

Maike Lins, Hans-Georg Becker

Open Data und Linked Data in einem Informationssystem für die Archäologie

Einleitung

Im Rahmen des Programms „Themenorientierte Informationsnetze“ fördert die DFG seit dem Jahr 2008 das Projekt „*ArcheoInf* – Informationsnetz zur umfassenden Bereitstellung von digitalen Forschungsdaten archäologischer Feldprojekte (Ausgrabungen, Feldsurveys)“. Neben den Universitätsbibliotheken Dortmund und Bochum, von deren Arbeit dieser Beitrag berichten wird, arbeiten das Archäologische Institut der Georg-August-Universität Göttingen, der Lehrstuhl für Software-Technologie der Technischen Universität Dortmund sowie der Fachbereich Geoinformatik/Geodäsie der Hochschule Bochum in diesem Projekt an der Entwicklung eines Informationssystems für die Archäologie.¹

ArcheoInf will der Silohaltung archäologischer Primärdaten entgegenwirken und zielt auf eine Verfügbarmachung wissenschaftlicher Forschungsdaten im Netz – Ziel ist die Ermöglichung des Zugriffs auf Karten, Bilder, Texte, schlicht alle zu den beteiligten Projekten verfügbaren Informationen – über eine einzige Oberfläche und von jedem Ort der Welt erreichbar. Doch *ArcheoInf* geht noch einen Schritt weiter: Es sollen hier nicht die Projekte isoliert durchsuchbar sein, sondern in einer Weise miteinander verknüpft werden, dass projektübergreifend Informationen zu einem gesuchten Thema gefunden und verbunden werden können.

Bei der eben beschriebenen Zielsetzung des Projektes stehen somit *Open Access* und *Open Data* für Publikationen und Forschungsdaten im Vordergrund. Der vorliegende Beitrag gibt zunächst einen Überblick über die archäologische Arbeitsweise. Anschließend wird gezeigt, dass *Open Data* nicht nur als *Output* eines solchen Projektes entsteht, sondern dass es auch auf der Seite des *Inputs* seine Berechtigung haben kann. Schließlich wird noch gezeigt, wie die Forschungsdaten in diesem Projekt mit den vorhandenen bibliographischen Informationen verknüpft werden.

1 Bubke/Lins: Zentrum für archäologische Feldforschungsdaten (2012). 122.



Was haben Archäologen mit Open Data zu tun?

Neue digitale Möglichkeiten und Technologien eröffnen Wissenschaftlern und Forschenden aller Fachrichtungen früher ungeahnte Wege, mit wissenschaftlichen Daten umzugehen. Um die Nutzung der Daten, ihre Qualität und Zugänglichkeit langfristig zu sichern, ist die Entwicklung von dringend benötigten Lösungen² seit einigen Jahren in vollem Gange. Trotzdem existieren in vielen Fachbereichen noch keine Patentrezepte zur Verfügbarmachung und langfristigen Erhaltung wissenschaftlicher Primärdaten.

Sind Forschungsdaten einmal erhoben – das Folgende gilt fächerübergreifend – werden diese ausgewertet und die resultierenden Forschungsergebnisse mit der Publikation öffentlich gemacht. In der Regel folgt darauf die Ablage der Daten – und dort, wo sie deponiert werden, bleiben die Daten, in ihrem ursprünglichen Format und anschließend vielleicht über Jahre ungenutzt.³

Für wen sind diese Daten bei einer solchen Vorgehensweise verfügbar? In der Regel ist dies nur ein sehr eng begrenzter Kreis, manchmal lediglich beschränkt auf die Wissenschaftler, die die Daten erhoben haben. Ein Problem kann auch die Lesbarkeit älterer Daten darstellen, die, je nach Alter, eingesetzten Speichermedien und ursprünglich verwendetem Datenformat, nicht immer gewährleistet werden kann. Als nur ein Beispielpiel mag hier der Fall der NASA dienen, die den Verlust der Aufzeichnungen aus mehreren Jahrzehnte der US-Raumfahrt zu beklagen hat: die zur Speicherung der Daten verwendeten Magnetbänder sind mit modernen Geräten nicht mehr zu lesen oder wurden bei dem Versuch, dies zu tun, mitsamt den enthaltenen Daten zum Teil unwiederherstellbar zerstört.⁴ Welche Arten von wissenschaftlichen Primärdaten sind in der archäologischen Forschung zu erwarten? Zwei wichtige Methoden in der Arbeit der Archäologen sind der Survey und die Ausgrabung.

Der Survey oder die Flächenerkundung dient der Untersuchung der Dimensionen, des Typs und der Anlage eines potenziellen Fundplatzes. Die Mitarbeiter eines Surveys untersuchen die fragliche Fläche systematisch, wobei sie die Bodenoberfläche nach besonderen Merkmalen absuchen und Funde registrieren und sammeln. Von großer Bedeutung ist die genaue Vermessung des Geländes und anschließende Kartierung. Bei einem Survey ergibt sich eine große Zahl von Daten: Funde, Vermerke über Bodenmerkmale, Vermessungsdaten, die zunächst

² Vgl. Deutsche Initiative für Netzwerkinformation e. V.: Positionspapier Forschungsdaten (2009), 5.

³ Bubke/Lins: Zentrum für archäologische Feldforschungsdaten (2012) 122.

⁴ Oßwald/Scheffel/Neuroth: Langzeitarchivierung von Forschungsdaten (2012), 13. Hammer-schmitt: Die Uhr läuft (2002).

lediglich aufgenommen, aber erst an einem späteren Punkt der Arbeit ausgewertet werden.

Die Ausgrabung, die nicht selten auf Basis der Erkenntnisse, die durch einen Survey gewonnen werden konnten, durchgeführt wird, spielt in der Feldarbeit der Archäologen eine zentrale Rolle – sie ist die Hauptmethode, um in der Archäologie Daten zu erheben.⁵ Das Freilegen der von Erd- oder Steinschichten überdeckten Funde und Befunde führt notwendigerweise zu einer Zerstörung des Befundes und muss deshalb mit äußerster Genauigkeit dokumentiert werden. Es fallen Daten in Form von Zeichnungen, Fotos, Messdaten über Größe eines Fundstückes, Lage und Ausrichtung sowie erste Beschreibungen oder sogar Interpretationen eines Stückes in Textform an. Neben diese Daten kommen weiterhin Karten und nicht zuletzt auch bibliographische Metadaten hinzu.

