehende Erweiterungen für das zvdd-Portal. Dokumente, die diesem Profil entsprechen, können sowohl durch den DFG-Viewer angezeigt als auch durch das zvdd-Portal verarbeitet und indexiert werden.“

angewendet, das in einigen Punkten leicht von dem METS Schema der Library of Congress abweicht. Es definiert nur fünf statt der sieben Hauptabschnitte, indem es den ‚METS-Header‘ und die ‚Behaviour Section‘ unberücksichtigt lässt. Im Folgenden sollen dennoch alle sieben Abschnitte kurz erläutert und eventuelle Besonderheiten des hier verwendeten METS-Profiles gegenüber dem METS-Schema erwähnt werden.

1. Der Kopfteil oder ‚**METS-Header**‘ enthält Metadaten, die das METS-Dokument beschreiben, sowie Angaben zum Bearbeiter oder Herausgeber des Dokuments. Der ‚METS-Header‘ ist in vorliegendem Beispiel nicht vorhanden.
2. Die Erschließungsangaben oder ‚**Descriptive Metadata Section**‘ (mets:dmdSec) können entweder durch den Verweis auf ein externes Dokument oder, wie im Fall des vorliegenden Beispiels, als in das METS-Dokument eingebettete Angaben – hier im MODS-Format⁴¹ - enthalten sein. Das eingebettete Format wird mit dem XML-Tag und eingefügtem Attribut <mets:mdWrap MDTYPE="MODS"> angezeigt.

```
<mets:dmdSec ID="DMDLOG_0000">
<mets:mdWrap MDTYPE="MODS">
<mets:xmlData>
<mods:mods>
<mods:location>
  <mods:shelfLocator>Ee 700-1086</mods:shelfLocator>
</mods:location>
<mods:originInfo>
  <mods:place>
    <mods:placeTerm type="text">Erfurti</mods:placeTerm>
  </mods:place>
  <mods:dateIssued encoding="w3cdtf" keyDate="yes">1656</mods:dateIssued>
  <mods:publisher>Dedekindus</mods:publisher>
</mods:originInfo>
<mods:originInfo>
  <mods:place>
    <mods:placeTerm type="text">Berlin</mods:placeTerm>
  </mods:place>
  <mods:dateCaptured encoding="w3cdtf">2011</mods:dateCaptured>
  <mods:publisher>
    Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, Germany
  </mods:publisher>
  <mods:edition>[Electronic ed.]</mods:edition>
</mods:originInfo>
<mods:classification authority="ZVDD">Historische Drucke</mods:classification>
<mods:classification authority="ZVDD">Theologie</mods:classification>
<mods:recordInfo>
  <mods:recordIdentifier source="gbv-ppn">PPN651724848</mods:recordIdentifier>
</mods:recordInfo>
<mods:identifier type="purl">
  http://resolver.staatsbibliothek-berlin.de/SBB000045E300000000
</mods:identifier>
<mods:identifier type="vd17">39:103909Q</mods:identifier>
```

⁴¹ „Sowohl der DFG-Viewer als auch das zvdd-Portal unterstützen lediglich deskriptive Metadatenabschnitte vom Typ MODS. Diese müssen in das METS Dokument eingebunden sein und sich innerhalb von <mdWrap> befinden.“ (zvdd/DFG-Viewer METS-Profil 2009, S.2)


```

<mods:identifizier type="PPNanalog">PPN091865476</mods:identifizier>
<mods:titleInfo>
  <mods:title>Christiani Vita Et Corona</mods:title>
  <mods:subTitle>
    Aus den Trostreichen Worten des heiligen Apostels Pauli in der II. an Timoth. am IV. v. 7. &
    8. Ich habe einen guten Kampf gekämpft/ [et]c. Bey der ... LeichBestattung Des ... Joachimi
    Gerstenbergers/ wohlverdienten ältesten Obristen Vierherrns in unserer Christlichen Polickey
    Welcher den 18. Ianuarii ... dieses 1656. Jahrs ... entschlaffen/ und hernach den 24. eiusdem ...
    in der PredigerKirchen beygesetzt worden
  </mods:subTitle>
</mods:titleInfo>
<mods:genre authority="aad" type="class">Leichenpredigt</mods:genre>
<mods:language>
  <mods:languageTerm authority="iso639-2b" type="code">de</mods:languageTerm>
</mods:language>
<mods:subject authority="gdz">
  <mods:topic>albechri</mods:topic>
</mods:subject>
<mods:relatedItem type="series">
  <mods:titleInfo>
    <mods:title>
      Preußen 17 digital - Digitalisierung des im VD 17 nachgewiesenen Bestandes
      preußischer Drucke der Staatsbibliothek zu Berlin
    </mods:title>
  </mods:titleInfo>
</mods:relatedItem>
<mods:name type="personal">
  <mods:role>
    <mods:roleTerm authority="marcrelator" type="code">aut</mods:roleTerm>
  </mods:role>
  <mods:namePart type="family">Alberti</mods:namePart>
  <mods:namePart type="given">Jeremias</mods:namePart>
  <mods:displayForm>Alberti, Jeremias</mods:displayForm>
</mods:name>
<mods:physicalDescription>
  <mods:digitalOrigin>reformatted digital</mods:digitalOrigin>
  <mods:extent>[36] Bl</mods:extent>
</mods:physicalDescription>
<mods:extension>
  <zvdd:zvddWrap>
    <zvdd:titleWord>
      Krone wahren Leichenbestattung Polizei Tränen Hintritt
    </zvdd:titleWord>
  </zvdd:zvddWrap>
</mods:extension>
</mods:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>

<mets:dmdSec ID="DMDLOG_0001">
<mets:mdWrap MDTYPE="MODS">
<mets:xmlData>
<mods:mods>
  <mods:identifizier type="PPNanalog">09176842X</mods:identifizier>
  <mods:titleInfo>
    <mods:title>
      Heisse Thränen über den Traurigen doch Seeligen Hintrit Deß Edlen ... Herrn
      Joachims Gerstenbergers Hiesiger Stadt ältesten und wohlverdienten Obristen
      Vier-Herrns
    </mods:title>
  </mods:titleInfo>

```

```

        <mods:language>
            <mods:languageTermauthority="iso6392b" type="code">de</mods:languageTerm>
        </mods:language>
    </mods:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>

```

Hier soll nur ein Ausschnitt der Descriptive Metadata Section gezeigt werden, der die Abschnitte „DMDLOG_0000“ und „DMDLOG_0001“ darstellt. Mit der ID „DMDLOG_XXXX“ wird auf Abschnitte mit deskriptiven Metadaten zu einer logischen Einheit verwiesen. Die Ebene _0000 steht für das gesamte Werk, _0001 für das erste Kapitel (hier die Epicedien), _0002 (hier nicht mehr abgebildet) für das zweite Kapitel (hier die Epicedien).⁴² Auf die deskriptiven Metadaten im MODS-Format wird unter 2.2 genauer eingegangen.

3. Die Verwaltungsangaben oder ‚**Administrative Metadata Section**‘ (mets:amdSec) können Informationen zur Herstellung und Speicherung der Dateien, zum Urheberrecht und zum Verhältnis zwischen Master und Derivaten enthalten. In dem im vorliegenden Beispiel verwendeten zvdd/DFG-Viewer METS-Profil kommt an dieser Stelle ein eigenes Schema zur Anwendung, um die Informationen über den Urheber im <mets:rightsMD> Element zu speichern:

Innerhalb des <mdWrap> Elements kommt ein eigenes Rechte-Schema zur Anwendung, welches seinen Inhalt in ein umschließendes <dv:rights> Element schreibt. Innerhalb dieses Elements gibt es folgende drei Unterelemente, die jeweils nur genau einmal vorkommen dürfen und müssen.

owner — Der Urheber des Digitalisats

logo — Eine URL des Logos des Urhebers; dieses Logo wird durch den DFG-Viewer sowie das zvdd-Portal entsprechend angezeigt.

homepage — Eine URL der Homepage des Urhebers.

(...)

Innerhalb des <mdWrap> Elements kommt ein eigenes Herkunfts-Schema zur Anwendung, welches seinen Inhalt in ein umschließendes <dv:links> Element schreibt. Innerhalb dieses Elements gibt es folgende zwei Unterelemente, die jeweils nur genau einmal vorkommen dürfen und müssen.

reference — Eine URL auf den Katalogeintrag des Digitalisats

presentation — Eine URL auf die Online-Präsentation des Digitalisats. (zvdd/DFG-Viewer METS-Profil 2009, S. 7ff)

⁴² Die analoge PPN (Pica Production Number) ist in diesem Beispiel nicht nur einmalig für das ganze Werk vergeben, sondern auch für die in „DMDLOG_0001“ und „DMDLOG_0002“ beschriebenen Epicedien.

```

<mets:amdSec ID="AMD">
<mets:rightsMD ID="RIGHTS">
<mets:mdWrap MDTYPE="OTHER" MIMETYPE="text/xml" OTHERMDTYPE="DVRIGHTS">
<mets:xmlData>
  <dv:rights>
    <dv:owner>
      Staatsbibliothek zu Berlin - Preußischer Kulturbesitz
    </dv:owner>
    <dv:ownerLogo>
      http://resolver.staatsbibliothek-berlin.de/SBB0000000100000000
    </dv:ownerLogo>
    <dv:ownerSiteURL>http://www.staatsbibliothek-berlin.de</dv:ownerSiteURL>
    <dv:ownerContact>mailto:info@sbb.spk-berlin.de</dv:ownerContact>
  </dv:rights>
</mets:xmlData>
</mets:mdWrap>
</mets:rightsMD>
<mets:digiprovMD ID="DIGIPROV">
<mets:mdWrap MDTYPE="OTHER" MIMETYPE="text/xml" OTHERMDTYPE="DVLINKS">
<mets:xmlData>
  <dv:links>
    <dv:reference>http://www.stabikat.de/DB=1/PPN?PPN=651724848 </dv:reference>
    <dv:presentation>
      http://digital.staatsbibliothek-berlin.de/dms/werkansicht/?PPN=PPN651724848
    </dv:presentation>
  </dv:links>
</mets:xmlData>
</mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>

```

4. Im Dateienabschnitt oder der ‚**File Section**‘ (mets:fileSec) findet sich eine Auflistung aller Dateien, aus denen das digitale Objekt besteht. Zusammengehörige Dateien können mit <file Grp> (File Group) zusammengefasst werden. Das <mets:file> Element ist ein Kindelement des <mets:fileGrp> Elements. Der Inhalt der Dateien (Bytestream), der außerhalb des METS-Dokuments gespeichert wird, ist persistent mit der METS-Datei durch das <FLocat> Element, das mit dem xlink:href Attribut mit einem URL referenziert wird, verknüpft. Der DFG-Viewer kann die Datei bei Bedarf aus dem ursprünglichen Repository laden.

Das Attribut „ID“ in jedem <file> Element dient als eindeutiger Verweis innerhalb des METS-Dokuments. Diese IDs werden von den <fptr> (filepointer) Elementen in dem Abschnitt <structMap> (s.u.) referenziert und weisen den Strukturelementen die entsprechenden <file> Elemente zu.⁴³ Auch an dieser Stelle kann nur ein kleiner Ausschnitt der ‚File Section‘ gezeigt werden. Sie nimmt den größten Teil des METS-Dokuments ein, da jede Dateigruppe ein komplettes Set an Imagedaten enthalten muss. Die ‚File Section‘ enthält im vorliegenden Beispiel sechs Dateigruppen (MIN, PRESENTATION, MAX, THUMBS, DEFAULT, LOCAL), die Derivate der Masterdateien

⁴³ Siehe dazu zvdd/DFG-Viewer METS-Profil 2009, S. 10

zur Anzeige in verschiedenen Größen enthalten. Dabei sind die Auflösungen MIN und DEFAULT für den DFG-Viewer vorgeschrieben.

```
<mets:fileSec>
  <mets:fileGrp USE="MIN">
    <mets:file ID="FILE_0000_MIN" MIMETYPE="image/jpg">
      <mets:FLocatLOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN651724848/500/0/00000001.jpg"/>
    </mets:file>
    <mets:file ID="FILE_0001_MIN" MIMETYPE="image/jpg">
      <mets:FLocatLOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN651724848/500/0/00000002.jpg"/>
    </mets:file>
    <mets:file ID="FILE_0002_MIN" MIMETYPE="image/jpg">
      <mets:FLocatLOCTYPE="URL" xlink:href="http://content.staatsbibliothekberlin.de/dms/PPN651724848/500/0/00000003.jpg"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>
```

- Die Strukturbeschreibung oder ‚**Structural Map**‘ (mets:structMap) bildet den inneren Aufbau des digitalen Objekts ab, indem sie Strukturelemente mit den Dateien, aus denen das digitale Objekt besteht, und den dazugehörigen Metadaten verbindet. Sie besteht aus zwei <StructMap> Elementen, deren TYPE Attribute die Werte „LOGICAL“ und „PHYSICAL“ haben und die einmal die logische und einmal die physische Struktur des Werks abbilden.

Die logische Einheit „LOG_0000“ wird der ‚Descriptive Metadata‘ ID „DMDLOG_0000“ mit dem Haupttitel der Leichenpredigt – „Christiani Vita Et Corona“ – zugewiesen. Es folgen die logischen Einheiten „LOG_0001“ bis „LOG_0008“, die den Strukturelementen Einband, Vorderdeckel, Spiegel, Vorsatz, Titelblatt, Predigt, Personalialia und Leichabdanke zugeordnet werden, usw.

```
<mets:structMap TYPE="LOGICAL">
  <mets:div ADMID="AMD" CONTENTIDS="http://resolver.staatsbibliothekberlin.de/SBB000045E30000000" DMDID="DMDLOG_0000" ID="LOG_0000" LABEL="Christiani Vita Et Corona" TYPE="monograph">
    <mets:div ID="LOG_0001" TYPE="binding">
      <mets:div ID="LOG_0002" TYPE="cover_front"/>
      <mets:div ID="LOG_0003" TYPE="paste_down"/>
      <mets:div ID="LOG_0004" TYPE="endsheet"/>
    </mets:div>
    <mets:div ID="LOG_0005" TYPE="title_page"/>
    <mets:div ID="LOG_0006" TYPE="sermon"/>
    <mets:div ID="LOG_0007" TYPE="personalialia"/>
    <mets:div ID="LOG_0008" TYPE="funeral_oration"/>
    <mets:div DMDID="DMDLOG_0001" ID="LOG_0009" LABEL="Heisse Thränen über den Traurigen doch Seeligen Hintritt Deß Edlen ... Herrn Joachims Gerstenbergers Hiesiger Stadt ältesten und wohlverdienten Obristen Vier-Herrns" TYPE="epicedia"/>
    <mets:div DMDID="DMDLOG_0002" ID="LOG_0010" LABEL="Iusta Funebria Beatis Manibus ... Joachimi Gerstenbergeri, Reipublicae Erffurtinae Supremi ac Senioris Quatuor-Viri ... a Quibusdam amoris, honoris & observantiae ergo soluta" TYPE="epicedia"/>
    <mets:div ID="LOG_0011" TYPE="colophon"/>
    <mets:div ID="LOG_0012" TYPE="binding">
  </mets:div>
```

```

    <mets:div ID="LOG_0013" TYPE="endsheet"/>
    <mets:div ID="LOG_0014" TYPE="paste_down"/>
    <mets:div ID="LOG_0015" TYPE="cover_back"/>
  </mets:div>
<mets:div ID="LOG_0016" TYPE="colour_checker"/>
</mets:div>
</mets:structMap>

```

Die physische ‚Structure Map‘ ordnet den einzelnen Seiten des Werks die dazugehörigen Images in den verschiedenen Auflösungen zu. Mit den Attributen ‚ORDER‘ und ‚PAGE‘ erfolgt die Paginierung gemäß der analogen Vorlage.

```

<mets:structMap TYPE="PHYSICAL">
  <mets:div CONTENTIDS="NULL" DMDID="DMDPHYS_0000" ID="PHYS_0000"
  TYPE="physSequence">
    <mets:div CONTENTIDS="NULL" ID="PHYS_0001" ORDER="1" TYPE="page">
      <mets:fptr FILEID="FILE_0000_MIN"/>
      <mets:fptr FILEID="FILE_0000_PRESENTATION"/>
      <mets:fptr FILEID="FILE_0000_MAX"/>
      <mets:fptr FILEID="FILE_0000_THUMBS"/>
      <mets:fptr FILEID="FILE_0000_DEFAULT"/>
      <mets:fptr FILEID="FILE_0000"/>
    </mets:div>
    <mets:div CONTENTIDS="NULL" ID="PHYS_0002" ORDER="2" TYPE="page">
      <mets:fptr FILEID="FILE_0001_MIN"/>
      <mets:fptr FILEID="FILE_0001_PRESENTATION"/>
      <mets:fptr FILEID="FILE_0001_MAX"/>
      <mets:fptr FILEID="FILE_0001_THUMBS"/>
      <mets:fptr FILEID="FILE_0001_DEFAULT"/>
      <mets:fptr FILEID="FILE_0001"/>
    </mets:div>
  </mets:div>