Die forschenden Archäologen organisieren diese Forschungsdaten in Datenbanken unterschiedlichster Ausprägung, die in der Regel nach Abschluss der Forschungen und nach der Publikation der resultierenden Ergebnisse bei den leitenden Mitarbeitern verbleiben. Andere Interessierte haben üblicherweise außerhalb der Publikationen keine Möglichkeit, auf diese Daten zuzugreifen oder die Forschungsergebnisse gar zu überprüfen.

Einer Übertragung oder Auswertung der Daten von anderer Seite werden zudem auch deshalb Steine in den Weg gelegt, da in den Projekten kaum Standards Anwendung finden – dies hat der Verlauf des Projekts *ArcheoInf* deutlich bestätigt.

Auch informatische Anforderungen finden keine Berücksichtigung: in den von den Archäologen selbst erstellten Datenbanken sind die Daten oft widersprüchlich oder uneindeutig, nicht mit anderen Systemen austauschbar oder stehen unverbunden nebeneinander.⁶

Die Infrastruktur von *ArcheoInf*

Die Grundlage des Projekts *ArcheoInf* bildeten von Beginn an die von den Partnerprojekten gelieferten Datenbanken.

Die Arbeit des Projekts gliedert sich in verschiedene Gruppen, die jeweils an Einzelementen des Gesamtsystems *ArcheoInf* arbeiten, um diese dann in einem letzten Schritt zu integrieren und als Ganzes nutzbar zu machen. Die Mitarbeitergruppen bearbeiten die Teile Thesaurus und Ontologie, Mediator, Geoserver und Dokumentenrepositorium.

⁵ Renfrew/Bahn: Basiswissen Archäologie (2009). 292. 83.

⁶ Bubke/Lins: Zentrum für archäologische Feldforschungsdaten (2012). 124.

Wesentlicher Bestandteil des *ArcheoInf*-Informationssystems ist der Mediator. Dieser erfüllt als Knotenpunkt die Schnittstellenrolle zu den übrigen Komponenten: Geoinformationssystem, Forschungsdaten- und Dokumentenrepositorium. Abbildung 1 veranschaulicht diese Architektur.

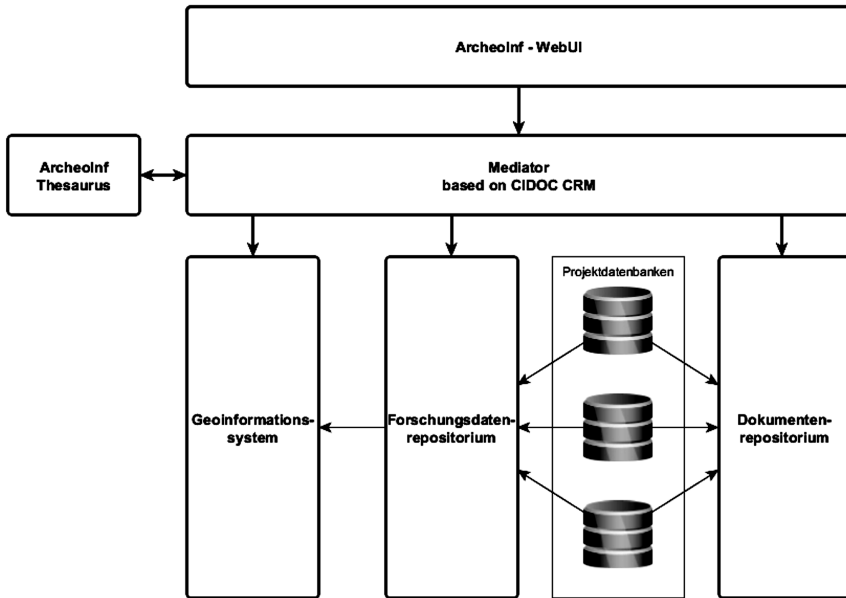


Abbildung 1: Architektur des *ArcheoInf*-Informationssystem

Die Kommunikation des Mediators mit den darunter liegenden Strukturen ermöglicht ein Netz aus *CIDOC CRM*-kompatiblen Ontologien. Unterstützt wird diese Struktur durch einen als Bindeglied zwischen den Erschließungsbegriffen in den Projektdatenbanken erstellten Thesaurus. Letzterer wurde von Beginn an in *SKOS* formuliert und so *Semantic Web*-fähig entwickelt.

Die ursprüngliche Idee, in *ArcheoInf* sämtliche Quellen, also die Projektdatenbanken, auf das *CIDOC CRM* abzubilden und somit eine semantische Interoperabilität im Sinne des *Linked Data* zu erreichen, konnte im Projekt bisher aufgrund der schlechten Datenlage bereits abgeschlossener Projekte nicht zufriedenstellend erreicht werden. Es folgte somit die Entscheidung für eine „Datenbergung“ und Aufbereitung der Daten sowie die Speicherung in einem separaten Forschungsdatenrepositorium.

Ein wichtiger Teil der in den Projektdatenbanken enthaltenen Daten sind bibliographische Angaben. Diese Angaben stehen in der Regel in engster Beziehung

zu den in den jeweiligen Projekten erforschten Fragestellungen und sind teilweise innerhalb der gepflegten Datenbanken mit den entsprechenden aufgenommenen Fundstücken verknüpft. Die hier zu findenden bibliographischen Metadaten reichen von der Angabe ganz allgemeiner thematisch relevanter Fachliteratur bis zu Texten, aus denen detailliert nachgewiesene Stellen- und Abbildungsverweise vorliegen. Im Projekt *ArcheoInf* wird ein Dokumentenrepositorium aufgebaut, in dem diese Daten und – wo möglich auch Volltexte – bereitgestellt werden.

Das Dokumentenrepositorium soll als integraler Bestandteil des *ArcheoInf*-Gesamtsystems die Arbeit der forschenden Archäologen und jedes Recherchierenden erleichtern, indem hier Literaturnachweise bereitgehalten werden, die sehr speziell auf die eingebunden Forschungsdaten ausgerichtet sind und so äußerst präzise und treffende Suchergebnisse ermöglichen sollen. Die Tatsache, dass durch die Archäologen vielfach auch Literatur aufgenommen wird, die dem Bereich der grauen oder nur regional verfügbaren Literatur zuzuordnen ist und durch eine Suche in anderen Rechercheinstrumenten nicht problemlos gefunden werden kann, bereichert die Inhalte des Repositoriums in besonderem Maße.

Die dem Mediator zugrunde liegenden Strukturen wurden mit Technologien des *Semantic Web* – sprich *Triple Stores* – realisiert. Da eine Recherche innerhalb dieser *Stores* derzeit weder performant genug noch mittels *SPARQL* oder ähnlichen Abfragesprachen dem Nutzer des Systems zumutbar ist, wird – zumindest bei der Literaturrecherche – zwischen dem Mediator und die Webapplikation eine Zwischenschicht mit Suchmaschinentechnologie gezogen. Um diese bis zur Fertigstellung des Gesamtsystems entwickeln zu können, wurde eine *Django*-Webapplikation mit darunter liegender Suchmaschine (Apache Solr) entwickelt. Als Metadatenformat für diese Schicht dient das *Metadata Object Description Schema (MODS)*. Für die Darstellung der bibliographischen Informationen innerhalb des Mediators wurde die *FRBRoo*-Erweiterung des *CRM* ausgewählt.

Nachdem nun erläutert worden ist, welchen Mehrwert das *ArcheoInf*-Dokumentenrepositorium seinen Benutzern vor allem in Verbindung mit dem Gesamtrecherchesystem bietet, stellt sich natürlich die Frage, wieso ein mit den bibliographischen Angaben der Projekte versorgtes Projekt die Einbindung von *Open Library Data* in Erwägung zieht.

Open Library Data als Quelle für das Repositorium

Die Aufnahme offener Daten war in den Anfängen des Projektes zunächst nicht in den Fokus der Arbeit einbezogen und resultierte aus dem Bemühen, bestimmte, unerwartet auftretende Schwierigkeiten der Datengewinnung zu überwinden und dem Projekt neuen Wert zu verschaffen.

Die Erfahrungen der *ArcheoInf*-Mitarbeiter während des Projektverlaufs haben einen deutlichen Bedarf der forschenden Archäologen nach modernen, technisch aktuellen und spezifisch auf die archäologischen Bedürfnisse ausgerichteten Möglichkeiten der Datenhaltung und -verwaltung gezeigt. Gleichzeitig sind – aufgrund des fehlenden Angebots – die verwendeten Werkzeuge für die Haltung der eigenen Daten in vielen Fällen selbst entwickelte Instrumente, die den Anforderungen an eine Datenbank in keiner Weise entsprechen. So ist es nicht nur äußerst zeitaufwendig, verfügbare Daten für eine Bereitstellung für die *ArcheoInf*-Recherche aufzubereiten und bereitzustellen – diese Arbeiten waren in der verfügbaren Projektlaufzeit nicht zu leisten – es hat sich auch insgesamt als sehr problematisch erwiesen, die Daten vergangener Feldforschungsprojekte für die Aufnahme in das System zu erhalten. Und ist die Zahl der erfassten Forschungsdaten schon als gering zu bezeichnen, so ist die Zahl der mit diesen mitgelieferten bibliographischen Metadaten umso geringer.

Für die Arbeitsgruppe, die das Rechercheportal aufbaut, hat sich so die Erschwernis ergeben, eine Infrastruktur zu entwickeln, die Anforderungen bezüglich der Nachhaltigkeit erfüllt, während relevantes Datenmaterial jedoch nicht zur Verfügung steht. Das Literaturrechercheinstrument konnte somit anfangs nicht ausreichend mit Daten gefüllt werden.

Um potentiellen Benutzern eine lohnenswerte Recherche in unserem System zu ermöglichen und nicht zuletzt auch, um das System hinreichend testen zu können, haben wir uns entschlossen, weitere verfügbare Quellen zur Erweiterung der Inhalte zu nutzen. Welche Quellen kommen an dieser Stelle eher in Frage als die Inhalte der *Open Library Data Cloud*?

Archäologisch relevante Quellen in der Open Library Data Cloud

Bei Beginn der Arbeiten an der Integration offener Daten in das Dokumenten-repositorium Anfang 2011 wurden aus der Menge der zu der Zeit verfügbaren offenen Daten eine Anzahl von möglichen Quellen identifiziert.

Eine gute Übersicht bot hier zur ersten Orientierung die Plattform *the Data Hub*, wo die *Open Knowledge Foundation* gleichsam einen Katalog von freigegebenen Daten bereitstellt. Die eingetragenen Datensets werden begleitet von ergänzenden Beschreibungen zu Formaten, freigebenden Institutionen, thematischer Ausrichtung der Daten und dergleichen mehr.⁷

⁷ Open Knowledge Foundation: About - the Data Hub (2012).

Eine Gruppenzuordnung erleichtert das Auffinden der für den eigenen Bedarf relevanten Datenbestände, in unserem Fall ist dies die Gruppe *Bibliographic Data*.⁸

Hier identifizierten wir zunächst als in Frage kommende Bestände die von den folgenden Einrichtungen zur Verfügung gestellten Daten:

- British National Bibliography der British National Library⁹
- Universitäts- und Stadtbibliothek Köln¹⁰
- Library of Congress^{11,12}
- Universitätsbibliothek Tübingen
- Open Library¹³

Bei der britischen Nationalbibliographie, sowie bei der *Open Library* und bei der Quasi-Nationalbibliothek *Library of Congress* war das Vorliegen auch archäologische relevanter Daten aufgrund des umfassenden Sammlungsziels ohnehin zu erwarten, bei den Universitätsbibliotheken war entscheidendes Kriterium bei der Auswahl das Vorhandensein eines archäologischen Fachbereichs, um mit relevanter Literatur rechnen zu können.

Die seither tatsächlich bearbeiteten und zur Integration vorbereiteten Datenbestände, auf die bezogen wir hier unser Vorgehen darstellen möchten, sind die BNB, die USB Köln und, da aktuell freigegeben und verfügbar, außerdem die gemeinschaftlich freigegebenen Katalogdaten von BVB und KOBV, der B3Kat.¹⁴

Unser Vorgehen umfasst in groben Umrissen das Herunterladen, die Filterung der Datenmenge nach archäologisch relevanten Daten, die Umwandlung in das benötigte Metadatenformat, die Indexierung und den Import in das Dokumentenrepository und schließlich perspektivisch die Verknüpfung der Forschungsdaten mit den bibliographischen Daten.

Die Arbeit mit den Daten

Nachdem die Auswahl der für *ArcheoInf* relevanten Quellen getroffen wurde, besteht die Notwendigkeit, die Daten zu analysieren und in das von uns verwen-

⁸ Open Knowledge Foundation: *Bibliographic Data - the Data Hub* (2012).

⁹ Lowery: *The British National Bibliography* (2012).

¹⁰ Universitäts- und Stadtbibliothek Köln: *Universitäts- und Stadtbibliothek Köln* (2012).

¹¹ Library of Congress: *Library of Congress Home* (2012).

¹² Universitätsbibliothek Tübingen: *Universitätsbibliothek - Home* (2012).

¹³ OpenLibrary: *Welcome to Open Library* (2012).