```

- Die Strukturverknüpfungen oder ‚**Structural Links**‘ (mets:structLink) speichern die Verknüpfungen zwischen logischer und physischer Struktur. Jede Verknüpfung erfolgt über ein eigenes <smLink> Element. Die xlink:from and xlink:to Attribute beinhalten die Werte der ID Attribute der jeweiligen Elemente aus der logischen und physischen Struktur. Aus unten aufgeführtem Ausschnitt der ‚Structural Links‘ wäre so zum Beispiel abzulesen, dass das Titelblatt („LOG_0005“) die Seiten 5 und 6 („PHYS_0005“ und „PHYS_0006“) umfasst – gezählt werden hier Vor- und Rückseite. Über die Strukturverknüpfungen wird eine Navigation in der Struktur des Werks ermöglicht.

```

<mets:structLink>
  <mets:smLink xlink:to="PHYS_0001" xlink:from="LOG_0000"/>
  <mets:smLink xlink:to="PHYS_0002" xlink:from="LOG_0000"/>
  <mets:smLink xlink:to="PHYS_0003" xlink:from="LOG_0000"/>
  <mets:smLink xlink:to="PHYS_0004" xlink:from="LOG_0000"/>
  <mets:smLink xlink:to="PHYS_0005" xlink:from="LOG_0000"/>
  (...)
  <mets:smLink xlink:to="PHYS_0001" xlink:from="LOG_0001"/>
  <mets:smLink xlink:to="PHYS_0002" xlink:from="LOG_0001"/>
  <mets:smLink xlink:to="PHYS_0003" xlink:from="LOG_0001"/>

```

```

<mets:smLink xlink:to="PHYS_0004" xlink:from="LOG_0001"/>
<mets:smLink xlink:to="PHYS_0001" xlink:from="LOG_0002"/>
<mets:smLink xlink:to="PHYS_0002" xlink:from="LOG_0003"/>
<mets:smLink xlink:to="PHYS_0003" xlink:from="LOG_0004"/>
<mets:smLink xlink:to="PHYS_0004" xlink:from="LOG_0004"/>
<mets:smLink xlink:to="PHYS_0005" xlink:from="LOG_0005"/>
<mets:smLink xlink:to="PHYS_0006" xlink:from="LOG_0005"/>

```

7. Die ‚**Behaviour Section**‘ verknüpft ausführbare Anweisungen für das Verhalten mit den Inhalten in METS-Objekten. Da sie in dem hier behandelten Beispiel nicht vorliegt, soll darauf an dieser Stelle nicht näher eingegangen werden.⁴⁴

2.2 MODS - Metadata Object Description Schema

Zum Abbilden der deskriptiven Metadaten wird im vorliegenden Beispiel das vom DFG-Viewer vorgeschriebene MODS-Format verwendet. Die bibliographischen Metadaten werden, wie im vorigen Abschnitt beschrieben, in die ‚Descriptive Metadata Section‘ der METS-Datei eingebettet. MODS steht für Metadata Object Description Schema und ist ein seit 2010 in der Version 3.4 vorliegendes, von der Library of Congress entwickeltes XML-Format für bibliographische Metadaten (MODS). Wie auch im Fall des METS-Profiles wird für die Beschreibung des hier behandelten Beispiels ein eigens definiertes MODS-Profil, das zvvdd MODS Anwendungsprofil in der Version 1.0, verwendet, das auf MODS 3.2 der Library of Congress von 2006 basiert (Otte et al. 2008). Das Anwendungsprofil beschreibt die Anforderungen an die einzelnen MODS-Elemente und weist ihnen einen Status (erforderlich, optional, unerwünscht etc.) zu. Darüber hinaus beinhaltet es Vorschläge für Mappings von unterschiedlichen Metadaten-Formaten auf das zvvdd MODS-Format.

Die bibliographischen Metadaten für die Digitalisate der Staatsbibliothek liegen zu Beginn der Digitalisierung im - von allen Verbundteilnehmern des GBV genutzten - PICA-Format vor. In einem ersten Schritt des Digitalisierungsprozesses wird in der Staatsbibliothek der PICA-Datensatz des zur Digitalisierung bestimmten analogen Werks kopiert und mit einer eigenen PPN (PPN digital) versehen. Im Verlauf der Digitalisierung erfolgt automatisch mit der in der Staatsbibliothek angewendeten Goobi-Software für die Workflow-Steuerung die Erstellung des METS-Datensatzes und das Mapping der in PICA vorliegenden Daten auf das MODS-Format. Zu den erforderlichen Elementen bzw. erforderlichen falls vorhandenen Elementen gehören in der bibliographischen Beschreibung laut zvvdd MODS Anwendungsprofil <titleInfo> mit den Subelementen <title> und <subtitle>, <name type="personal" authority="..."> mit den Subelementen <role> und <roleTerm type="code" authority="marcrelator">, <originInfo> mit den Subelementen <place>, <placeTerm type="text">, <publisher>, <dateIssued keyDate="yes"

⁴⁴ Für eine ausführliche Darstellung siehe <http://www.loc.gov/standards/mets/METSOverview.v2.html>

encoding="w3cdtf"> und <edition>. ⁴⁵ Am Beispiel der bibliographischen Angaben zum Verfasser soll kurz das Mapping der Daten vom PICA- auf das MODS-Format erläutert werden:

GBV-Pica+	Pica-Inhalt	zvdd-MODS
028A \$B 028B \$B	Verfasser (Funktionsbezeichnung)	<pre><name type="personal"><role> <roleTerm type="code" authority="marcrelator "></pre>
028C \$B	Sonstige beteiligte Person (Funktionsbezeichnung)	
028F	Gefeierte Person (Funktionsbezeichnung)	
028G \$B	Sonstige nichtbeteiligte Person (Funktionsbezeichnung)	
028L/00	Widmungsempfänger	
028L/01	Zensor	
028L/02	Lit., künstl., musikal. Beiträger	
028L/03 \$B	Sonstige nichtbeteiligte bzw. im Sachtitel genannte Personen (Funktionsbezeichnung)	
028M \$B	Konkurrenz-Verfasser (MBW) (Funktionsbezeichnung)	

Tabelle 1 - Mapping Pica – ZVDD-MODS (aus: Otte et al. 2008 S. 15)

Der Name des Verfassers oder anderer Beiträger befindet sich in den PICA-Daten in den Feldern 028(...). Für die Abbildung im MODS-Format dienen das Element <name> und das Attribut „type“ mit dem Wert „personal“. Die PICA-Funktionsbezeichnungen werden mit den Subelementen <role> und <roleTerm> und den Attributen „authority“ mit dem Wert „marcrelator“ und „type“ mit dem Wert „code“ in Marcrelator-Codes umgewandelt:

```
<mods:name type="personal">
  <mods:role>
    <mods:roleTerm authority="marcrelator" type="code">aut</mods:roleTerm>
  </mods:role>
  <mods:namePart type="family">Alberti</mods:namePart>
  <mods:namePart type="given">Jeremias</mods:namePart>
  <mods:displayForm>Alberti, Jeremias</mods:displayForm>
</mods:name>
```

In vorliegendem Beispiel wird das Element <name> noch durch die Subelemente <namePart> und <displayForm> detailliert.

Hier fällt auf, dass andere in den PICA-Daten erfasste Beiträger der Leichenpredigt in MODS nicht erfasst sind, obwohl es das zvdd MODS Anwendungsprofil ermöglicht. So könnte beispielsweise die verstorbene Person mit dem Marcrelatorcode „hnr“ als gefeierte Person verzeichnet und anschließend als solche gesucht werden:

⁴⁵ Für eine vollständige Auflistung aller obligatorischen bzw. obligatorischen falls vorhandenen Elemente siehe Otte et al. 2008 S.4f

```

<mods:name type="personal">
  <mods:role>
    <mods:roleTerm authority="marcrelator" type="code">hnr</mods:roleTerm>
  </mods:role>
  <mods:namePart type="family">Gerstenberger</mods:namePart>
  <mods:namePart type="given">Joachim</mods:namePart>
</mods:name>

```

Die Probleme, die sich durch die Reduktion der Daten im Verlauf des Mappings vom PICA- auf das MODS-Format für die Suche ergeben, werden in den folgenden Kapiteln zur Sprache kommen.

2.3 Mögliche Schwierigkeiten für die Datenmodellierung

Bevor im nächsten Kapitel die Modellierung der Daten mit dem Europeana Data Model erfolgt, soll der exemplarisch zu mappende Datensatz auf potentielle Probleme hin untersucht werden, die sich mit dem Mapping auf das EDM ergeben können.

Eine grundsätzliche Schwierigkeit besteht darin, dass die Metadaten im MODS-Format stellenweise sowohl das analoge, als auch das digitalisierte Werk beschreiben. Das Element <Shelflocator> etwa bezeichnet die Signatur des analogen Werks:

```

<mods:location>
  <mods:shelfLocator>Ee 700-1086</mods:shelfLocator>
</mods:location>

```

Parallel sind als <Identifizier> die PPN für den Metadatensatz zum analogen und zum digitalen Werk aufgeführt. Andere Metadaten, wie Autor oder Titel, gelten gleichermaßen für das analoge, wie auch das digitale Werk.

Das Element <OriginInfo> ist zwei Mal vorhanden und beschreibt in einem Fall das analoge, im anderen das digitale Werk, das von dem analogen mit dem Subelement <mods:edition>[Electronic ed.]</mods:edition> unterschieden wird:

```

<mods:originInfo>
  <mods:place>
    <mods:placeTerm type="text">Erfurt</mods:placeTerm>
  </mods:place>
  <mods:dateIssued encoding="w3cdtf" keyDate="yes">1656</mods:dateIssued>
  <mods:publisher>Dedekindus</mods:publisher>
</mods:originInfo>
<mods:originInfo>
  <mods:place>
    <mods:placeTerm type="text">Berlin</mods:placeTerm>
  </mods:place>
  <mods:dateCaptured encoding="w3cdtf">2011</mods:dateCaptured>
  <mods:publisher>
    Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, Germany
  </mods:publisher>
  <mods:edition>[Electronic ed.]</mods:edition>
</mods:originInfo>

```


Allgemein sei auf die Schwierigkeit hingewiesen, die Metadaten, das analoge und das digitale Werk betreffend, klar zu trennen: dem Digitalisat sind – da es nicht digital „geboren“ wurde – immer zwei Ursprünge inhärent: zum einen handelt es sich bei dem vorliegenden Beispiel um den Druck aus dem Jahr 1656, gedruckt von Dedekindus in Erfurt, zum anderen um ein Digitalisat, das 2011 von der Staatsbibliothek in Berlin veröffentlicht wurde. Das Europeana Data Model hält die beiden Perspektiven auf das Werk auseinander. Wie diese Unterscheidung realisiert wird, soll im folgenden Kapitel näher erläutert werden. Eine wichtige Anforderung an die Inhaltsanbieter, die Metadaten an Europeana liefern, besteht darin, die Metadaten des Digitalisats und des analogen Exemplars betreffend zu trennen:

One of the most important pieces of information expected from providers is the distinction between metadata that applies to the object itself, and the metadata that applies to the digital representations (...). (Isaac 2010A, S.17f)

Die Reduktion, die die bibliographischen Metadaten durch das Mapping vom PICA- auf das MODS-Format erfahren, stellt insofern eine Einbuße dar, als sämtliche in den PICA-Daten enthaltene Beiträge, die gefeierte Person und die Verlinkung mit den Normdaten⁴⁶ verloren gehen und so in den Digitalen Sammlungen der Staatsbibliothek weder nach Verfassern einzelner Teile der Leichenpredigt, noch nach dem Verstorbenen oder unterschiedlichen Schreibweisen der Namen, die die Normdaten bereithalten, gesucht werden kann.⁴⁷ Über die Modellierung der Daten mit dem Europeana Data Model soll die Verbindung zu den verlorenen Informationen wieder hergestellt werden.

Kleinere Unregelmäßigkeiten in den Metadaten betreffen etwa den Autor-Titel-Schlüssel, der mit den Elementen <subject> und <topic> kodiert wird:

```
<mods:subject authority="gdz">  
<mods:topic>albechri</mods:topic>  
</mods:subject>
```

Da die Verwendung beider Elemente in diesem Kontext dem internen Gebrauch dient und sich darüber hinaus schwer erschließt, bleiben sie bei dem Mapping unberücksichtigt.

⁴⁶ Die in vorliegendem Beispiel verwendete METS/MODS Version ist nicht in der Lage, die Verlinkung mit den Normdaten abzubilden. Die erst kürzlich veröffentlichte METS/MODS Version 3.4 ermöglicht die Abbildung der Verlinkung.

⁴⁷ Die Staatsbibliothek weiß um diese Probleme und arbeitet bereits seit längerem an einer Lösung.

3. Die Leichenpredigt „Christiani Vita Et Corona“: Modellierung mit dem Europeana Data Model

Nachdem in den vorangegangenen Kapiteln das Europeana Data Model, die auf dieses Modell abzubildenden Daten und die XML-basierten Austauschformate METS und MODS beschrieben wurden, soll nun eine graphenbasierte Modellierung der Daten mit dem Europeana Data Model erfolgen.

3.1 Die Objektrepräsentation mit dem Europeana Data Model

Zunächst wird die Objektrepräsentation entsprechend der obersten logischen Ebene des Datensatzes – ID="LOG_0000", die das gesamte Werk beschreibt, modelliert:

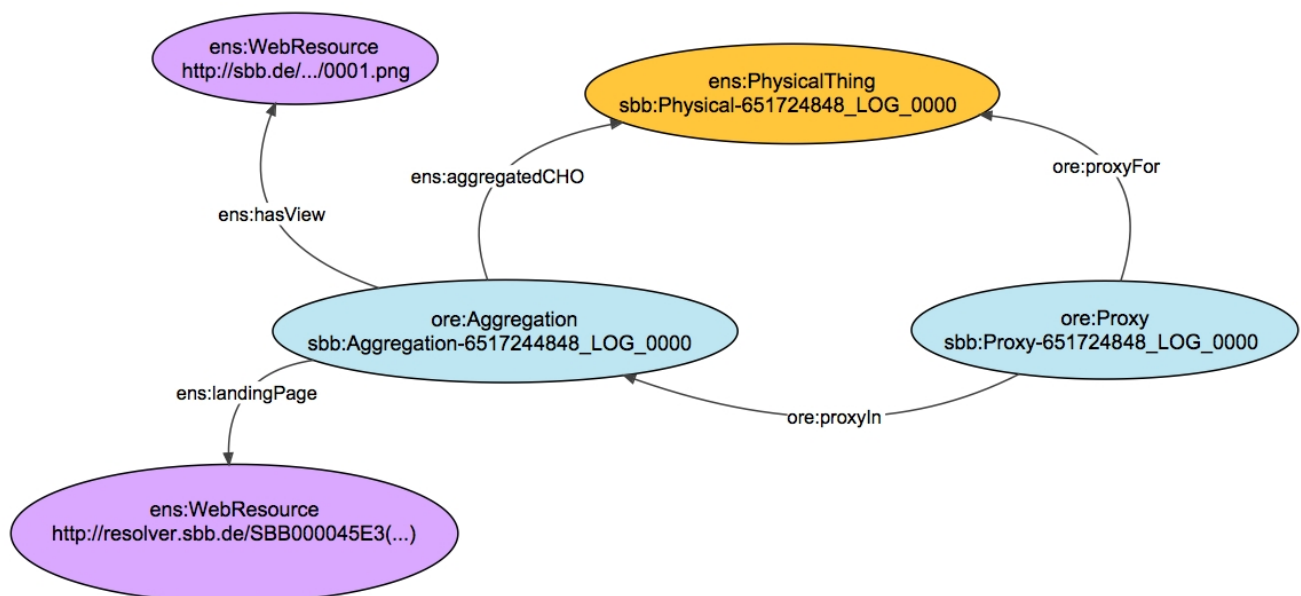


Abbildung 3.1 - Objektrepräsentation in EDM

Die auf dem Resource Description Framework basierende Modellierung mit den aus dem Namensraum OAI-ORE übernommenen Klassen **ore:Aggregation** und **ore:Proxy** und den mit dem Europeana Namespace eingeführten Klassen **ens:PhysicalThing** und **ens:WebResource** bildet den Kern einer jeden Objektrepräsentation im Informationsraum Europeana.

Die Unterscheidung in ein ‚PhysicalThing‘ und einen ‚Proxy‘⁴⁸ ermöglicht die Einnahme unterschiedlicher Perspektiven auf ein Objekt. Mit der Klasse **ens:WebResource** wird die

⁴⁸ Bei Vorträgen zum Europeana Data Model, beispielsweise in der Staatsbibliothek zu Berlin oder auf dem KIM-DINI-Kickoff-Workshop in Mannheim, zeigte sich wiederholt, dass die Bedeutung des ‚Proxy‘, wie er im Sinne des EDM genutzt wird, nicht weithin gebräuchlich ist. Die „Definition of the Europeana Data Model Elements“ beschreibt den Einsatz des Proxy wie folgt: „Europeana uses proxies as place-holders for Cultural Heritage Objects within Aggregations (...) to the end of making assertions about the corresponding Cultural Heritage Objects while

Differenzierung zwischen dem Objekt, auf dem die Objektrepräsentation basiert, und seinen digitalen Repräsentationen möglich. (Vgl. Kapitel 1.3 der vorliegenden Arbeit). Über die ore:Aggregation werden die verschiedenen Objektinformationen miteinander verbunden.⁴⁹

Um aus den Entitäten ‚Resources‘ gemäß den Vorgaben des RDF Schema zu machen, müssen sie über einen persistenten URI verfügen, der von Europeana erstellt wird. In obigem Beispiel (Abbildung 3.1) sind die URIs mit sbb:... angedeutet. sbb kennzeichnet eine ‚Resource‘ aus dem Namensraum der Staatsbibliothek zu Berlin. Die weiteren Teile der, in diesem Fall fiktiven, URIs sind vor allem unter dem Aspekt der Lesbarkeit und Verständlichkeit des Graphen gewählt: aggregation oder proxy, da es sich um die ore:Aggregation bzw. den ore:Proxy handelt. Die digitale PPN (651724848) wurde gewählt, da es sich dabei um einen stabilen ‚Identifizier‘ im Zusammenhang mit dem digitalen Objekt handelt. Es wäre auch die Wahl der VD 17-Nummer denkbar gewesen, die ebenfalls in den MODS Metadaten aufgeführt ist⁵⁰ und als ‚Identifizier‘ für das Objekt über die Verwendung in der Staatsbibliothek hinaus Gültigkeit hat. Die ens:WebResource, die über ens:landingPage mit der ore:Aggregation verbunden ist, verfügt über einen URI, der in den Metadaten als PURL⁵¹ aufgeführt ist:

```
<mods:identifizier type="purl">  
http://resolver.staatsbibliothek-berlin.de/SBB000045E300000000  
</mods:identifizier>
```

Der Link führt zu der Abbildung des Vorderdeckels des hier als Beispiel verwendeten Objekts „Christiani Vita Et Corona“:

distinguishing the provenance of these assertions.” (Definition 2010, S.7) Dieses Vorgehen sollte in der LOD-Community weiter an Transparenz gewinnen.

⁴⁹ Die „Definition of the Europeana Data Model elements“ definiert eine ore:Aggregation mit „A set of related resources (Aggregated Resources), grouped together such that the set can be treated as a single resource. This is the entity described within the ORE interoperability framework by a Resource Map.” (Definition 2010, S.6)

⁵⁰ </mods:identifizier>

<mods:identifizier type="vd17">39:103909Q</mods:identifizier>

⁵¹ PURL steht für Persistent Uniform Resource Locator. Es handelt sich dabei um einen Identifizier in Form eines Uniform Resource Locator, der nicht direkt auf die Internet-Ressource verweist, sondern auf einen Linkresolver, der die aktuelle URL bereithält. Damit wird persistent auf Inhalte im Netz verwiesen, auch wenn sich Adressen von Websites ändern.



Abbildung 3.2 - „Christiani Vita Et Corona“ Vorderdeckel

Eine `ens:WebResource` mit der Eigenschaft `ens:landingPage` darf laut der „Definition of the Europeana Data Model elements“ in der Objektrepräsentation nur ein Mal vorkommen.⁵² Über die ‚Property‘ `ens:hasView` können beliebig viele Ansichten des Objekts als `ens:WebResource` mit der `ore:Aggregation` verbunden werden. Mindestens eine Voransicht – oder ein ‚Thumb‘ - des Objekts wird von Europeana zur Anzeige auf dem Europeana Portal bereitgehalten.

In oben modelliertem Beispiel (Abbildung 3.1) stammt der URI für die Voransicht aus der ‚File Section‘ der METS Datei aus der File Group „THUMBS“:

```
<mets:fileGrp USE="THUMBS">
<mets:file ID="FILE_0000_THUMBS" MIMETYPE="image/png">
<mets:FLocatLOCTYPE="URL" xlink:href="http://content.staatsbibliothek-berlin.de/dms/PPN651724848/100/0/00000001.png"/>
</mets:file>
```

Die Voransicht zeigt den oben abgebildeten Vorderdeckel des Objekts (Abbildung 3.2), der auch auf der ‚Landing Page‘ erscheint. Diese Ansicht ist leider weder attraktiv noch aussagekräftig, da es sich bei dem Vorderdeckel im Fall des vorliegenden Beispiels um einen nachträglich angebrachten Einband handelt. Die Staatsbibliothek zeigt aus diesem Grund (bei einem Zugriff auf das Objekt von ihren Digitalen Sammlungen aus) das Titelblatt als erste Ansicht (Abbildung 3.3).

⁵² Siehe dazu Definition 2010, S.25: „This property captures the relation between an aggregation representing a Cultural Heritage Object and the Web Resource representing that Object on the provider’s web site. (...) A (...) Europeana Aggregation may have at most one web resource as landing page, and a web resource may be the landing page of at most 1 Europeana Aggregation.”

The screenshot shows a digital library interface. At the top, there is a search bar and navigation links. Below that, the title 'Christiani Vita Et Corona, 1656' is displayed. On the left side, there is a list of metadata items, including 'Titelblatt', 'Vorderdeckel', 'Spiegel', 'Vorsatz', 'Titelblatt', 'Predigt', 'Personalia', 'Leichabdanckung', 'Heisse Thränen über den Traurigen doch Seeligen H...', 'Iusta Funebria Beatis Manibus ... Joachimi Gerste...', 'Kolophon', 'Einband', 'Vorsatz', 'Spiegel', 'Rückdeckel', and 'Farbinformation'. The main content area shows a thumbnail of the title page of the book, which is a Latin text with a decorative border. The title page text includes 'CHRISTIANI VITA ET CORONA' and 'Joachimi Gerstenbergers'.

Abbildung 3.3 - „Christiani Vita Et Corona“ Titelblatt

Wünschenswert wäre diese Ansicht auch für das Europeana Portal. Die Realisierung scheitert derzeit an dem Aufbau der METS ‚Structural Map‘ und den ‚Structural Links‘, die der obersten logischen Ebene des Werks - `<mets:dmdSec ID="DMDLOG_0000">` - sämtliche Files von `_0000` bis `_0081` als gleichwertige Kindelemente zuordnen, so dass eine Verlinkung nur auf das erste File (`_0000`) auf dieser Ebene möglich ist.

Denkbar ist es, der Objektrepräsentation auf der obersten logischen Ebene Voransichten zu allen Seiten des Werks über `ens:hasView` beizufügen. Die Reihenfolge zwischen den Voransichten wird in dem Fall über `ens:isNextInSequence` modelliert. Aus Gründen der Darstellbarkeit wurde in oben dargestelltem Graphen (Abbildung 3.1) auf das Abbilden aller 81 Voransichten verzichtet.

3.2 Der ore:Proxy: Deskriptive Metadaten

In einem nächsten Schritt der Modellierung werden die Metadaten das gesamte Werk betreffend, die in der METS/MODS Datei auf der Ebene „DMDLOG_0000“ aufgeführt sind, über `rdfs:properties` mit dem Proxy verbunden. Hier wird die Differenzierung der Metadaten, die das analoge Objekt, von jenen, die das digitale Objekt beschreiben, relevant: der Proxy bezieht sich (`ore:proxyFor`) auf das ‚PhysicalThing‘. Berücksichtigung finden in der Modellierung entsprechend die deskriptiven Metadaten, die das analoge Werk beschreiben. Die unter `<mods:edition>[Electronic ed.]</mods:edition>` angegebenen Daten beschreiben das digitale Werk und bleiben unberücksichtigt. Unberücksichtigt bleiben auch jene Metadaten, die nur für

den internen Gebrauch der Staatsbibliothek von Bedeutung sind, wie die Signatur für das analoge Werk und der Autor-Titel-Schlüssel.

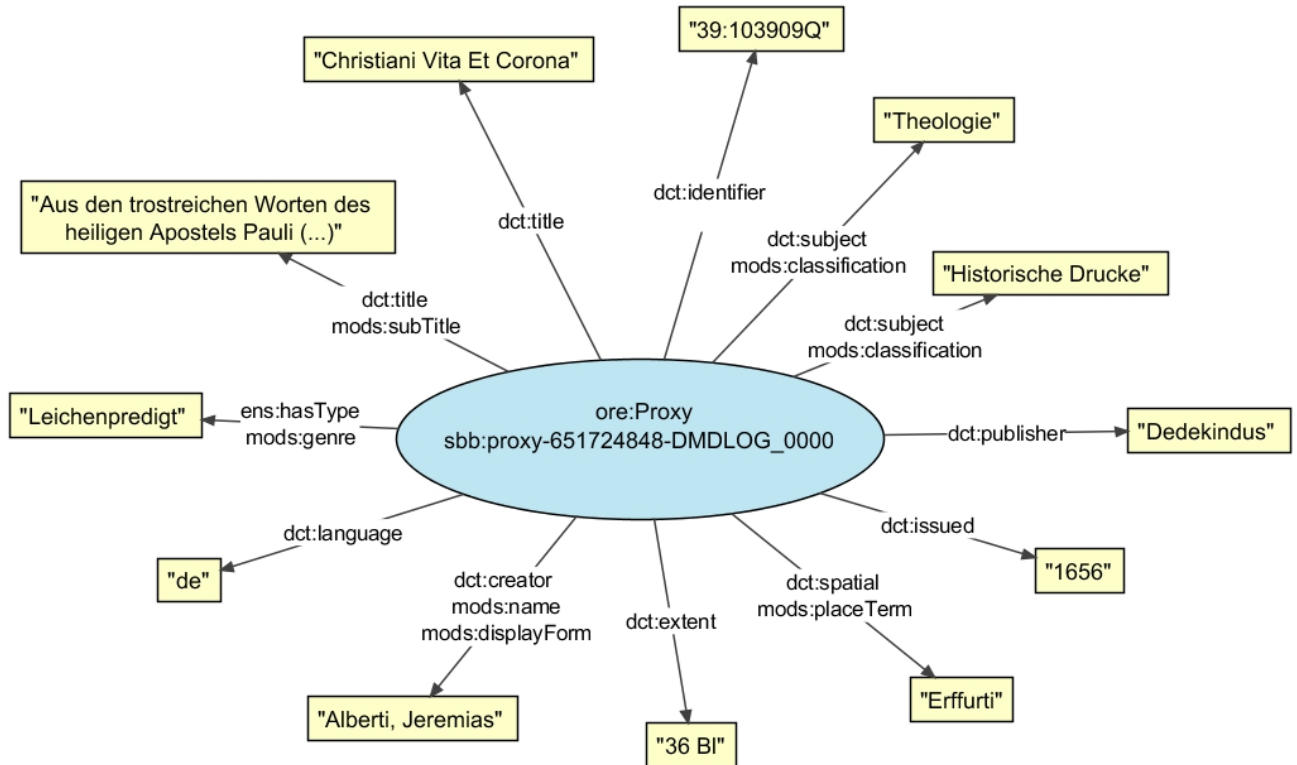


Abbildung 3.4 - ore:Proxy, deskriptive Metadaten

Die zur Beschreibung der Metadaten vorliegenden Elemente im MODS Format werden auf EDM-,Properties' – vor allem DCMI Metadata Terms (dct) - gemappt. Sind die Elemente des Ausgangsformats nicht deckungsgleich mit den im EDM verwendeten, bleiben sie als rdfs:subproperties neben den ,Properties' des EDM bestehen (wie im Beispiel mods:placeTerm in Abbildung 3.4).

In dem Graphen (Abbildung 3.4) sind alle relevanten Metadaten aus der ,Descriptive Metadata Section' in Form von ,Literals' als Objekte der RDF-,Statements' enthalten. Das Ziel von Europeana ist es, aus so vielen ,Literals' wie möglich, ,Resources' zu machen, um sie durch die Verlinkung mit der ,Semantic Data Layer' semantisch anreichern zu können:

Objects ingested by Europeana often use simple strings as values for the metadata field. Europeana hopes to update that information by linking objects to fully-fledged resources that are thoroughly described and are themselves connected to other resources, such as authority files for persons and thesauri for subjects. These resources enable richer functions, such as query expansion (e.g., using alternatives for a creator's name), recommendations of objects using semantic relations between them (objects created by connected artists), etc. (Isaac 2010, S.14f)

Die Daten im PICA-Format gehen zwar teils über „simple strings“ („Literals“) hinaus, da sie mit Normdaten verlinkt sind. Die Links gehen jedoch bei der Erstellung der METS/MODS Datei verloren. Auf die „Reparatur“ dieses Verlusts sei an späterer Stelle eingegangen. Hier soll zunächst eine Alternative zu der objektzentrierten Modellierung (Abbildung 3.4) der deskriptiven Metadaten gezeigt werden, die anschlussfähig für neue Verlinkungen ist.

3.3 Die Ereigniszentrierte Objektbeschreibung

Das Europeana Data Model ermöglicht neben einer, der bibliothekarischen Tradition verhafteten, objektzentrierten Beschreibung eine ereigniszentrierte Beschreibung des Objekts. Dabei fokussiert die Beschreibung auf die Ereignisse, in die das Objekt involviert war. Ein typisches Beispiel für eine solche Beschreibung aus der Museumswelt beträfe etwa ein Artefakt, das zu einem bestimmten Zeitpunkt erschaffen, zu einem späteren verschüttet und schließlich bei einer Grabung wieder zu Tage gefördert wurde. Die Idee für den Perspektivwechsel von dem Fokus auf das Objekt hin zu einem Fokus auf die Ereignisse, in die das Objekt involviert war, stammt bereits aus der Mitte der 1990er Jahre und wurde zunächst vor allem im Museumsbereich, zum Beispiel über CIDOC-CRM, realisiert.⁵³

Das Europeana Data Model realisiert die ereigniszentrierte Modellierung mit der Klasse `ens:Event` und den ‚Properties‘ `ens:wasPresentAt`, `ens:occuredAt` und `ens:happenedAt` (Abbildung 3.5).