¹⁴ Siehe zum B3Kat den Beitrag von Ceynowa et. al in diesem Band.

deten *MODS*-Format zu konvertieren. Der letzte Schritt nutzt die bereits in den Universitätsbibliotheken Bochum und Dortmund etablierten Verfahren zur Indexierung von *MODS*-Daten in eine Suchmaschine.

Wie im Metadatenumfeld üblich, gilt auch hier: *MODS* ist nicht gleich *MODS*. Wir haben uns insbesondere bei der Abbildung der bibliographischen Hierarchien eine gemeinsame Struktur von „Titelaufnahme“ und der zugehörigen „Über- und Unterordnungen“ überlegt. Im Gegensatz zu der in unseren Systemen üblichen Aufteilung der einzelnen Hierarchien in separate Datensätze, ist es bei *MARC* und somit auch bei *MODS* üblich, alle Elemente in einem Datensatz zu führen. *MODS* verwendet hierfür *relatedItem*-Elemente, die verschiedene Formen annehmen können. Dazu zählen *host*, *series* und *constituent*. Wir haben uns für folgende Aufteilung entschieden:¹⁵

```
<mods>
<!-- Band eines mehrbändig begrenzten Werks (mbW)/Volume/Jahrgang/
einbändiges Werk nicht Teil eines mbW (1) -->
<relatedItem type="series">
  <!-- fortlaufendes Sammelwerk (2) -->
</relatedItem>
<relatedItem type="constituent">
  <!-- Band/Heft/Beitrag/Artikel (3) -->
  <relatedItem type="constituent">
    <!-- Beitrag/Artikel (4) -->
  </relatedItem>
</relatedItem>
</mods>
```

Tabelle 1: Beispiele für Hierarchien im *MODS*-Datensatz.

	Beispiel 1	Beispiel 2	Beispiel 3	Beispiel 4
(1)	Monographie	Monographie	LNCS ¹⁴ -Band	Jahrgang einer Zeitschrift
(2)	--	Reihe/Serie	LNCS	Zeitschrift
(3)	--	--	Beiträge	Heft
(4)	--	--		Artikel

¹⁵ Bei den Publikationstypen folgen wir den Ausführungen in Gantert/Hacker: Bibliothekarisches Grundwissen (2008). Vgl. dazu auch den Beitrag „FRBR, Serials und CIDOC CRM“ in diesem Band.

¹⁶ LNCS = Lecture Notes in Computer Science ist eine Reihe von Konferenzbänden.

In **Tabelle 1** wird dies durch Beispiele veranschaulicht. Die Vorteile dieser Struktur liegen in der Übertragbarkeit auf das in Suchmaschinen übliche flache Datenschema und – wie wir später noch sehen werden – in der „sauberen“ Übertragbarkeit auf die Work-Entitäten der *FRBRoo*.

Mapping British National Bibliography (BNB)

Die Wahl für das erste zu bearbeitende Datenset fiel auf die Open-Data-Bestände, die die *British National Library* zur Verfügung stellt. Nachdem bereits früher Daten der *British Library* für Forschungszwecke – zum damaligen Zeitpunkt aber zu restriktiven Bedingungen¹⁷ – bereitgestellt worden waren, erfolgte die neuerliche Freigabe im Jahr 2010 unter einer *Creative Commons Zero* Lizenz¹⁸.

Bei der *British National Bibliography* handelt es sich um den Teilbestand des Katalogs der *British Library*, der die in Großbritannien und Irland seit 1950 erschienenen Publikationen umfasst.¹⁹ Für britische und irische Verlage besteht eine Abgabepflicht ihrer Publikationen, so dass mit den BNB-Daten ein umfassender Bestand von etwa 3 Millionen Titeldaten verfügbar gemacht worden ist.²⁰

Aufgrund des Sammlungsprofils der BNB stand bereits vor dem Beginn unserer Arbeit an den Daten fest, dass hier eine große Zahl von für unsere Zwecke nutzbaren Daten zu erwarten wäre. Eine ungefähre Einschätzung der zu erwartenden Menge an archäologisch und damit für unsere Zwecke verwertbaren Daten wurde zunächst durch eine Recherche im Online-Katalog der BNB²¹ ermittelt. Die Suche nach einschlägigen Schlagwörtern sowie den relevanten Klassifizierungen ergab eine zu erwartende Treffermenge von weit mehr als 4000 archäologisch relevanten Titeln.

Bei den zu Beginn der Bearbeitung verfügbaren Datensätzen²² handelte es sich um einfache *RDF/XML*-Repräsentationen, die unter der Verwendung der

17 Flimm: JISC gibt die Britische Nationalbibliographie als Open Data unter CC0 frei (2012).

18 Informationen zu Creative Commons Lizenzen: Creative Commons: Creative Commons Deutschland (2012).

19 Details und Ausnahmen siehe: McKinley: The British National Bibliography Exclusions Policy (2012).

20 Lowery: The British National Bibliography (2012).

21 The British Library: Explore the British Library (2012).

22 Die vorliegende Beschreibung bezieht sich auf eine ältere Datenversion. Die Katalogdaten der BNB liegen aktuell auf Basis eines geänderten Datenmodells vor. Das neue British Library Data Model setzt, neben den oben genannten, zahlreiche weitere verbreitete Ontologien sowie die eigens erstellten British Library Terms ein, die Klassen und Beziehungen der ausgewählten Schemata ergänzt. Die British Library hat zudem eigene URIs geprägt, so etwa für Autoren/

folgenden Vokabulare erstellt wurden: *Dublin Core*²³, *ISBD*²⁴, *SKOS*²⁵, *OWLT*²⁶, *RDFS*²⁷.

Zur Vorbereitung der Konvertierung des Ausgangsformats in das Zielformat *MODS* entschieden wir uns zunächst für die Erstellung einer Konkordanz, um später eine genaue Abbildung der BNB-XML-Elemente auf die *MODS*-Struktur zu erreichen. Hierzu nutzten wir einige ausgewählte Beispieldatensätze, die eine Höchstzahl der zu erwartenden Inhalte und auch häufiger zu erwartende Hindernisse enthielten. Im Mittelpunkt dieser Arbeit stand vor allem die Analyse und Verarbeitung der in den einzelnen Elementen angegebenen bibliographischen Daten, die in dieser Form nicht den Anforderungen an die Inhalte der *MODS*-Elemente entsprachen.