⁵³ Siehe dazu Doerr et al. 2010, S.9: „Since the mid 1990’s, several communities world-wide began to understand (IndeCs, CRM, ABC, OPM) that the complexity of describing historical relationships can be normalized and dramatically be simplified, if, instead of putting the objects in the centre of the description, historical events, which „mediate“ in a sense all dynamic relationships between people, things, time and space, are put in the centre of documentation.“

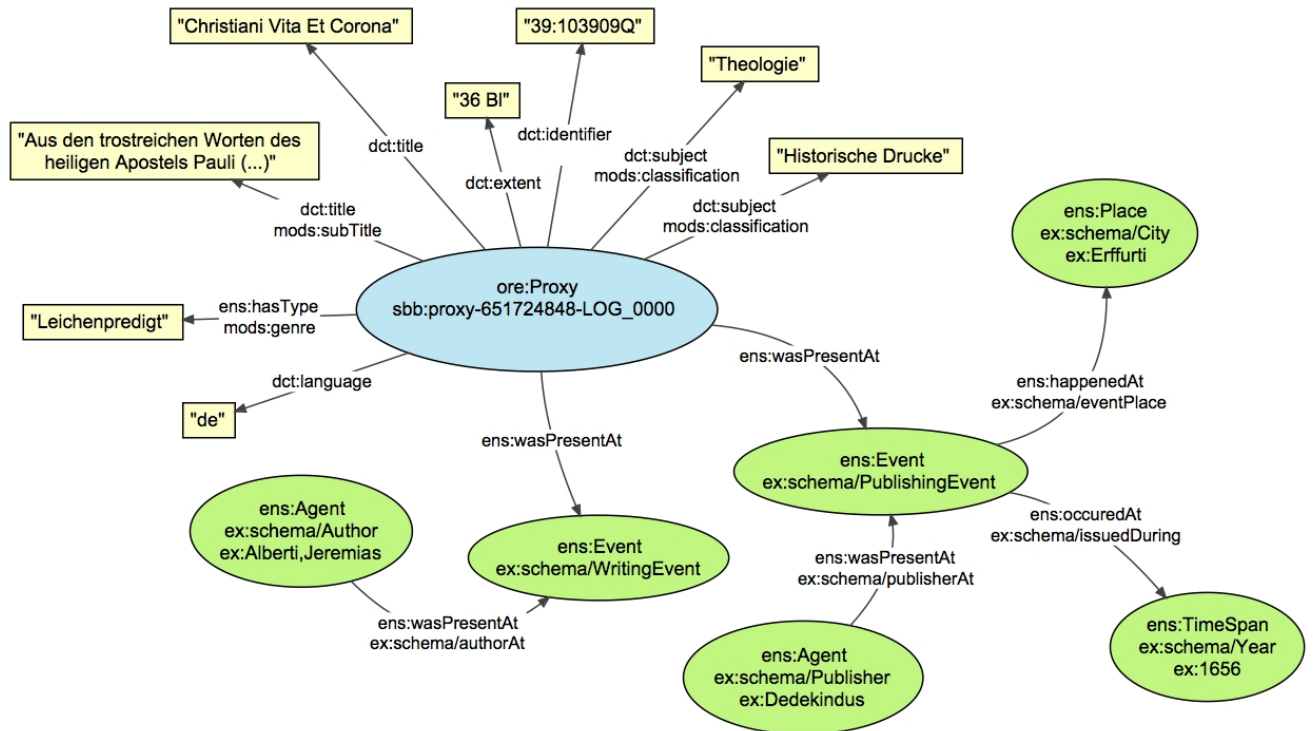


Abbildung 3.5 - Ereigniszentrierte Objektbeschreibung

Indem aus dem Schreib- und Druckvorgang `ens:Events` modelliert werden, sind Anschlussmöglichkeiten für Verbindungen mit weiteren ‚Resources‘ zu neuen RDF-‚Triples‘ geschaffen. Auf diese Weise können die Angaben zu Autor, Drucker, Erscheinungsjahr und Erscheinungsort mit den Klassen `ens:Agent`, `ens:TimeSpan` und `ens:Place` als ‚Resources‘ ausgedrückt werden und eröffnen so die Möglichkeit einer Verknüpfung mit den kontrollierten Vokabularen der ‚Semantic Data Layer‘. Mit „`ex:schema...`“ wird hier angedeutet, dass die ‚Events‘ mit einem eigens entwickelten Schema noch näher spezifiziert werden können. Diese Spezifizierung wird als `rdfs:subClass` aufgenommen. Eine andere Möglichkeit der Spezifizierung des ‚Events‘, die derzeit in der Diskussion ist, bedient sich der ‚Property‘ `dct:hasType` und verbindet damit das `ens:Event` mit einem entsprechenden Term eines kontrollierten Vokabulars. Auf die Möglichkeit der semantischen Anreicherung wird im Zusammenhang mit den zu reintegrierenden Normdaten noch einmal eingegangen.

3.4 Die Beschreibung komplexer Objekte

In einem nächsten Schritt werden die Strukturinformationen des hier exemplarisch behandelten Objekts „*Christiani Vita Et Corona*“ modelliert. Es handelt sich dabei um ein komplexes Objekt mit - dem strengen Aufbau einer Leichenpredigt folgend - diversen Abschnitten und den üblichen den Einband des Buches beschreibenden Strukturelementen. Die Angaben zur Struktur stammen aus der ‚Structural Map‘ der METS/MODS Datei:


```

<mets:structMap TYPE="LOGICAL">
  <mets:divADMID="AMD"CONTENTIDS="http://resolver.staatsbibliothekberlin.de/SBB000045E30
  000000"DMDID="DMDLOG_0000" ID="LOG_0000" LABEL="Christiani Vita Et Corona"
  TYPE="monograph">
    <mets:div ID="LOG_0001" TYPE="binding">
      <mets:div ID="LOG_0002" TYPE="cover_front"/>
      <mets:div ID="LOG_0003" TYPE="paste_down"/>
      <mets:div ID="LOG_0004" TYPE="endsheet"/>
    </mets:div>
    <mets:div ID="LOG_0005" TYPE="title_page"/>
    <mets:div ID="LOG_0006" TYPE="sermon"/>
    <mets:div ID="LOG_0007" TYPE="personalia"/>
    <mets:div ID="LOG_0008" TYPE="funeral_oration"/>
  <mets:div DMDID="DMDLOG_0001" ID="LOG_0009" LABEL="Heisse Thränen über den
  Traurigen doch Seeligen Hintrit Deß Edlen ... Herrn Joachims Gerstenbergers Hiesiger Stadt
  ältesten und wohlverdienten Obristen Vier-Herrns" TYPE="epicedia"/> (...)

```

Die oberste logische Hierarchieebene „LOG_0000“ aggregiert die Strukturelemente der darunter liegenden Ebene „LOG_0001“ – „LOG_0016“. Aus Gründen der Darstellbarkeit sollen hier nur zwei der aggregierten Strukturelemente – „LOG_0001“ und „LOG_0009“⁵⁴ – abgebildet werden:

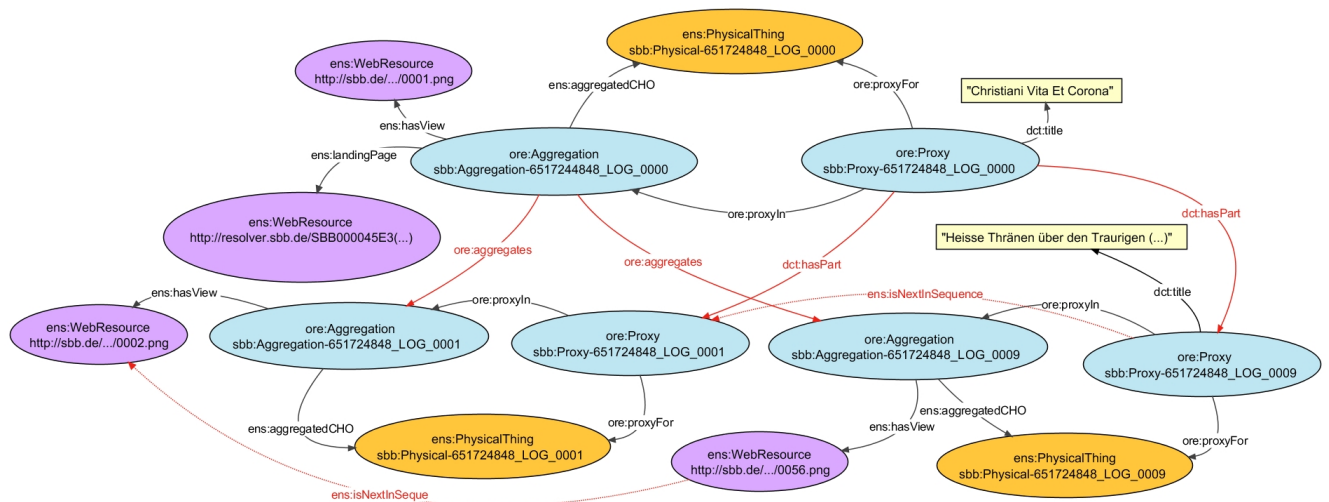


Abbildung 3.6 - Beschreibung der hierarchischen Struktur

Unter Verwendung der (hier in rot dargestellten) ‚Properties‘ ore:aggregates, dct:hasPart und ens:isNextInSequence, können mit dem EDM Kompositobjekte, in diesem Fall ein Objekt, das aus mehreren Abschnitten besteht, dargestellt werden. Dabei aggregiert die dem ganzen Werk entsprechende oberste hierarchische Ebene (in der logischen Struktur der XML-Datei „LOG_0000“) sämtliche Elemente der darunter liegenden Ebene mit ore:aggregates. Der stellvertretend für das gesamte Werk stehende ‚Proxy‘ schafft die Verbindung zu den einzelnen Abschnitten mit dct:hasPart. Die ‚Property‘ ens:isNextInSequence zeigt die Reihenfolge an, in

⁵⁴ „LOG_0009“ wurde ausgewählt, da es sich hierbei um den ersten Teil der zu der Leichenpredigt gehörigen Epicedien handelt. Sie werden mit einer eigenen analogen PPN wie selbstständige Werke katalogisiert und haben im Fall des ersten Teils auch einen eigenen Autor. Über die Reintegration der PICA-Daten können die bibliographischen Angaben der METS/MODS-Datei ergänzt und mit den Normdaten angereichert werden.

der die Abschnitte zueinander stehen. Sie verbindet sowohl die ‚Proxies‘ wie auch die ‚WebResources‘ miteinander. In dem Graphen (Abbildung 3.6) ist die Verbindung über `ens:isNextInSequence` lediglich angedeutet, da die beiden exemplarisch verwendeten Strukturelemente nicht direkt aufeinander folgen.

3.5 „*Christiani Vita Et Corona*“ im PICA3-Format

Im Digitalisierungsprozess kam es bei der Erstellung der METS/MODS-Datei aus den PICA-Daten zu einem Informationsverlust. Im Folgenden soll mit einem Blick auf einen Auszug der vom GBV stammenden Daten im PICA3-Format das Ausmaß des Datenverlusts festgestellt werden:

```
0100 091865476
1500 /1ger
2277 VD17 39:103909Q
3000 Jeremias@Alberti!378676598! IJeremias@Alberti ; PND-ID: 12872370X N
3010 Joachim@Gerstenberg!153121947! IJoachim@Gerstenberg ; PND-ID: 102525838 N
3062 Johann@Möller [Beiträger]
3062 Nicolaus@Stenger [Beiträger]
3062 Zacharias@Hogel [Beiträger]
3062 Christian Timotheus@Duftius [Beiträger]
3062 Samuel@Boccius [Beiträger]
3062 Johann Melchior@Starckloff [Beiträger]
3062 Wilhelm Hieronymus@Stenger [Beiträger]
3062 Joachim@Gerstenberg [Beiträger]
3062 Joachim@Schefer [Beiträger]
3062 Johann Georg@Schäffer [Beiträger]
3062 Georg Nicolaus@Stenger [Beiträger]
3062 @M. H. G." [Beiträger]
3062 Johannes@Jacobi [Beiträger]
3062 Joachim@Cramer [Beiträger]
3260 Beschreibung des Lebens und Crone eines waren Christen
4000 Christiani Vita Et Corona = Beschreibung des Lebens und Crone eines waren Christen :
Aus den Trostreichen Worten des heiligen Apostels Pauli in der II. an Timoth. am IV. v. 7. ý& 8.
Ich habe einen guten Kampf gekämpft/ [et]c. Bey der ... LeichBestattung Des ... Joachimi
Gerstenbergers/ wohlverdienten ältesten Obristen Vierherrns in unserer Christlichen Polickey
Welcher den 18. Ianuarii ... dieses 1656. Jahrs ... entschlaffen/ und hernach den 24. eiusdem ...
in der PredigerKirchen beygesetzt worden / Auf Anordnung und Begehren ein und angeführet
von M. Jeremia Alberti, der Christlichen Gemeinde zu den Predigern Pastore
4030 Erfurti : Dedekindus
4033 Erfurt
4043 !221265686! IFriedrich Melchior@Dedekind. - Erfurt (1627-1668) N
4060 [36] BI
4217 Erscheinungsvermerk im Kolophon. - Erscheinungsjahr nach dem Sterbejahr angegeben
4222 Enth. außerdem: [1] Heisse Thränen über den Traurigen doch Seeligen Hintrit Deß ... Herrn
Joachims Gerstenbergers ... / vergossen von desselben Ergebenen jüngsten Sohne Sigismundo
Gerstenbergern. [2] Iusta Funebria Beatiss Manibus ... Joachimi Gerstenbergeri ...
4243 Digitalisierte Ausg.: Alberti, Jeremias: Christiani Vita Et Corona N=GBV *651724848*
5570 !09663216X! ILeichenpredigt N
4803 Gestempelter Besitzvermerk: Fürstlich Stolberg. Bibliothek Wernigerode
```

Über den in der METS/MODS Datei angegebenen ‚Identifizier‘

```
<mods:identifizier type="PPNanalog">PPN091865476</mods:identifizier>
```

kann die Verbindung zu dem PICA-Datensatz des GBV hergestellt werden. In PICA3 befindet sich die PPN zu dem analogen Werk in dem Feld 0100.⁵⁵ Neben dem Verfasser der Leichenpredigt (PICA3 3000) sind in diesem Datensatz die gefeierte Person (PICA3 3010) und sämtliche Beiträger (PICA3 3062) aufgeführt. Im Fall des Verfassers Jeremias Alberti und der gefeierten Person Joachim Gerstenberg sind auch die PND-Nummern⁵⁶ angegeben. Den Erscheinungsort Erfurt führt der PICA-Datensatz nicht nur, wie es in METS/MODS der Fall ist, in der Vorlageform (Erfurti), sondern auch in der heute gebräuchlichen Form Erfurt auf. Ebenso verhält es sich mit dem Namen des Druckers, der in der METS/MODS Datei mit Dedekindus angegeben ist, in PICA mit vollem Namen als Friedrich Melchior Dedekind (PICA3 4043). Und schließlich sind die Leichenpredigten in den PICA-Daten über den Gattungsbegriff zu recherchieren (PICA3 5570). Zwar ist der Gattungsbegriff in dem hier untersuchten Beispiel auch in MODS vorhanden. Eine Recherche in den Digitalisierten Sammlungen der Staatsbibliothek über den Suchbegriff „Leichenpredigt“ ergibt jedoch gegenüber den etwa 1800 zugänglichen Leichenpredigten lediglich eine Treffermenge von 172 Werken.⁵⁷ Das lässt darauf schließen, dass der Gattungsbegriff in MODS nicht konsequent vergeben wird.

0100 **09176842X**

3000 **Sigismund@Gerstenberg**!546774164! ISigismund@Gerstenberg ; **PND-ID: 133484548 N**
4000 #1#!091865476! IChristiani Vita Et Corona : Aus den Trostreichen Worten des heiligen Apostels Pauli in der II. an Timoth. am IV. v. 7. y& 8. Ich habe einen guten Kampf gekämpft/ [et]c. Bey der ... LeichBestattung Des ... Joachimi Gerstenbergers/ wohlverdienten ältesten Obristen Vierherrns in unserer Christlichen Polickey Welcher den 18. Ianuarii ... dieses 1656. Jahrs ... entschlaffen/ und hernach den 24. eiusdem ... in der PredigerKirchen beygesetzt worden / Alberti, Jeremias. - Erfurti : Dedekindus, 1656 N

4004 Heisse Thränen über den Traurigen doch Seeligen Hintrit Deß Edlen ... Herrn Joachims Gerstenbergers Hiesiger Stadt ältesten und wohlverdienten Obristen Vier-Herrns

In den PICA-Daten, die den ersten Teil der Epicedien (mit einer eigenen PPN) beschreiben, ist – anders als in MODS („DMDLOG_0001) – der Verfasser angegeben (PICA3 3000). Über die angegebene PND-ID lässt sich recherchieren, dass es sich bei dem Verfasser Sigismund Gerstenberg um den Sohn des Verstorbenen Joachim Gerstenberg handelt.

In den Graphen werden die PICA3-Metadaten mit einem weiteren ‚Proxy‘ integriert. Über die ‚Property‘ ore:proxyFor wird er mit dem ens:PhysicalThing verbunden. Auf diese Weise können aus der Perspektive zweier Quellen (sbb:Proxy-651724848_LOG_0000 und gbv:Proxy-091865476 bzw. sbb:Proxy-651724848_LOG_0009 und gbv:Proxy-09176842X) Aussagen über

⁵⁵ Ergänzend zu dem Mapping von METS/MODS auf das EDM im Rahmen dieser Arbeit und des Prototyping des EDM wurde von Steffen Hennicke ein Mapping von PICA3 auf das EDM erstellt.

⁵⁶ PND steht für Personennamendatei und ist eine Normdatei die von der Deutschen Nationalbibliothek verwaltet und von allen deutschen und österreichischen Bibliotheksverbänden gemeinsam geführt wird. Für jede in die PND aufgenommene Person wird ein Datensatz mit einem eindeutigen ‚Identifier‘ angelegt, über den auf den Datensatz verwiesen werden kann (DNB PND).

⁵⁷ Diese Zahl stammt vom 11.05.2011. Da derzeit täglich neu digitalisierte Bestände online zugänglich gemacht werden, unterliegt die Trefferanzahl ständigen Veränderungen.

eine ‚Resource‘ getroffen werden. Die mit den PICA-Daten neu hinzugekommenen Informationen sind in dem Graphen (Abbildung 3.7) rot markiert. Bei den Beiträgern handelt es sich – aus Platzgründen - um eine Auswahl.

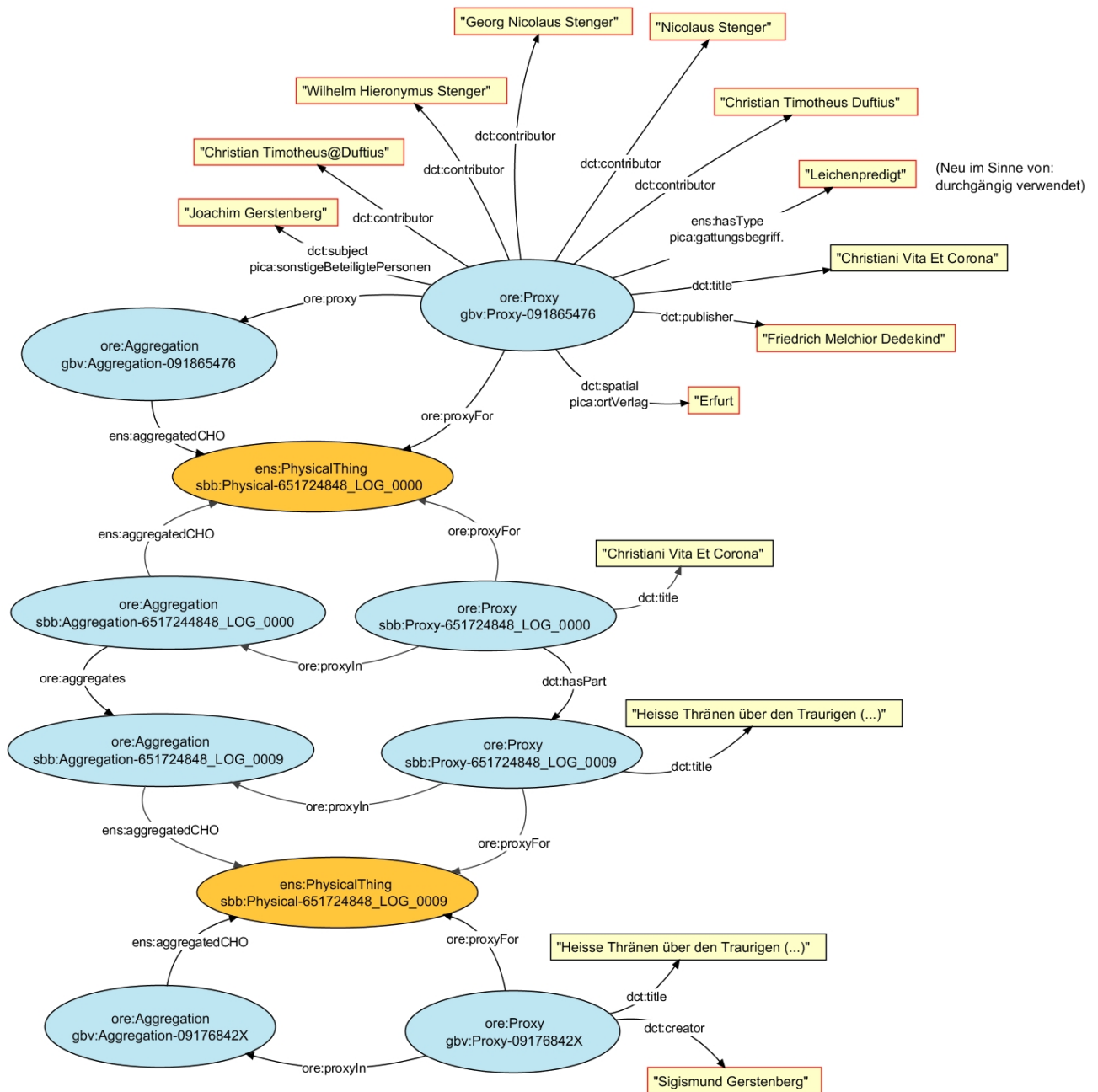


Abbildung 3.7 - Zwei Sichten auf ein Objekt

In dem Graphen (Abbildung 3.7) sind die Metadaten als ‚Literals‘ modelliert. Folgend soll ein Ausblick auf mögliche Anreicherungsszenarien gegeben werden.

Da es sich bei der Leichenpredigt um eine Sammlung von Beiträgen handelt, die im Zusammenhang mit einem Begräbnis verfasst und vorgetragen wurden, bietet sich hier eine ereigniszentrierte Modellierung der Daten um ein ens:Event an.

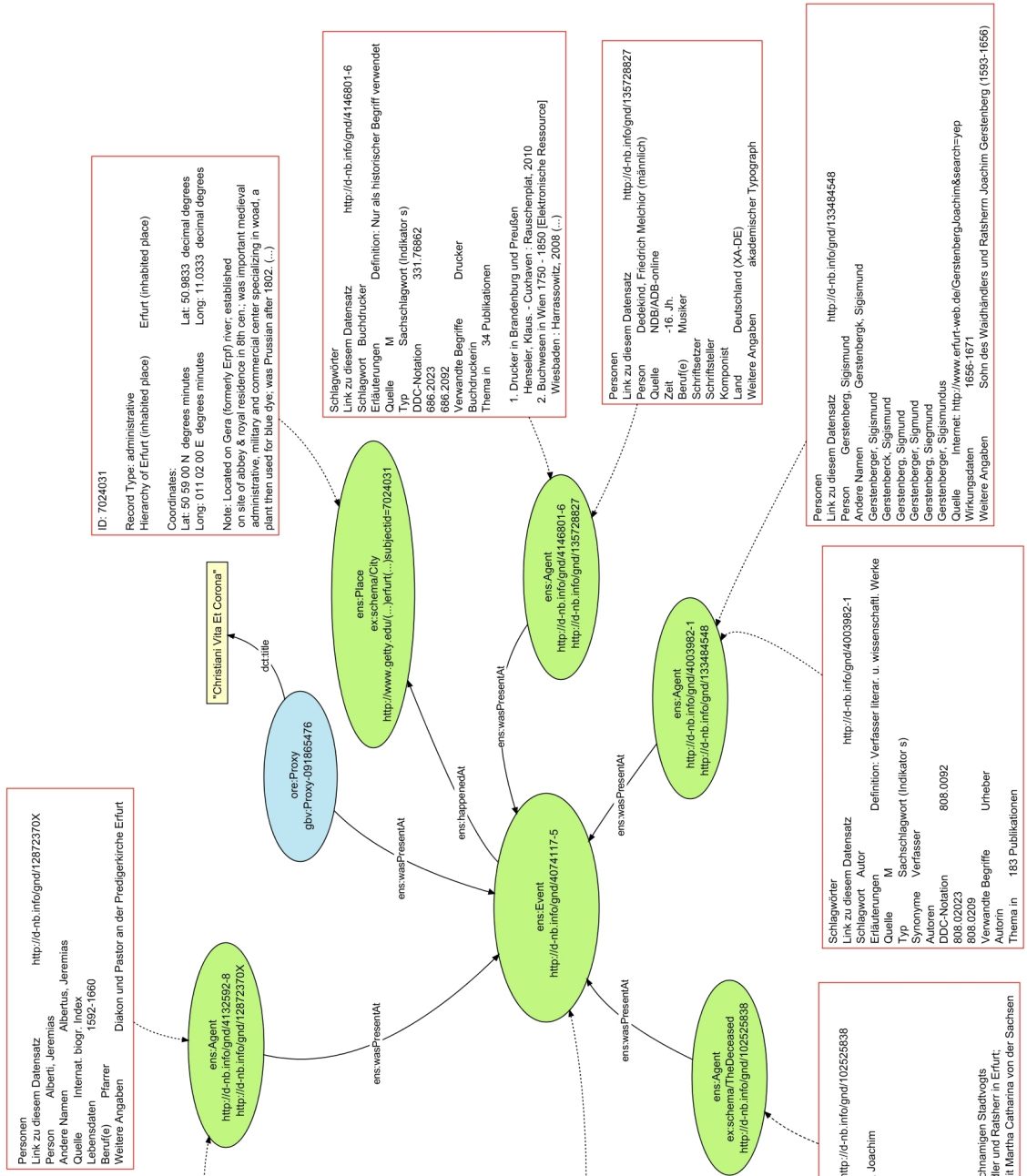


Abbildung 3.8 - Eindeutige Referenzierung der Personen mit der PND-ID

Abbildung 3.8

Für eine eindeutige Referenzierung der Personen kann zum einen die VIAF-ID herangezogen werden oder – und das scheint im Fall der vorliegenden Daten sinnvoller – die PND-ID. Die PND-ID ist für drei Personen aus dem hier behandelten Beispiel in den PICA-Daten bereits angegeben. Für die weiteren Personen – den Drucker und die Beiträger – können die PND-IDs über einen String-Abgleich mit Einträgen in der Personennamendatei ermittelt werden. Die Angaben in der Personennamendatei ermöglichen die Anreicherung der Informationen zu den Personen, beispielsweise um Berufe und Verwandtschaftsbeziehungen. Auch der Erscheinungsort und die Textgattung werden mit einem ‚Identifier‘, etwa aus dem Getty Thesaurus of Geographic Names (TGN) oder der Schlagwortnormdatei (SWD), referenzierbar und lassen sich mit den dort hinterlegten Angaben anreichern.

Der Graph (Abbildung 3.8) zeigt beispielhaft, wie eine Referenzierung und Anreicherung über URIs aussehen könnte. Das `ens:Event` erhält den ‚Identifier‘ zu dem Schlagwort „Leichenpredigt“ aus der Schlagwortnormdatei. Der Datensatz der Schlagwortnormdatei gibt zu dem Schlagwort „Leichenpredigt“ die Oberbegriffe „Predigt“ und „Personalschrift“ an, die über `skos:broader` (Miles et al. 2009) mit dem `ens:Event` verbunden werden können. Die weiteren Metadaten werden – hier nur ausschnittsweise – um das Ereignis „Leichenpredigt“ gruppiert. Der Pfarrer und Verfasser der Leichenpredigt, Jeremias Albert, wird als an dem Ereignis beteiligter `ens:Agent` modelliert. Ein ‚Identifier‘ verweist auf einen Eintrag in der Schlagwortnormdatei zu dem Schlagwort „Pfarrer“. Dort findet sich neben der DDC-Notation der Hinweis auf Publikationen zu diesem Schlagwort. Der zweite ‚Identifier‘ verlinkt den `ens:Agent` mit dem Eintrag zu Jeremias Alberti in der Personennamendatei. Die PND-ID sorgt hier für eine eindeutige Referenzierbarkeit der Person und verlinkt mit weiteren relevanten Informationen: verschiedenen Schreibweisen des Namens, die über `skos:prefLabel` und `skos:altLabel` in den Graphen integriert und suchbar gemacht werden können und der Angabe, dass Jeremias Alberti Diakon und Pastor an der Predigerkirche Erfurt gewesen ist. Derartige Informationen lassen sich automatisch aus den PND-Einträgen extrahieren und können zu neuen Verlinkungen – etwa mit dem Wikipedia-Eintrag zu der Predigerkirche Erfurt – führen. Erfurt ist als `ens:Place` über den ‚Identifier‘ aus dem Getty Thesaurus of Geographic Names mit geographischen und historischen Angaben über die Stadt verbunden.

In dem Datensatz der Personennamendatei zu Sigismund Gerstenberg, ein weiterer an dem `ens:Event` beteiligter `ens:Agent`, findet sich neben verschiedenen Schreibweisen seines Namens die Aussage, dass es sich bei ihm um den Sohn des Verstorbenen handelt.

Bereits an der Verlinkung mit der (als LOD verfügbaren) Gemeinsamen Normdatei (PND und SWD) lässt sich das Potential semantischer Nutzungsszenarien für die Suche in den mit dem EDM modellierten Daten erahnen. Nähere Ausführungen dazu folgen im nächsten Kapitel.

4. Ein Ausblick auf neue Suchoptionen und semantische Nutzungsszenarien

Nachdem die vorliegende Arbeit den Fokus auf die Datenmodellierung mit dem Europeana Data Model einnahm, gilt es nun den Mehrwert, den die Suche aus dieser Modellierung gegenüber den Suchmöglichkeiten in den mit dem Format METS/MODS beschriebenen Daten ergibt.

4.1 Das Integrationspotential des Europeana Data Model

Ein wesentlicher Mehrwert für die Suche resultiert aus der Reintegration der Metadaten und der Links zu den Normdaten, die bei der Erstellung der METS/MODS-Datei aus den PICA3-Daten abhandeln kommen. Die Reintegration der Daten verdankt sich dem Integrationspotential des EDM-Ansatzes: der ‚Proxy‘-Mechanismus ermöglicht die Sichtweisen verschiedener Inhaltsanbieter auf ein Objekt und führt im Falle des hier beschriebenen Beispiels die Metadaten der Staatsbibliothek mit jenen des GBV zusammen. Daraus ergeben sich neue Möglichkeiten für die Suche: die Suche nach Objekten in den Digitalisierten Sammlungen der Staatsbibliothek beschränkt sich derzeit auf die einfache Suche (Abbildung 4.1). Dabei besteht die Option, die Suche in den Materialarten und Fächern auf eine oder mehrere Kategorien einzuschränken oder eine Kategorie direkt anzusteuern.

The screenshot shows the search interface for 'Digitalisierte Sammlungen der SBB'. The page title is 'Digitalisierte Sammlungen der SBB'. Below the title, there is a search bar and a 'Suche starten!' button. The text below the search bar reads: 'Hier finden Sie Digitalisate von Büchern, Handschriften und anderen Medien, die sich physisch im Bestand der Staatsbibliothek zu Berlin befinden. Derzeit sind dies insgesamt 11267 digitalisierte Werke.' Below this, there are two sections for filtering the search results: 'Materialarten' and 'Fächer'. Each section has a list of categories with checkboxes and counts. The 'Materialarten' list includes: Historische Drucke (10101), Einblattdrucke (25), Inkunabeln (28), Kinder- und Jugendbücher (5), Handschriften (19), Nachlässe und Autographe (119), Orientalische Handschriften (220), Karten (5), Musiknoten (513), and Porträts, Bildmaterialien (3). The 'Fächer' list includes: Allgemeines/Wissenschaftskunde /Literaturgeschichte (422), Theologie (3392), Politik/Staat/Gesellschaft/Wirtschaft (119), Rechtswissenschaft (1672), Kriegskunst (16), Medizin (315), Naturwissenschaften/Mathematik (119), Aberglaube/Mystische Philosophie (63), Pädagogik (45), Philosophie/Psychologie (266), and Kunst (20). There are also links for 'einschränken' and 'stöbern'.

Abbildung 4.1 - Sucheinstieg Digitalisierte Sammlungen (Staatsbibliothek)

Angezeigt wird das Objekt nur über Suchbegriffe, die in der METS/MODS-Datei enthalten sind. Das schränkt die Möglichkeiten für den direkten Einstieg über die Digitalisierten Sammlungen erheblich ein: das Objekt lässt sich weder über den Namen der gefeierten Person, noch über die Beiträge aufrufen. Der Name des Verstorbenen kommt zwar in den Titeln vor, allerdings nur in rektierter Form - eine Suche nach dem Nominativ zeigt folglich ein Null-Treffer-Ergebnis an. Der Druckort und der Drucker sind in der METS/MODS-Datei in der historischen Vorlageform aufgeführt, so dass auch die Suche nach diesen Angaben nur in der Vorlageform („Erfurti“ und „Dedekindus“) Erfolg hat.

Mit der Integration der GBV-Daten (Abbildung 3.7 und 3.8) kann das Objekt auch über die gefeierte Person und die Beiträge aufgerufen werden. Die Normdaten reichern die Daten um verschiedene Schreibweisen der Personen- und Ortsnamen an. Da der Gattungsbegriff in den GBV-Daten durchgängig vergeben ist, liefert er in der Suche ein vollständiges Set der Leichenpredigten.