Als Beispiel soll hier die Titelangabe aus einem der ausgewählten Datensätze dienen. Es ist deutlich zu erkennen, dass später getrennt aufzuführende Angaben im Sinne einer bibliographischen Hierarchie in einem Element zusammengefasst sind und für die weitere Verarbeitung getrennt werden müssen.

```
<dcterms:title>
Ancient Naukratis. Vol.2, The survey at Naukratis and Environs. Pt.1, The survey at Naukratis
</dcterms:title>
```

Ähnliches gilt für andere Elemente, etwa das die Person beschreibende:

```
<dcterms:contributor>
<rdf:Description>
<rdfs:label>
Coulson, William D. E., 1942-2001.
</rdfs:label>
```

Beitragende (Personen und Organisationen), verschiedener Kategorien von Schlagworten und Klassifikationen, Überordnungsbeziehungen und Ereignisse im Kontext der Publikation Einen Überblick über das Datenmodell findet sich unter: British Library Data Model - Book (2012). bzw. British Library Data Model - Serials (2012)... Ein Beispieltitel ist hier verfügbar: : British National Bibliography (BNB) - Linked Open Data / Example resource RDF - ttl/json/html - default html - Datensatz - Ressource - the Data Hub (2012). In der Folge der genannten Änderungen musste das Mapping in das ArcheoInf-Zielformat stark angepasst werden. Die oben beschriebenen Vorgehensweise kann aber weiterhin beispielhaft verdeutlichen, welche Probleme und Schwierigkeiten im Konvertierungsprozess auftreten können.

23 DCMI Metadata Terms (2012).

24 The Registry: ISBD Elements (2012).

25 SKOS Simple Knowledge Organization System Namespace Document - HTML Variant, 18 August 2009 Recommendation Edition (2011).

26 OWL Time ontology (2006).

27 RDFS (2004).

```
</rdf:Description>
</dcterms:contributor>
```

Ebenso die Angabe von Überordnungen:

```
<dcterms:isPartOf>
<rdf:Description>
  <rdfs:label>Oxbow monograph ; 60</rdfs:label>
</rdf:Description>
</dcterms:isPartOf>
```

Diese Bestandteile sollen im Ergebnisdatensatz die folgende Struktur aufweisen:

```
<mods version="3.0"
  xsi:schemaLocation="http://www.loc.gov/mods/v3
    http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
<titleInfo>
  <title>Ancient Naukratis</title>
</titleInfo>
<relatedItem type="series">
  <titleInfo>
    <title>Oxbow monograph</title>
  </titleInfo>
</relatedItem>
<part>
  <detail type="volume">
    <number>60</number>
  </detail>
</part>
<relatedItem type="constituent">
  <titleInfo>
    <title>The survey at Naukratis and Environs</title>
    <partNumber>Pt.1</partNumber>
    <partName>The survey at Naukratis</partName>
  </titleInfo>
  <part>
    <detail type="volume">
      <number>2</number>
    </detail>
  </part>
  ...
  <name>
    <namePart>Coulson, William D. E.</namePart>
    <namePart type="date">1942-2001</namePart>
    <role>
```

```

<roleTerm type="text" authority="marcrelator">
  contributor
</roleTerm>
<roleTerm type="code" authority="marcrelator">
  ctb
</roleTerm>
</role>
</name>
</relatedItem>
</mods>

```

Nach der Analyse der Datenstruktur bestand die besondere Herausforderung bei der Weiterverarbeitung der freigegebenen Daten nun darin, die in den BNB-Datensätzen vorliegenden Zeichenketten im Rahmen der Erstellung eines *XSLT-Stylesheets*²⁸ zuverlässig so zu trennen, dass die benötigten Bestandteile im Endformat im jeweils korrekten Element zu finden sind. Dies wiederum erforderte eine genaue Prüfung der übrigen Datensätze, um ein Set von verlässlichen Kriterien zur Trennung der einzelnen Bestandteile zu ermitteln. Insbesondere die Trennung des Titels in seine verschiedenen Einheiten einschließlich der Aufschlüsselung der in den vorliegenden Angaben dargestellten Verhältnisse der Über- und Unterordnungen und der korrekten Abbildung im Metadatenformat *MODS* stellten hierbei eine Herausforderung dar.

In den *RDF/XML*-Daten der BNB fehlte zudem ein eindeutiger Bezeichner, der die verzeichneten Publikationen zweifelsfrei als bestimmte Dokumententypen erkennen ließ. Ohne Unterschied waren alle Publikationen – ob *Proceedings*, Monographie, Teilband eines mehrbändigen Werkes und dergleichen mehr – mit dem Label *monographic* versehen. Der Charakter eines Titels als konkreter Publikationstyp ist deshalb ausschließlich und soweit möglich aus den angegebenen Informationen zu Titel sowie gegebenenfalls Über- und Unterordnungen zu entnehmen.

Auch die Personenangaben ließen an dieser Stelle keine weiteren Schlussfolgerungen zu, da die Daten dort ebenfalls keine näheren Spezifizierungen der Rollen der angegebenen Personen enthielten. Die einheitliche Bezeichnung aller Beteiligten mit *dcterms:contributor* und ohne weitere Definition erlaubte es uns weder, den Publikationstyp näher zu bestimmen, noch den Beitragenden in unserem Zielformat eine genauer bezeichnete Rolle zuzuweisen.

Am Ende der Bearbeitung der BNB-Daten im Rahmen des *Mappings* stand schließlich die Entscheidung darüber, wie die Menge der verfügbaren Daten nach solchen Titeln gefiltert werden sollte, die für die Verwendung in einem archäolo-

²⁸ XSL Transformations (XSLT) (2003).

gischen Informationssystem geeignet sein würden. Neben der Verschlagwortung durch die *Library of Congress Subject Headings* sind die Titel der BNB durch die *Dewey Decimal Classification* klassifiziert. Die Wahl fiel hier schnell auf eine Filterung der Daten anhand von Klassifikationen, da es auf diese Weise möglich ist, mit einer relativ geringen Menge von Filterkriterien eine maximale Anzahl an relevanten Daten bei einer äußerst geringen Anzahl an unbrauchbaren Daten zu erhalten. Eine Filterung über Schlagwörter würde sich an dieser Stelle als deutlich arbeitsintensiver erweisen.

Das Filterverfahren mit Hilfe von Klassifikationen wurde auch bei den später bearbeiteten Datenbeständen beibehalten und entsprechend der jeweils verwendeten Klassifikationssysteme angepasst.

Mapping USB Köln

Die Inhalte des Bibliothekskatalogs der Universitäts- und Stadtbibliothek Köln wurden der Öffentlichkeit zu Beginn des Jahres 2010 in Kooperation mit dem Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz) zur freien Nutzung zur Verfügung gestellt. Mit mehr als 3 Millionen Titelaufnahmen²⁹ und einem der größten archäologischen Institute Deutschlands an der Philosophischen Fakultät der Universität zu Köln war auch in diesen Daten ein nennenswerter Bestand an für unsere Zwecke verwendbaren Titelaufnahmen zu vermuten.

Die Kölner Daten liegen in einem Metadatenformat vor, das auf den Kategorien des MAB2 beruht und dessen Struktur in einem eigens erstellten Wiki-Artikel dokumentiert ist.³⁰ Hier findet sich eine kurze Erläuterung des Aufbaus des Formats sowie eine Aufstellung der verwendeten Kategorien und der zugehörigen Inhalte. Die hier vorliegende Schlüssel-Wert-Struktur bedingte einige Anpassungen in unserer Vorgehensweise.

Die Bearbeitung der Kölner Daten gestaltete sich zunächst so, dass auch hier eine Anzahl von Datensätzen ausgewählt wurde, die alle zu erwartenden Publikationstypen repräsentierten. Diese wurden anschließend im Hinblick auf die tatsächlichen Inhalte der verwendeten Kategorien untersucht und in einem entsprechenden Modell-MODS-Datensatz abgebildet.

²⁹ Universitäts- und Stadtbibliothek Köln: Open Bibliographic Data an der Universitäts- und Stadtbibliothek Köln (2011). Daneben sind in den freigegebenen Daten 1.5 Millionen Personenaufnahmen, 156 Tausend Körperschaftsaufnahmen, 40 Tausend Notationen sowie 243 Tausend Schlagworte enthalten.

³⁰ Metadaten-Format – OpenBib Wiki (2012).

Als Hürde erwiesen sich auch hier die Abbildungen der bibliographischen Hierarchien. Konkret meint dies beispielsweise die Aufnahme der einzelnen Bände mehrbändiger Werke, die jedoch wiederum als Ganzes (d. h. alle Teile des mehrbändigen Werkes) Teil einer ungezählten Reihe bilden, wobei jeder Teilband einen eigenen Teil einer gezählten Reihe darstellt.³¹ Die Aufnahme der jeweiligen Teilbände einschließlich sowohl der Angabe der Bandnummer der gezählten Reihe als auch des übergeordneten Titels einschließlich der Angabe der diesem zugeordneten ungezählten Reihe in getrennten Datensätzen erforderten zum Zwecke der späteren korrekten Darstellung in unserem Zielformat *MODS* ein Ansprechen der verschiedenen Datensätze, um die einzelnen Bestandteile zu einem Datensatz zusammenzuführen.

Als verbindendes Element und als eindeutiger Bezeichner erwiesen sich die Identifikationsnummern der einzelnen Datensätze. Diese ermöglichten den Zugriff auf die Informationen, die wir zur Vervollständigung unserer *MODS*-Datensätze benötigten. (s. u. Konvertierung)

Ein weiterer Unterschied zu den zuvor bearbeiteten Daten stellte die Tatsache dar, dass sich in den herunterzuladenden Beständen neben der Datei,³² die die Titeldaten des Katalogs der USB enthält, weitere gesondert bereitgestellte Dateien finden, die Informationen zu Normdaten enthalten. Diese Normdaten sind wiederum mit Identifikationsnummern versehen, die in den Titeldatensätzen verwendet werden, um auf die benötigten Daten, wie etwa Schlagwörter oder Personennamen, Bezug zu nehmen. Um später in den für *ArcheoInf* verwendeten Daten Namen sowie Schlagwörter als sprechenden Text darstellen zu können, muss zusätzlich ein Zugriff auf die Normdaten-Dateien der USB erfolgen, um diesen die den *IDs* zugeordneten Inhalte zu entnehmen und in die *MODS*-Daten zu überführen und dort anzuzeigen.

31 Bsp. Schweitzer, Marcell u. Franz Xaver Bea: Allgemeine Betriebswirtschaftslehre. 1 Grundlagen. 5. Aufl. Stuttgart: Fischer 1990 (=Uni-Taschenbücher 1081). und Schweitzer, Marcell u. Franz Xaver Bea: Allgemeine Betriebswirtschaftslehre. 2 Führung. 4. Aufl. Stuttgart, Stuttgart: Fischer; Lucius & Lucius 1989 (=Uni-Taschenbücher 1082)., wobei der übergeordnete Titel Schweitzer, Marcell: Allgemeine Betriebswirtschaftslehre. Hrsg. von Franz Xaver Bea u. Erwin Dichtl (Hrsg.). Stuttgart: Lucius & Lucius (=Grundwissen der Ökonomik. Betriebswirtschaftslehre)., Teil der ungezählten Reihe Grundwissen der Ökonomik. Betriebswirtschaftslehre ist.

32 Neben dem Gesamtbestand der USB Köln sind auch Teilbestände gebildet worden, die sich einzeln herunterladen lassen. Für unsere Zielsetzung war es sinnvoll, die Daten zu verwenden, die den gesamten Bestand umfassen. Vgl. zu den Teilbeständen: Universitäts- und Stadtbibliothek Köln: Open Bibliographic Data an der Universitäts- und Stadtbibliothek Köln (2011).

Mapping B3Kat

Die Freigabe der Daten des B3Kat, des gemeinsamen Katalogs der Bayerischen Staatsbibliothek, des Bibliotheksverbundes Bayern und des Kooperativen Bibliotheksverbund Berlin-Brandenburg als *Linked Open Data* zum Ende des Jahres 2011 war für uns Grund genug, uns nach Beendigung der Arbeit an den oben beschriebenen Paketen direkt der Bearbeitung dieses Angebotes zuzuwenden. Die Datenmenge von mehr als 23 Millionen Medien³³ sowie die vertretenen Bestände ließen einen entsprechend umfangreichen Teil an für *ArcheoInf* verwertbaren Daten vermuten.³⁴

Ganz gemäß den Anforderungen an *Linked (Open) Data* wird in den Daten des B3Kat eine große Zahl an Links zu den entsprechend relevanten Datenbeständen verwendet.³⁵ Dies wiederum erforderte bei unserem Mapping zu *MODS* den Zugriff auf die jeweils dort verlinkten Bestände, um – ähnlich, wie es oben bei den Daten der USB Köln der Fall war – in der *ArcheoInf*-Literaturrecherche für die Benutzersicht den Inhalt in Textform abbilden zu können. Für die Einrichtung des Zugriffs auf das entsprechende Element in den verlinkten Dokumenten wurde dieses zunächst aufgerufen und – sofern verfügbar – das zugehörige *RDF/XML*-Dokument in Autopsie auf seine Struktur hin analysiert. Anschließend wurde unter Verwendung eines *XSLT-Stylesheets* eine Regel für die Ansprache des benötigten Elements und die Ausgabe des zu verwendenden Wertes erstellt. Im Falle etwa eines Autors bedeutet dies, dass auf die *RDF/XML*-Repräsentation des GND-Datensatzes zugegriffen wird um den Inhalt des Elements *gnd:preferredNameForThePerson* ausgeben zu können.