4.2 SPARQL – Eine RDF Anfrage Sprache

SPARQL – eine Abkürzung für SPARQL Protocol and RDF Query Language - ist eine Sprache zur Anfrage auf RDF-Graphen. Seit Januar 2008 hat sie den Status einer offiziellen W3C- ‚Recommendation‘ (Prud'hommeaux et al. 2008). SPARQL-Queries bieten die Möglichkeit, Anfragen über mehrere Graphen verteilt zu stellen. Das prädestiniert SPARQL für die Verwendung in Szenarien, in denen ein Zugriff auf verteilte Datenbestände erforderlich ist. Die Kommunikation zwischen RDF-Daten und SPARQL erfolgt über das SPARQL Protocol for RDF mit sogenannten SPARQL-Endpoints. Ein Beispiel für einen SPARQL-Endpoint bietet die DBpedia.org, eine Datenbank, die Wikipedia-Inhalte in RDF bereitstellt. SPARQL umfasst mehrere Anfragetypen, die Ergebnisse in verschiedener Form - etwa als Tabelle mit Variablenwerten („Select“) oder als RDF-Graph („Construct“) - liefern. Bei einer Anfrage erfolgt ein Graphmuster-Vergleich, der zu einer Menge von Ergebnissen führt. Jedes Ergebnis liefert eine Menge an Variablen (Schlönvoigt 2009).

Die Suchoptionen auf der Grundlage von SPARQL übertreffen die Suche in konventionellen Bibliothekskatalogen. Im Fall des in Kapitel 3 modellierten Graphen (Abbildung 3.8) ließe sich beispielsweise – vorausgesetzt die Daten liegen vollständig in RDF beschrieben vor – mit SPARQL ein Set aller Drucker und Druckorte des VD 17 in Verbindung mit der Zeitangabe, zu der an ihnen gewirkt wurde, generieren. Für die Präsentation des Anfrageergebnisses bietet sich eine ‚Timeline‘ an, ähnlich jener, die im Europeana Portal eingesetzt wird (Abbildung 4.2).

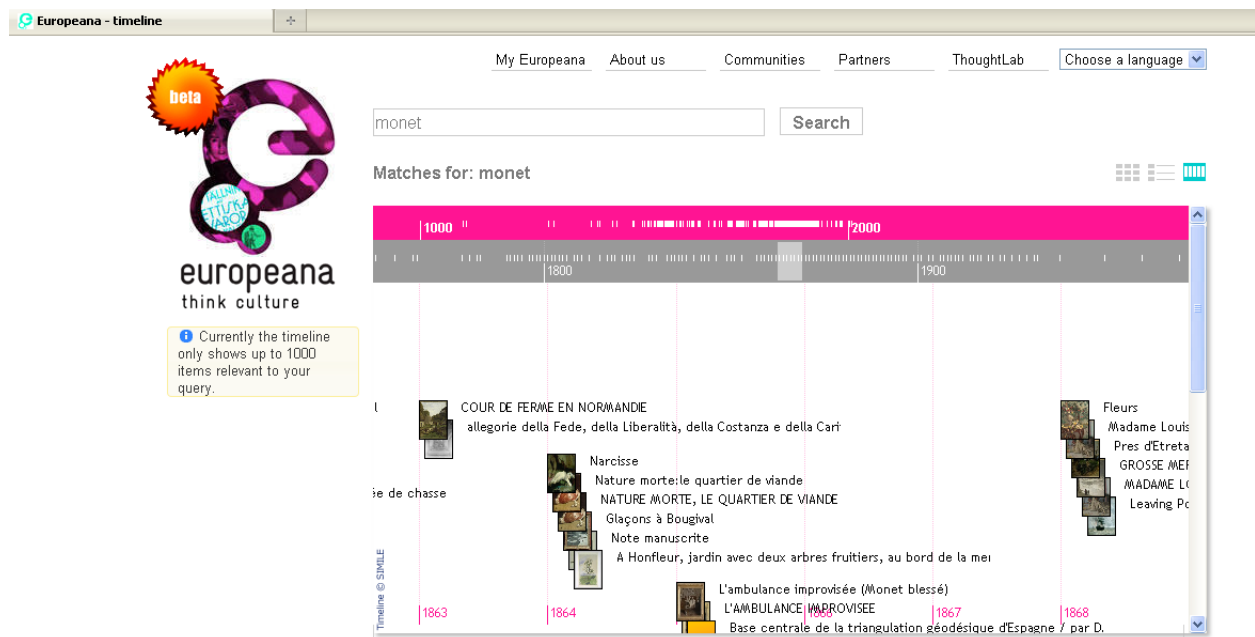


Abbildung 4.2 - Visualisierte Darstellung der Suchergebnisse auf einer ‚Timeline‘ (Europeana)

4.3 Personennetze mit dem Resource Description Framework

In der Verlinkung der deskriptiven Metadaten mit Normdaten liegt ein wichtiges Potential für die Schaffung von Mehrwerten. Die Deutsche Nationalbibliothek stellt die Gemeinsame Normdatei als Linked Data in RDF zur Verfügung. Über einen ‚String‘-Abgleich mit den Daten der GND können beispielsweise die Beiträge in dem hier behandelten Beispiel mit Normdaten angereichert werden (Abbildung 3.8). Dabei ergibt sich mit der automatischen Extraktion⁵⁸ der Verwandtschafts- und Berufsbezeichnungen aus den Normdaten und ihrer Beschreibung in RDF die Grundlage für die visuelle Darstellung eines Personennetzes, aus dem sich die Beziehungen der beteiligten Personen untereinander ablesen lassen.

Als Beispiel ist der GND-Eintrag von Anna Christina Gerstenberg aufgeführt, die, wie das Personennetz zeigen wird (Abbildung 4.3), eine Schlüsselrolle in dem Beziehungsgeflecht der Beiträge der Leichenpredigt „Christiani Vita Et Corona“ einnimmt.

⁵⁸ Eine automatische Extraktion von Informationen aus Texten kann mit sogenannten „semantic tagging tools“ auf der Basis von Named Entity Recognition erfolgen. Beispiele dafür bieten Open Calais, eine proprietäre Web Anwendung, die ihre Nutzung bis zu einem bestimmten Umfang kostenlos zur Verfügung stellt, und GATE, eine Open Source Anwendung. Beide Anwendungen evaluieren Gerber et al. (2011) ausführlich. Ein Problem stellt in der Named Entity Recognition nach wie vor die zum großen Teil bestehende Beschränkung auf die englische Sprache dar. „Named Entity Recognition is an important preprocessing step for many NLP tasks. It finds usage in applications like Textual Entailment, Question Answering, and Information Extraction. As is often the case for NLP tasks, most of the work has been done for English. To our knowledge, at this time there is no single “off-the-shelf” NER system for German freely available for academic purposes. A major reason for this situation is the (un-)availability of labelled development data in the respective languages. For English, many large corpora annotated with named entities are available from a number of shared tasks and bakeoffs, including CoNLL 2003, MUC 2006/2007 and ACE 2008.” (Faruqui et al. 2010) Mehrsprachige Textanalyse-Lösungen bietet die Firma Temis mit ihrer Software-Plattform Luxid. Luxid kann Inhalte in Texten automatisch mit relevanten Informationen anreichern und vernetzen und unterstützt so die Orientierung in großen Textmengen. (Temis)

Link zu diesem Datensatz	http://d-nb.info/gnd/128882948
Person	Gerstenberg, Anna Christina Gerstenberg, Anna Christiana Gerstenberg, Anne Christine Gerstenberger, Anne Christine Gerstenberger, Anna Christiana
Andere Namen	Gerstenberger, Anna Christina Stenger, Anna Christina [Früherer Name] Stenger, Anne Christine Stenger, Anna Christiana
Lebensdaten	1645-1673
Weitere Angaben	Tochter des Pfarrers Nicolaus Stenger; Ehefrau des Erfurter Kaufmanns Georg Heinrich Gerstenberg (1643-1673)

Tabelle 2 - Datensatz der PND zu Anna Christina Gerstenberg (aus: Gemeinsame Normdatei der Deutschen Nationalbibliothek)

Neben der GND-ID (in dem Link) lassen sich dem Eintrag verschiedene Namensschreibweisen, die Lebensdaten und verwandtschaftlichen Beziehungen entnehmen.

Die Beschreibung des Datensatzes mit RDF/XML enthält noch weitere Informationen:

```
<rdf:RDF><rdf:Description rdf:about="http://d-nb.info/gnd/128882948">
<gnd:variantNameForThePerson rdf:parseType="Resource"><gnd:foreName>AnneChristine
</gnd:foreName><gnd:surname>Stenger</gnd:surname><gnd:usedRules>RAKW</gnd:usedRules>
</gnd:variantNameForThePerson>
<rdaGr2:placeOfBirthxml:lang="de">Erfurt</rdaGr2:placeOfBirth>
<gnd:variantNameForThePerson>Gerstenberger, Anna Christina</gnd:variantNameForThePerson>
<gnd:publicationOfThePerson xml:lang="de">
Andreae, Johann Wilhelm: Christlicher Leich-Sermon von Rechtschaffener Christen ungleichem
Zustande, Aus der Offenbarung Johannis c. 7. v. 13. &c. ... - 1674</gnd:publicationOfThePerson>
<gnd:variantNameForThePerson>Gerstenberg, Anna Christiana</gnd:variantNameForThePerson>
<owl:sameAs rdf:resource="http://viaf.org/viaf/3537303"/>
<rdaGr2:biographicalInformation xml:lang="de">
Tochter des Pfarrers Nicolaus Stenger; Ehefrau des Erfurter Kaufmanns Georg Heinrich Gerstenberg
(1643-1673)</rdaGr2:biographicalInformation>
<gnd:preferredNameForThePerson rdf:parseType="Resource">
<gnd:foreName>Anna Christina</gnd:foreName> <gnd:surname>Gerstenberg</gnd:surname>
<gnd:usedRules>RAK-WB</gnd:usedRules>
</gnd:preferredNameForThePerson>
<gnd:variantNameForThePerson rdf:parseType="Resource">
<gnd:foreName>Anna Christiana</gnd:foreName>
<gnd:surname>Stenger</gnd:surname>
<gnd:usedRules>RAK-WB</gnd:usedRules>
</gnd:variantNameForThePerson>
<gnd:variantNameForThePerson>Gerstenberger, Anne Christine</gnd:variantNameForThePerson>
<gnd:variantNameForThePerson>Stenger, Anne Christine</gnd:variantNameForThePerson>
<gnd:variantNameForThePerson>Gerstenberg, Anne Christine</gnd:variantNameForThePerson>
<gnd:preferredNameForThePerson>Gerstenberg, Anna Christina</gnd:preferredNameForThePerson>
<gnd:variantNameForThePerson rdf:parseType="Resource">
<gnd:foreName>AnneChristine</gnd:foreName>
```

```

<gnd:surname>Gerstenberger</gnd:surname>
<gnd:usedRules>RAK-WB</gnd:usedRules>
</gnd:variantNameForThePerson>
(...)

```

Die unterschiedlichen Schreibweisen des Namens werden hier mit `gnd:variantNameForThePerson` – einem Äquivalent zu `skos:altLabel` – beschrieben. Mit `gnd:publicationOfThePerson` wird auf weitere Publikationen verwiesen, die zu der betreffenden Person in Verbindung stehen – in diesem Fall vermutlich auf die 1674 veröffentlichte Leichenpredigt für Anna Christina Gerstenberg, die 1673 verstorben ist.

`owl:sameAs` verbindet den GND-Datensatz mit dem entsprechenden ‚Identifizier‘ der Virtual International Authority Files (VIAF).

Auf der Grundlage der - in diesem Fall intellektuellen – Auswertung einiger exemplarischer Datensätze der Gemeinsamen Normdatei, die im Zusammenhang mit der hier als Beispiel verwendeten Leichenpredigt stehen, ergibt sich ein Personennetz (Abbildung 4.3).

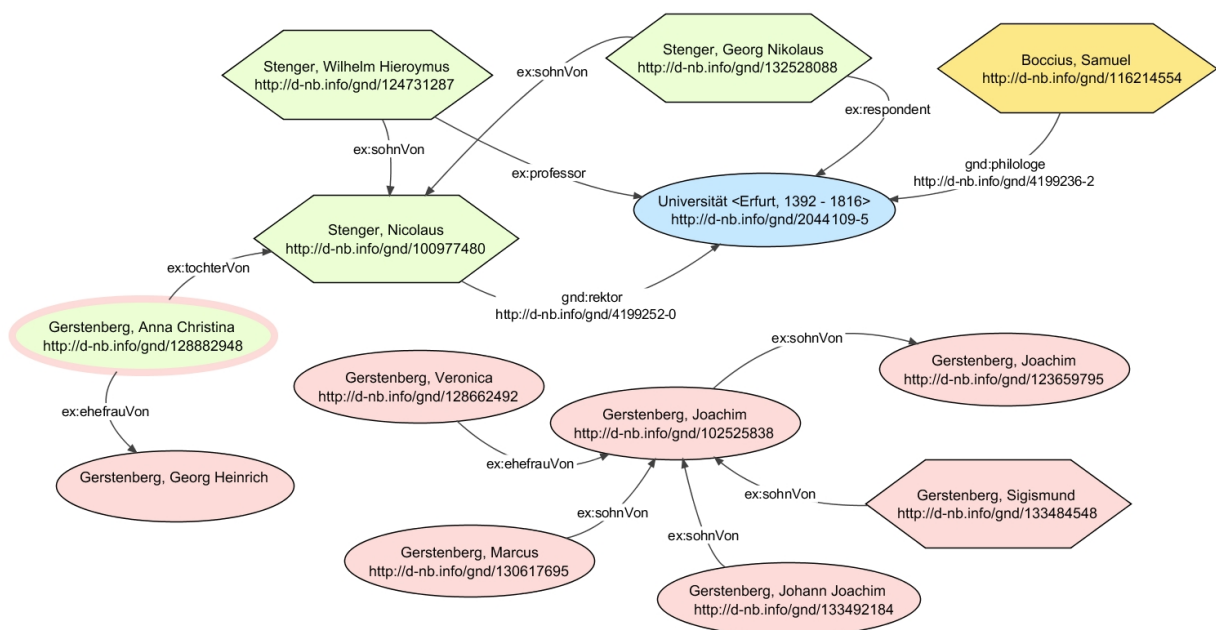


Abbildung 4.3 - Beziehungsnetz der mit der Leichenpredigt verbundenen Personen

Die Farben dienen in der Darstellung der besseren Unterscheidbarkeit der Familienangehörigkeit. Mit der eckigen Form sind die in den deskriptiven Metadaten genannten Beiträge der Leichenpredigt gekennzeichnet. Die Personen sind – mit einer Ausnahme, da in diesem Fall kein GND-Eintrag vorliegt – mit dem GND-‚Identifizier‘ versehen. Die ‚Properties‘ sind hier mit dem Namensraum `ex:` lediglich angedeutet. Im Fall einer Verlinkung der Person mit einer genormten Berufsbezeichnung (Rektor, Philologe), wurde sie als ‚Property‘ genutzt und beispielhaft als GND Namensraum bezeichnet (`gnd:`).

Die graphische Darstellung bietet Ansatzpunkte für eine Interpretation des Beziehungsgeflechts, das die Personen miteinander verbindet: einer der Beiträger, Sigismund Gerstenberg, ist der Sohn des Verstorbenen. Der Beiträger Nicolaus Stenger ist der Vater von Anna Christina Gerstenberg, der Ehefrau Georg Heinrich Gerstenbergs. Eine mögliche Interpretation dieser Ehe wäre, dass sie die Verschwägerung der Familien Stenger und Gerstenberg begründet, weshalb sich unter den Beiträgern neben Nicolaus Stenger auch dessen Söhne Georg Nicolaus und Wilhelm Hieronymus befinden. Ein weiterer Verbindungspunkt in diesem Geflecht scheint die Universität Erfurt zu sein. Neben den Beiträgern aus der Familie Stenger ist dort ein weiterer Beiträger, Samuel Boccius, tätig.