Im Falle von Autoren- oder Herausgeberangaben erwies sich dieses Vorgehen als vergleichsweise unkompliziert, da alle hier angegebenen Links auf den Datenpool der GND verweisen und die Struktur dieser Daten zuverlässig einheitlich ist. Nach der Einführung der GND-Ontologie im April 2012 waren allerdings einige Anpassungen des ursprünglichen *Mappings* erforderlich.

Eine umfangreichere Bearbeitung war bei den Schlagwortangaben bzw. bei den Angaben der Klassifikationen vonnöten – es sollten hier sowohl die Schlagwörter der GND als auch die Notationen der RVK übernommen werden. Die Übernahme der Klassifikationen der *DDC* erwies sich als unproblematisch, da diese hier in Textform angegeben sind. Bei der Ausgabe der Schlagwörter der GND war

33 Bayerische Staatsbibliothek: URIs und Datenmodell - Dokumentation des LinkedOpenDataService lod.b3kat.de (2012).

34 Zur Freigabe der Daten des B3Kat siehe den Beitrag in diesem Band.

35 Bayerische Staatsbibliothek: URIs und Datenmodell - Dokumentation des LinkedOpenDataService lod.b3kat.de (2012).

in erster Linie die unterschiedliche Struktur der *RDF*-Repräsentationen der personenbezogenen bzw. sachbezogenen Schlagwörter zu beachten und das Mapping entsprechend anzupassen. In Bezug auf die Daten der GND erwies sich hier in den Anfängen der Arbeit die Tatsache als Hindernis, dass nicht alle verzeichneten Schlagwörter auch als *RDF*-Repräsentationen erfasst waren.

Im Gegensatz zu den Daten der *British Library* eröffneten die B3Kat-Daten uns die Möglichkeit, auch die Rolle der angegebenen Personen in einigen Fällen näher zu bezeichnen. Neben *dc:creator* findet *bibo:editor* Verwendung, das uns erlaubt, die Herausgeber etwa von Sammelwerken auch als solche zu bestimmen und anzuzeigen.

Gleiches gilt hier auch für die Angaben zum Publikationstyp. Neben der Angabe `<rdf:type rdf:resource="http://purl.org/ontology/bibo/Book"/>` sind hier auch zu finden: `<rdf:type rdf:resource="http://purl.org/ontology/bibo/Proceedings"/>` sowie `<rdf:type rdf:resource="http://purl.org/ontology/bibo/Thesis"/>`, was uns gestattet, auch auf der Ebene des Publikationstyps in unserer Anzeige eine nähere Spezifizierung vorzunehmen.

Konvertierung

Um nun die Quelldaten in das *MODS*-Format umzuwandeln, sind verschiedene Programme entstanden. Mit dem Ziel, die Anzahl dieser Programme möglichst klein zu halten, wurde verstärkt auf die Verarbeitung von *XML*-Formaten mittels der *XSLT*-Technologie gesetzt. Durch diese Technologie ist es möglich, *XML*-Daten durch die Verwendung von *XML*-basierten *Stylesheets* in andere (*XML*-) Formate umzuwandeln (vgl. Abbildung 2). Aufgrund der großen Datenmengen und der Arbeitsspeicher-intensiven Verarbeitung der *XML*-Daten, kommen bei uns *Streaming*-Mechanismen des *SaxonEE*-Prozessors³⁶ zur Anwendung.

Für die Daten der USB Köln ist ein *Preprocessing* notwendig, da die Daten aus dem lokalen Bibliothekssystem *SunRise* – wie oben beschrieben – in einem text-basierten, Schlüssel-Wert-Schema vorliegen. Ferner sind dort für die Abbildung der bibliographischen Hierarchien *IDs* abgelegt, die für eine weitere Verarbeitung aufgelöst bzw. die in den zugehörigen Datensätzen vorliegenden Informationen ausgelesen werden müssen. Hierzu wurde in der UB Dortmund ein Java-Programm geschrieben, welches auf mehrere parallele Prozesse verteilt, die Daten in *MODS* umwandelt.³⁷

³⁶ Web-Seite: The SAXON XSLT and XQuery Processor (2012).

³⁷ Weitere Informationen zu den Mappings und zu den Konvertern folgen auf der Open Data-Webseite der UB Dortmund: Universitätsbibliothek Dortmund: Offene bibliographische Daten an

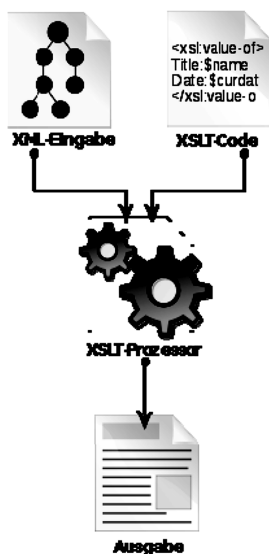


Abbildung 2: Vereinfachte Darstellung der Funktionsweise von XSLT.³⁸

Indexierung

Die auf diese Weise erzeugten *MODS*-Daten werden im finalen Schritt für zwei Arten der Indexierung nochmals umgewandelt. Für den Anwendungsfall Suchmaschine werden die *MODS*-Daten in das für Apache Lucene/Solr³⁹ notwendige Importformat umgewandelt. Darin werden die Hierarchien des *MODS*-Datensatzes auf entsprechende Felder übertragen. Für die Titelangaben der Beispiele aus **Tabelle 1** gilt beispielsweise: *title* für (1) aus Beispiel 1 und 2, für (3) aus Beispiel 3 und (4) aus Beispiel 4; *serialtitle* für (2) aus Beispiel 2 und 3; *journaltitle* für (2) aus Beispiel 4; ...

Für die Ebene des Mediators wenden wir das Verfahren aus dem Beitrag „FRBR, Serials und CIDOC CRM“ in diesem Band an. Durch die Übertragung der mit dem Verfahren des vorigen Abschnitts erzeugten *MODS*-Daten in ein *FRBRoo*-konformes *RDF-XML*, wird die Anbindung der bibliographischen Daten an das *CIDOC CRM* erreicht. Hierbei sind die in den vorigen Abschnitten erzeugten *MODS*-Datensätze Instanzen des *E31 Document* aus dem *CIDOC CRM*. Diese

der Universitätsbibliothek Dortmund (2012).