Diese punktuelle Auswertung einiger exemplarischer GND-Datensätze soll das Potential andeuten, das in einer automatischen Auswertung großer Datenmengen liegt. Die automatische Generierung und visuelle Darstellung der Beziehungsgeflechte der Personen, die im Zusammenhang mit den Leichenpredigten des 17. Jahrhunderts stehen, unterstützen das prosopographische Forschungsinteresse, das den Leichenpredigten bereits seit dem frühen 20. Jahrhundert entgegengebracht wird. Als Quellenreservoir sind die personalen Gelegenheitschriften im Allgemeinen und die Leichenpredigten auf Grund ihres Umfangs im Besonderen für die Geschichtswissenschaft unter vielfältigen Aspekten von großem Interesse:

Neben ihrer Bedeutung für Literaturgeschichte und Predigtforschung bieten die Leichenpredigten dem Sozial-, Kultur- und Universitätshistoriker eine Fülle statistisch auswertbarer Daten, dem Kunsthistoriker und Heraldiker reiches Material zur Ikonographie und Emblematik, dem Musikwissenschaftler eine ganze Reihe von nur hier überlieferten Trauerkompositionen; die teilweise sogar mit Abbildungen versehenen Krankheitsberichte schließlich liefern dem Medizinhistoriker wertvolle Informationen. (Staatsbibliothek, Personale Gelegenheitschriften)

Viele dieser Aspekte kommen nur mit einer Auswertung des Volltextes zum Tragen. Aus dem Grund wird derzeit an der Staatsbibliothek zu Berlin ein Pilotprojekt zur automatisierten Volltexterfassung größerer Funeralschriftenbestände durchgeführt. Dabei wird ein Wörterbuch mit Wörtern aufgebaut, die in Funeralschriften häufig Verwendung finden - ein Corpus, das zukünftig die Möglichkeiten für Named Entity Recognition im Zusammenhang mit dieser Textgattung verbessern könnte.⁵⁹

Neben der Auszeichnung der Volltexte auf der Grundlage bestehender Thesauri – zum Beispiel THELO und THEPRO der Forschungsstelle für Personalschriften Marburg und der Liste

⁵⁹ Auf die sprachliche Beschränktheit der NER-Software wurde an anderer Stelle bereits hingewiesen. Alte Sprachen, wie Latein oder Altgriechisch, und historische Formen bestehender Sprachen – wie im Fall des vorliegenden Beispiels Deutsch aus dem 17. Jahrhundert – werden dabei noch überhaupt nicht bedient. Dazu kommt, dass die Textcorpora, auf denen die NER-Technik fußt, sehr domänenspezifisch sind. Kleine, sehr spezielle Domänen, wie die Funeralschriften, finden besonders schwer Berücksichtigung. Siehe dazu auch Gerber et al. 2011 S.20: “Although named entity recognition techniques are improving, they are still limited to a relatively small number of domains (news, sports, business). Domain-specific approaches are able to achieve reasonable performance for limited numbers of entity types on well-circumscribed corpora. However they don’t easily permit customization or tailoring for domains other than the one for which they were designed.”

historischer Krankheitsbezeichnungen des Vereins für Computergenealogie – hat die Staatsbibliothek im Rahmen dieses Projekts ein Metadatenmodell für Funeralschriften entwickelt, das eine umfassende strukturelle und inhaltliche Erschließung vorsieht und damit die oben genannten Forschungsinteressen bedient (Staatsbibliothek, Funeralschriften).

Aus den deskriptiven Metadaten und Normdaten lassen sich neben den beruflichen und verwandtschaftlichen Beziehungen der beteiligten Personen auch Informationen darüber generieren, welche Prediger besonders aktiv waren und in welchem Radius sie gewirkt haben. Eine Software, die diese Informationen netzartig visualisieren kann, ist Protovis.⁶⁰ Protovis ist eine freie Java Script-basierte Open Source Software, die an der University of Stanford entwickelt wurde. Graphisch ermöglicht sie die farbige Kennzeichnung verschiedener Personengruppen. Die Größe der Knoten steht im Verhältnis zu der Anzahl der Beziehungen die in sie eingehen. (Abbildung 4.4).

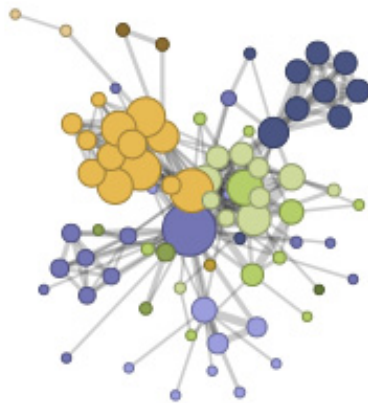


Abbildung 4.4 - Visualisierungssoftware (Protovis)

⁶⁰ Protovis wird beispielsweise im Rahmen des Projekts „Aufbau eines Repositoriums für biografische Daten historischer Personen des 19. Jahrhunderts“ – kurz: Personendaten-Repository – der Telota-Initiative der Berlin-Brandenburgischen Akademie der Wissenschaften eingesetzt. „Es erforscht anhand von Personeninformationen des „langen 19. Jahrhunderts“ (1789–1914), wie sich heterogene Datenbestände miteinander verbinden und präsentieren lassen. Ziel des Projektes ist die Bereitstellung eines dezentralen Softwaresystems, welches Lehr- und Forschungseinrichtungen, Archiven und Bibliotheken ermöglicht, biographische Informationen aus verschiedenen Beständen über einen gemeinsamen Zugang zu nutzen.“ (Telota)

4.4 Die explorative Suche

Ein häufiger Kritikpunkt an Bibliotheken und ihren (Online-)Katalogen besteht in der mangelnden Sichtbarkeit und Zugänglichkeit ihrer Bestände. Das Problem der (Un-)Sichtbarkeit hängt mit den Datenformaten zusammen, die zur Beschreibung der Bestände verwendet werden. Sie sind stark dem Paradigma des analogen Katalogs verhaftet und verhindern eine Verlinkung in WWW-Architekturen. Einer der Gründe für die schlechte Zugänglichkeit von Bibliotheksbeständen liegt darin, dass sich die Suchfunktionen der Bibliothekskataloge nicht intuitiv erschließen. OPACs bieten kaum die Möglichkeit des ‚Browsens‘. Sie setzen - so drückt es Karen Coyle etwas polemisch aus – einen Nutzer voraus, der schon weiß, welches Buch er sucht, wenn er die Bibliothek konsultiert (Coyle 2010).

Die Beschreibung der Daten mit RDF und ihre Vernetzung als Linked (Open) Data schafft – und das ist ein entscheidender Mehrwert – die Voraussetzung für eine grundsätzlich andere Art der Suche: die explorative Suche. Sie bietet dem Nutzer die Möglichkeit, an den vernetzten Informationen entlang zu wandern und den bisher geschlossenen Raum bibliothekarischer Daten zu verlassen, die Grenzen des Katalogs beständig zu überschreiten und dabei bibliothekarische Informationsressourcen mit Daten aus dem WWW zu kontextualisieren. Neben die Suche tritt das Entdecken, der Fund nach dem Serendipitäts-Prinzip (Sack 2010). Über die Vernetzung der Objekte werden Beziehungen zwischen ihnen sichtbar (Abbildung 4.5).

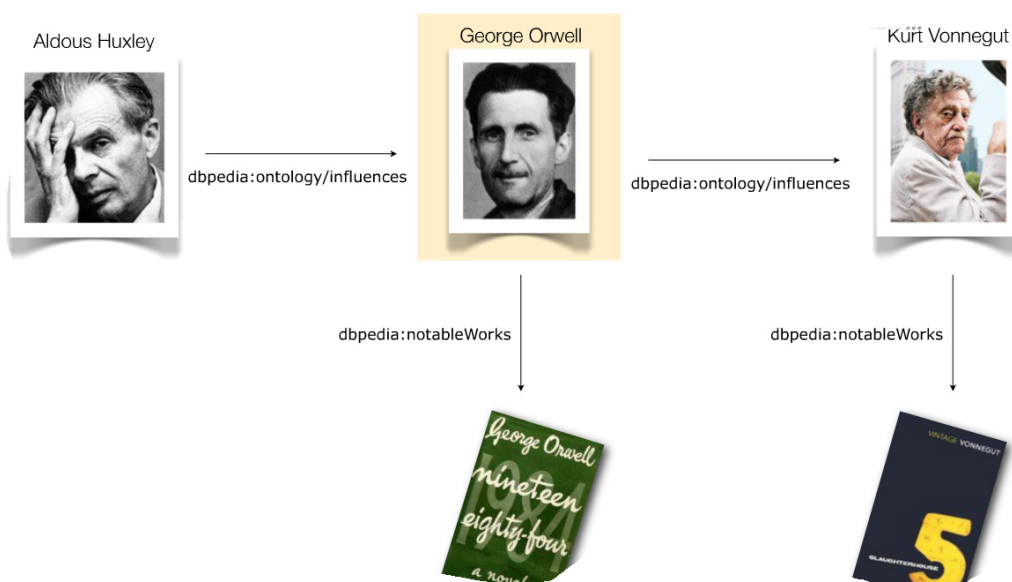


Abbildung 4.5 - Objektvernetzung in der DBpedia (Sack 2010)

Ein Problem, das den Nutzer bei seiner Bewegung in dem geöffneten Informationsraum der LOD-Cloud behindern kann, besteht in der Masse an ungewichtetem Wissen, dem er begegnet:

so finden sich etwa in der DBpedia über 600 Fakten zu dem Schriftsteller Aldous Huxley in RDF ausgedrückt, ohne jede Gewichtung, was Relevanz oder Reihenfolge betrifft. Hier bedarf es heuristischer Verfahren zur Relevanzbewertung von Linked Data Fakten, um die Informationen in einem ‚Interface‘ nutzerfreundlich aufzubereiten (Sack 2010). Ein solches ‚Interface‘ bietet das Europeana Portal mit der Implementierung des EDM. Mit der geplanten Integration von Europeana in die LOD Cloud öffnet Europeana seinen Informationsraum und erweitert damit die Kontextualisierungsoptionen für die repräsentierten Objekte.

Für das Beispiel der hier verwendeten Leichenpredigt bedeutet das neben Anreicherungen mit biographischen und geographischen Informationen auch die Vernetzung mit personellen Gelegenheitsschriften anderer Inhaltsanbieter. Die Universitäts- und Landesbibliothek Sachsen-Anhalt ist im Besitz der 1613 anlässlich der Hochzeit von Joachim und Veronica Gerstenberg gedruckten Gelegenheitsschrift und macht das Digitalisat unter anderem über Europeana zugänglich. Das gemeinsame Auffinden der Hochzeits- und Leichenpredigt Joachim Gerstenbergs ist angesichts der oben genannten Forschungsinteressen an personellen Gelegenheitsschriften ein Mehrwert, der das Potential der Entgrenzung von Informationsräumen für die explorative Suche verdeutlicht.

The screenshot shows the Europeana search interface. At the top, there are navigation links: 'My Europeana', 'About us', 'Communities', 'Partners', 'ThoughtLab', and a language selector. A search bar contains the text 'joachim gerstenberg' with a 'Search' button and a 'Advanced search' link. Below the search bar, it says 'Matches for: joachim gerstenberg'. The main content area is titled 'Item details' and shows a thumbnail of a document cover on the left. To the right of the thumbnail, the title is 'Epithalamia In Honorem Nuptiarum ... Dn. Joachimi Gerstenberck Budstadiensis Sponsi, Cum ... Veronica, ... Viri, Dn Johannis Martini ... Filia : 15. Novemb. A. 1613. celebratarum, Scripta a Fautoribus & Amicis'. Below the title, there are fields for 'Title', 'Date: 1613', 'Creator', 'Description', 'Language: la', 'Format: [4] Bl.; 4"', 'Rights: Universitäts- und Landesbibliothek Sachsen-Anhalt, Halle (Saale)', and 'Data provider: Universitäts- und Landesbibliothek Sachsen-Anhalt, Halle (Saale)'. On the left side, there is a 'Related content' section with three items and an 'Actions' section with buttons for 'Add a tag', 'Save to My Europeana', 'Embed', 'Translate details', 'Login', and 'Register'.

Abbildung 4.6 - Hochzeitsschrift Joachim und Veronica Gerstenberg (Europeana)

5. Fazit

“Stop hugging your data” lautet ein Ausspruch von Tim Berners-Lee. Damit wirbt er für die Verwendung offener Standards bei der Bereitstellung von Informationsressourcen im World Wide Web. Der Einsatz persistenter URIs und die Beschreibung von Daten im Netz mit RDF öffnet die Möglichkeit für die Verlinkung mit anderen Informationsressourcen – und den Weg zu einem „Web of Data“. Bibliothekarische Einrichtungen müssen sich, wollen sie in ihrer Rolle als Informationsvermittler künftig nicht an Relevanz verlieren, mit ihren Beständen in die digitale Informationslandschaft integrieren. Die von Bibliotheken gepflegten ‚Datensilos‘ verhindern die Informationsvernetzung und die Integration in eine WWW-basierte Informationsarchitektur. Eine Reorganisation ihrer Informationsressourcen ist daher essentiell. Einen Weg für die Integration bibliothekarischer Bestände in das WWW bietet die Beschreibung der Daten mit dem Europeana Data Model. Es basiert auf offenen, vom W3C empfohlenen Standards und schafft für die an Europeana beteiligten Institutionen einen gemeinsamen Rahmen, der es erlaubt, über die Grenzen einzelner Domänen hinweg, Informationsressourcen zu vernetzen und neu zu kontextualisieren. Mit der Verwendung semantischer Technologien geschieht die Kontextualisierung nicht auf rein begrifflicher Ebene, sondern auf der Grundlage von Wissensmodellen und Ontologien.

Die Beschreibung bibliothekarischer Informationsressourcen mit dem graphenbasierten Europeana Data Model schafft – so das Ergebnis der vorliegenden Arbeit – Mehrwerte gegenüber einer Beschreibung mit den in Bibliotheken verwendeten XML-Formaten METS und MODS. Zum einen ermöglicht die Verwendung des Resource Description Frameworks den Einsatz der Anfragesprache SPARQL, die in ihrem Potential bei der Bildung von Suchsets anderen in Bibliothekskatalogen verwendeten Anfragesprachen überlegen ist. Zum anderen spielt das Integrationspotential des EDM-Ansatzes bei der Schaffung von Mehrwerten eine wesentliche Rolle: die Beschreibung der Objekte mit RDF erzeugt ein Netzwerk von ‚Resources‘. Mit der Nutzung des Proxy-Ansatzes aus dem Modell OAI-ORE können die Beschreibungen verschiedener Inhaltsanbieter zu einem Objekt zusammengeführt werden. Für das in der vorliegenden Arbeit untersuchte Beispiel der Leichenpredigt „Christiani Vita Et Corona“ ergeben sich aus der Zusammenführung der Beschreibungen zweier Institutionen deutliche Vorteile für die Suchoptionen. Die Objektrepräsentation der Leichenpredigt kann im Informationsraum Europeana mit Normdaten verlinkt und anderen, mit ihr in Zusammenhang stehenden Objektrepräsentationen kontextualisiert werden.

Das EDM schafft die Voraussetzungen für die Öffnung von Informationsräumen. Mit der Organisation der Daten in Form von Netzgraphen ermöglicht es eine neue Art der Suche:

neben die Suche des bereits Bekannten tritt die explorative Suche, die Entdeckung neuer, mit dem bekannten Objekt vernetzter Informationen.

Die geplante Integration von Europeana in die Linking Open Data Cloud birgt das Potential einer Erweiterung sowohl des Informationsraums Europeana als auch der Kontextualisierungsoptionen für die über Europeana zugänglichen Informationsobjekte.

Für bibliothekarische Einrichtungen stellt die Partizipation an der Informationsvernetzung mit Linked Open Data eine Chance dar, sich mit ihrem Wissen und ihren Beständen aktiv am Aufbau neuer Informationsarchitekturen zu beteiligen und so im WWW sichtbar zu werden.

Abkürzungsverzeichnis

API	Application Programming Interface
DCMI	Dublin Core Metadata Initiative
DDC	Dewey Decimal Classification
DFG	Deutsche Forschungsgemeinschaft
EDM	Europeana Data Model
Ens	Europeana Namespace
ESE	Europeana Semantic Elements
GBV	Gemeinsamer Bibliotheksverbund
GND	Gemeinsame Normdatei
HTML	Hypertext Markup Language
JISC	Joint Information Systems Committee
LLD XG	W3C Library Linked Data Incubator Group
LOD	Linking/ Linked Open Data
MAB	Maschinelles Austauschformat für Bibliotheken
MARC	Machine-Readable Catalogue
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OPAC	Online Public Access Catalogue
OWL	Web Ontology Language
PND	Personennamendatei
PPN	PICA Production Number
PURL	Persistent Uniform Resource Locator
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SKOS	Simple Knowledge Organization System
SPARQL	Simple Protocol And RDF Query Language
SWD	Schlagwortnormdatei
TGN	Getty Thesaurus of Geographic Names
URI	Uniform Resource Identifier
VIAF	Virtual International Authority Files
VD 17	Verzeichnis der Drucke des 17. Jahrhunderts
WWW	World Wide Web
W3C	World Wide Web Consortium
XML	Extensible Markup Language

Bibliographie

(Barbera et al. 2008) Barbera, Michele/ Nucci, Michele/ Hahn, Daniel, A Semantic Web Powered Distributed Digital Library System, in: Chan, Leslie/ Mornati, Susanna: Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing, Toronto 2008

http://elpub.scix.net/cgi-bin/works/Show?id=130_elpub2008&sort=DEFAULT&search=barbera&hits=3

(Becker et al. 2010) Becker, Hans-Georg/ Hoepfner, Iris/ Kirsch, Christian, Universitätsbibliothek Dortmund 2.0, in: Bergmann, Julia/ Danowski, Patrick (Hrsg.), Handbuch Bibliothek 2.0, Berlin/ New York, 2010

(Bérard 2011) Bérard, Raymond, Free Library Data ? in: Liber Quarterly 20, no.3/4, 2011, <http://liber.library.uu.nl/publish/articles/000512/article.pdf>

(Bermès et al. 2011) Bermès, Emmanuelle/ Eckert, Kai/ Harper, Corey et al., Draft for the final report intended to capture the high level benefits of Library Linked Data, W3C 2011, http://www.w3.org/2005/Incubator/ld/wiki/Draft_Benefits

(Berners-Lee 2006) Berners-Lee, Tim, Linked Data, W3C 2006, <http://www.w3.org/DesignIssues/LinkedData.html>

(Berners-Lee et al. 2001) Berners-Lee, Tim/ Hendler, James/ Lassila, Ora, The semantic web, in: Scientific American 284, No. 5, 2001

(Biblios.net) <http://biblios.net/>

(Bizer et al. 2009) Bizer, Christian/ Berners-Lee, Tim/ Heath, Tom, Linked data – the story so far, in: International Journal on Semantic Web and Information Systems 5, no. 3, 2009

(Blumauer et al. 2006) Blumauer, Andreas/ Pellegrini, Tassilo, Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen, in: dies. (Hrsg.), Semantic Web, Wege zur vernetzten Wissensgesellschaft, Berlin/ Heidelberg 2006

(Blumauer et al. 2009), Blumauer, Andreas/ Pellegrini Tassilo, Semantic Web Revisited - Eine kurze Einführung in das Social Semantic Web, in: dies. (Hrsg.), Social Semantic Web : Web 2.0 - was nun? Berlin [u.a.] 2009

(Brickley et al. 2004) Brickley, Dan/ Guha, R.V., RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 2004, <http://www.w3.org/TR/rdf-schema/>

(CIDOC CRM) <http://www.cidoc-crm.org/>

(Concordia et al. 2010) Concordia, Cesare/ Gradmann, Stefan/ Siebinga, Sjoerd, Not just another portal, not just another digital library: A portrait of Europeana as an application program interface, Library Associations and Institutions 36, no. 1, 2010

(Coyle 2010) Coyle, Karen, Yes We Can! Libraries and the Semantic Web, Vortrag, Köln 2010, http://swib.org/swib10/vortraege/swib10_coyle.ppt

(Data.gov) <http://www.data.gov/>

(Data.gov.uk) <http://data.gov.uk/>

(DBpedia) <http://dbpedia.org/About>

(Definition 2010) Definition of the Europeana Data Model elements, Version 5.2, 2010

(Dekkers et al. 2009) Dekkers, Makx/ Gradmann, Stefan/ Meghini Carlo, Europeana Outline Functional Specification. For development of an operational European Digital Library, Version 1.7, 2009

(Description of Work 2009) Description of Work, EuropeanaConnect, Version 30.4.2009

(DFG-Viewer) <http://dfg-viewer.de/>

(DNB GND) <http://www.d-nb.de/standardisierung/normdateien/gnd.htm>

(DNB PND) <http://www.d-nb.de/standardisierung/normdateien/pnd.htm>

(DNB SWD) <http://www.d-nb.de/standardisierung/normdateien/swd.htm>

(Doerr et al. 2010) Doerr, Martin/ Gradmann, Stefan/ Henniecke, Steffen et al. (2010) The Europeana Data Model (EDM), in: World Library and Information Congress: 76th IFLA General Conference and Assembly, <http://www.ifla.org/files/hq/papers/ifla76/149-doerr-en.pdf>

(Doerr et al. 2011) Doerr, Martin/ Gradmann, Stefan/ Henniecke, Steffen et al., Europeana Data Model, Vortrag, Mannheim 2011, <http://www.kim-forum.org/material/schulungen/KIM-DINI-Kickoff-Workshop.htm>

(Dublin Core) <http://dublincore.org/documents/dces/>

(Dublin Core Metadata Terms) <http://dublincore.org/documents/dcmi-terms/>

(DuCharme 2008) DuCharme, Bob, (semantic web) - semantics = linked data?, in: bobdc-blog 2008, <http://www.snee.com/bobdc.blog/2008/10/semantic-web-semantics-linked.html>

(Eckert 2011) Eckert, Kai, Linked Open Data an der UB Mannheim, in: Neues aus der UB Mannheim 2011, <http://blog.bib.uni-mannheim.de/Aktuelles/?p=3172>

(Europeana) Europeana, <http://www.europeana.eu/portal/>

(Europeana Thoughtlab) http://www.europeana.eu/portal/thoughtlab_semanticsearching.html.

(Europeana Semantic Elements 2010) Europeana Semantic Elements specifications Version 3.2.2, 2010, http://www.version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602

(Eustis 2009), Eustis, Jennifer, NGC4LIB on Tim Berners-Lee and the Semantic Web, <http://celeripedean.wordpress.com/2009/11/09/ngc4lib-on-tim-berners-lee-and-the-semantic-web/>

(Faruqui et al. 2010) Faruqui, Manaal/ Padó, Sebastian, Training and Evaluating a German Named Entity Recognizer with Semantic Generalization, 2010, http://www.nlpado.de/~sebastian/pub/papers/konvens10_faruqui.pdf

(Feigenbaum et al. 2008), Feigenbaum, Lee et al., Mein Computer versteht mich – allmählich, in: Spektrum der Wissenschaft, no. 11, 2008

(FRBR) <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>

(Gantz et al. 2008) Gantz, John F./ Chute, Christopher, The diverse and exploding digital universe: an updated forecast of worldwide information growth trough 2011, International Data Corporation, Framingham 2008

(Gerber et al. 2011) Gerber, Anna/ Gao, Lianli/ Hunter, Jane, A Scoping Study of (Who, What, When, Where) Semantic Tagging Services, The University of Queensland 2011, <http://itee.uq.edu.au/~eresearch/projects/ands/W4SemanticTagging-report-2011-02.pdf>

(Goobi) <http://www.goobi.org/>

(Google Books) <http://books.google.de/>

(Google Scholar) <http://scholar.google.de/schhp?hl=de>

(Gradmann 2009 A) Gradmann, Stefan, Signal, Information, Zeichen: zu den Bedingungen des Verstehens in semantischen Netzen, Antrittsvorlesung 28. Oktober 2008, in: Öffentliche Vorlesungen / Humboldt-Universität zu Berlin 2009. <http://edoc.hu-berlin.de/humboldt-vl/157>.

(Gradmann 2009 B) Gradmann, Stefan, Warum sollten sich kulturelle Einrichtungen mit dem Semantic Web beschäftigen?, Vortrag, Köln 2009, http://www.swib09.de/vortraege/20091124_gradmann.pdf

(Gradmann 2010), Gradmann, Stefan, Linked Open Europeana: Das Europeana Data Model (EDM), Vortrag, Köln 2010, http://swib.org/swib10/vortraege/swib10_gradmann.pdf

(Gradmann et al. 2009) Gradmann, Stefan/ Olensky, Marlies, Linked Data und Semantic Web-basierte Funktionalität in Europeana, Vortrag, Köln 2009, http://www.swib09.de/vortraege/20091125_gradmann_olensky.pdf

(Gradmann et al. 2010), Gradmann, Stefan/ Horstmann, Wolfram, OAI Object Reuse and Exchange, in: The KIM Technology Watch Report 2010, <http://metadaten-twr.org/2010/12/14/oai-ore>

(van Harmelen 2006) van Harmelen, Frank, Semantic Web Research anno 2006: main streams, popular fallacies, current status and future challenges, in: Cooperative Information Agents X. 10th International Workshop, CIA 2006 Edinburgh, Berlin/Heidelberg 2006

(Hennicke et al. 2011) Hennicke, Steffen/ Olensky, Marlies/ de Boer, Viktor/ Isaac, Antoine/ Wielemaker, Jan, A data model for cross-domain data representation. The "Europeana Data Model" in the case of archival and museum data In: Schriften zur Informationswissenschaft 58, Proceedings des 12. Internationalen Symposiums der Informationswissenschaft, 2011

(Isaac 2010 A) Isaac, Antoine, Europeana Data Model Primer, Europeana v1.0, 2010

(Isaac 2010 B) Isaac, Antoine, W3C Library Linked Data Incubator Group (LLD XG), Vortrag, Köln 2010, http://swib.org/swib10/programm_29-november.html

(Isaac et al. 2009) Isaac, Antoine/ Summers, Ed, SKOS Simple Knowledge Organization Primer, W3C 2009, <http://www.w3.org/TR/skos-primer/>

(JISC) <http://obd.jisc.ac.uk/>

(Klyne et al. 2004) Klyne, Graham/ Carroll, Jeremy J., Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation 2004 <http://www.w3.org/TR/2004/REC-rdfconcepts-20040210/>

(Lagoze et al. 2008 A) Lagoze, Carl/ Van de Sompel, Herbert, ORE User Guide - Primer, 2008, <http://www.openarchives.org/ore/1.0/primer>

(Lagoze et al. 2008 B) Lagoze, Carl/ Van de Sompel, Herbert/ Johnston, Pete/ Nelson, Michael/ Sanderson, Robert, et al., ORE Specification - Abstract Data Model, 2008, <http://www.openarchives.org/ore/1.0/datamodel#Proxy>

(LOD) <http://richard.cyganiak.de/2007/10/lod/>

(MAB) <http://www.d-nb.de/standardisierung/formate/mab.htm>

(MARC) <http://www.loc.gov/marc/>

(Mashup Australia) <http://mashupaustralia.org/>

(METS) <http://www.loc.gov/standards/mets/>

(Miles et al. 2009) Miles, Alistair/ Bechhofer, Sean, SKOS Simple Knowledge Organization System Namespace Document, W3C 2009, <http://www.w3.org/TR/skos-reference/skos.html>

(MODS) <http://www.loc.gov/standards/mods/>

(Neffe 2009) Neffe, Jürgen, Es war einmal. Die Ära des gedruckten Buches geht zu Ende. Kein Grund zur Trauer, in: Die Zeit, Nr. 18 – 23 vom 23.04.2009, <http://www.zeit.de/2009/18/L-Buch/komplettansicht>

(OAI 2008) Open Archives Initiative Object Reuse and Exchange, ORE Specifications and User Guides, 2008, <http://www.openarchives.org/ore/1.0/toc.html>

(OCLC) <http://www.oclc.org/de/de/global/default.htm>

(OCLC 2005) Perception of Library and Information Resources, OCLC 2005, <http://www.oclc.org/reports/2005perceptions.htm>

(Open Knowledge Foundation) <http://okfn.org/>

(Open Library) <http://openlibrary.org/>

(Otte et al. 2008) Otte, Karsten/ Levergood, Barbara, zvd MODS Anwendungsprofil Version 1.0, Göttingen 2008

(Pédauque 2003) Pédauque, Roger T., Document: Form, Sign and Medium: As reformulated for Electronic Documents, 2003, http://halshs.archives-ouvertes.fr/docs/00/06/22/28/PDF/sic_00000594.pdf

(Pohl 2011) Pohl, Adrian, Linked Data und die Bibliothekswelt, 2011, http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2011_linked-data_ODOK.pdf

(Pohl et al. 2010) Pohl, Adrian/ Ostrowski, Felix, Zur Konzeption und Implementierung einer Infrastruktur für freie bibliografische Daten, in: Ockenfeld, Marlies (Hrsg.), Semantic Web & Linked Data, Elemente zukünftiger Informationsinfrastrukturen, Tagungsband der 1. DGI-Konferenz/ 62. Jahrestagung der DGI 2010, Frankfurt a.M. 2010

(Protovis) Protovis, <http://vis.stanford.edu/protovis/>

(Prud'hommeaux et al. 2008) Prud'hommeaux, Eric/ Seaborne, Andy, SPARQL Query Language for RDF, W3C Recommendation 2008, <http://www.w3.org/TR/rdf-sparql-query/>

(Sack 2010) Sack, Harald, Vom Suchen und Finden in multimedialen Archiven der nächsten Generation, Vortrag, München 2010, <http://www.slideshare.net/lysander07/explorative-und-semantische-suche-mit-mediaglobe>

(Schlönvoigt 2009) Schlönvoigt, Steffen, Semantic Web Technologien, Konstanz 2009, <http://www.slideshare.net/seso81/7-sprachen-des-semantic-web-sparql>

(Skyriver) <http://theskyriver.com/>

(Staatsbibliothek) <http://digital.staatsbibliothek-berlin.de/dms/>

(Staatsbibliothek, Funeralschriften) <http://staatsbibliothek-berlin.de/abteilung-historische-drucke/aufgaben-profil/projekte/funeralschriften.html>

(Staatsbibliothek, Personale Gelegenheitsschriften) <http://staatsbibliothek-berlin.de/abteilung-historische-drucke/sammlungen/bestaende/personale-gelegenheitsschriften.html>

(Svensson 2011) Svensson, Lars, Linked Data aus der Kulturdomäne – Sprungbrett in das semantische Web?, Vortrag auf der 2. I-Science Tagung der FH Potsdam im März 2011

(TGN) <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>

(THELO) <http://www.personalschriften.de/datenbanken/thelo.html>

(THEPRO) <http://www.personalschriften.de/datenbanken/thepro.html>

(Telota) <http://pdr.bbaw.de/>

(Temis) <http://www.temis.com>

(VD 17) <http://www.vd17.de/>

(VIAF) <http://viaf.org/>

(Voß 2009) Voß, Jakob, Einführung in das Semantic Web, Vortrag, Köln 2009, http://www.swib09.de/vortraege/20091124_voss.pdf

(W3C 2010) W3C. "Semantic Web," 2010. <http://www.w3.org/standards/semanticweb/>

(W3C Semantic Web) <http://www.w3.org/2001/sw/>

(zvdd/DFG-Viewer METS-Profil 2009) zvdd/DFG-Viewer METS-Profil 2009 – Version 2.0, 2009, http://dfgviewer.de/fileadmin/groups/dfgviewer/METS_Anwendungsprofil_2.0.pdf

Letzter Zugriff auf Hyperlinks am 24.05.2011

Abbildungsverzeichnis

Abbildung 1.1 - Schichtenmodell der Objektrepräsentationen in Europeana (aus: Doerr et al. 2011)	16
Abbildung 1.2 - RDF-, Triple'	17
Abbildung 1.3 - Der Gebrauch von ‚Literals‘ und URIs im Zusammenhang mit dem RDF.....	18
Abbildung 1.4 - Vier Klassen in der Datenmodellierung mit dem EDM	21
Tabelle 1 - Mapping Pica – ZVDD-MODS (aus: Otte et al. 2008 S. 15).....	31
Abbildung 3.1 - Objektrepräsentation in EDM.....	34
Abbildung 3.2 - ‚Christiani Vita Et Corona‘ Vorderdeckel.....	36
Abbildung 3.3 - ‚Christiani Vita Et Corona‘ Titelblatt	37
Abbildung 3.4 - ore:Proxy, deskriptive Metadaten.....	38
Abbildung 3.5 - Ereigniszentrierte Objektbeschreibung.....	40
Abbildung 3.6 - Beschreibung der hierarchischen Struktur.....	41
Abbildung 3.7 - Zwei Sichten auf ein Objekt	44
Abbildung 3.8 - Eindeutige Referenzierung der Personen mit der PND-ID.....	45
Abbildung 4.1 - Sucheinstieg Digitalisierte Sammlungen (Staatsbibliothek)	47
Abbildung 4.2 - Visualisierte Darstellung der Suchergebnisse auf einer ‚Timeline‘ (Europeana)	49
Tabelle 2 - Datensatz der PND zu Anna Christina Gerstenberg (aus: Gemeinsame Normdatei der Deutschen Nationalbibliothek).....	50
Abbildung 4.3 - Beziehungsnetz der mit der Leichenpredigt verbundenen Personen.....	51
Abbildung 4.4 - Visualisierungssoftware (Protovis).....	53
Abbildung 4.5 - Objektvernetzung in der DBpedia (Sack 2010).....	54
Abbildung 4.6 - Hochzeitsschrift Joachim und Veronica Gerstenberg (Europeana).....	55

Tabellenverzeichnis

Tabelle 1 - Mapping Pica – ZVDD-MODS (aus: Otte et al. 2008 S. 15).....	31
Tabelle 2 - Datensatz der PND zu Anna Christina Gerstenberg (aus: Gemeinsame Normdatei der Deutschen Nationalbibliothek).....	50