³⁸ Wikipedia: XSL Transformations (XSLT) (2003).

³⁹ Apache Lucene - Apache Solr (2012.); verwendet wird derzeit Version 3.5.

beschreiben bibliographische Objekte der Form *Manifestation*, welche wiederum als *F3 Manifestation Product Type* vom *CRM* abgeleitet ist.⁴⁰

Bei der Konvertierung der *MODS*-Daten nach *RDF* wird die *CIDOC CRM* und *FRBRoo*-Darstellung des Projektes *Erlangen CRM*⁴¹ verwendet. Dabei werden die Hierarchien der *MODS*-Datensätze aus den Beispielen aus **Tabelle 1** wie folgt auf die *FRBRoo*-*Work*-Entitäten abgebildet: *F14 Individual Work* für (1) aus Beispiel 1 und 2; *F17 Aggregation Work* für (1) aus Beispiel 3 und (3) aus Beispiel 4; *F18 Serial Work* für (2) aus den Beispielen 2,3 und 4; *F15 Complex Work* für (1) aus Beispiel 4.

In beiden Fällen wird vor der Konvertierung die weiter oben beschriebene Filterung der Daten nach fachlichen Bereichen der Klassifikationen *DDC*, *RVK* und Basisklassifikation durchgeführt.

Verbindung von bibliographischen Daten zu Forschungsdaten

Nachdem nun die Aggregation und Indexierung von Daten diskutiert wurde, wird noch der Aspekt der Verbindung von bibliographischen Daten zu Forschungsdaten betrachtet. Bei der Erfassung von Forschungsdaten kommt es nicht selten vor, dass auch Literaturstellen mit diesen verknüpft werden. Mögliche Szenarien sind dabei Nachweise zu Quellen, die zu einer Einordnung eines Objektes führen, oder Publikationen zum gleichen Forschungsthema. Solche Verknüpfungen sind auch in den in *ArcheoInf* vorliegenden Daten der archäologischen Feldforschungsprojekte vorhanden und müssen somit ebenfalls in die Modellierung des Mediators einfließen.

Das *CRM* lässt mindestens drei Ebenen der Verknüpfung von bibliographischen Daten und Forschungsdaten zu:

1. Referenzierung von einer Publikation zu einem Objekt
2. Zuordnung über die Verwendung inhaltserschließender Konzepte
3. Referenzierung von objektbeschreibenden Dokumenten

Die Zuordnung von Publikationen zu (Forschungs-)objekten geschieht über das Ereignis *E13 Attribute Assignment*, zu dem es in der *CRM*-Spezifikation heißt: „This class comprises the actions of making assertions about properties of an

⁴⁰ Näheres in „Titelaufnahmen und Linked Data“ aus der Dokumentation der Linked Open Data-Plattform der UB Dortmund:Universitätsbibliothek.

⁴¹ Schiemann/Oischinger: Erlangen CRM / OWL - An OWL DL 1.0 implementation of the CIDOC Conceptual Reference Model (CIDOC CRM) (2011).

object or any relation between two items or concepts.“ Dabei referenziert ein *E89 Propositional Object*, zu dem auch die *F19 Publication Works* zählen, ein anderes Objekt in dem Sinne, dass sich die Publikation auf dieses Objekt bezieht. Abbildung 3 stellt dies schematisch dar.

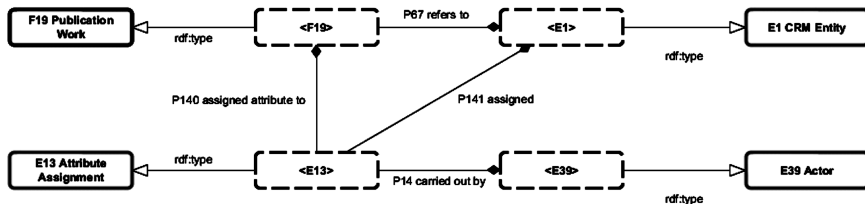


Abbildung 3: direkte Relationierung von Publikationen und (Forschungs-)Objekten.

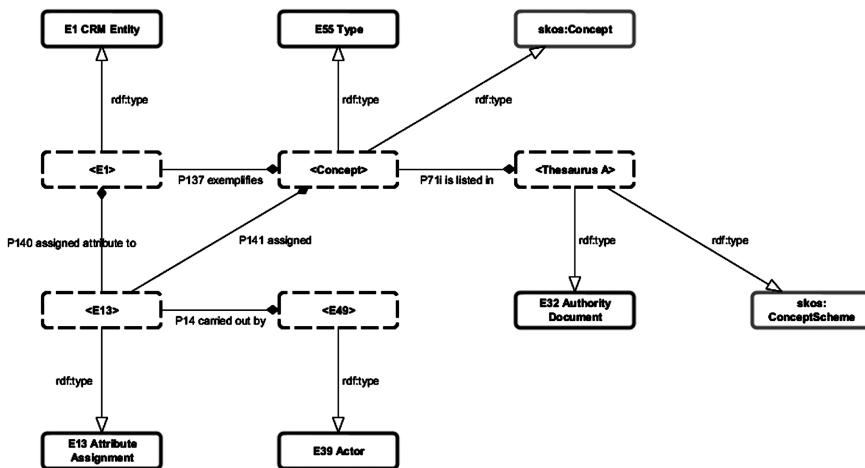


Abbildung 4: Inhaltliche Erschließung eines (Forschungs-)Objektes.

Eine Spezialisierung dieser Möglichkeit stellt die inhaltliche Zuordnung von Konzepten zu einem Objekt dar. Hierbei werden beispielsweise Begriffe aus Thesauri oder Klassifikationselemente den Objekten zugeordnet. Dies gilt für alle Arten von Objekten innerhalb des *CRM*. Allerdings kann dies auf zwei verschiedene Weisen geschehen. Zum einen kann die Aussage getroffen werden, dass ein Objekt ein Exemplar eines Konzepts ist (*P137 exemplifies*; z.B. „Object BM000098044 of the Clayton Herbarium [...] exemplifies *Spigelia marilandica*[...]“⁴²; vgl. Abbildung 4). Zum anderen kann die Aussage getroffen werden, dass ein *E89 Propositional*

42 Vgl. Crofts/Doerr: Definition of the CIDOC Conceptual Reference Model (2011). 72.

nal Object – also auch hier wieder Publikationen – von einem Konzept handelt. Hierzu wird *P129 is about* verwendet (vgl. Abbildung 5).

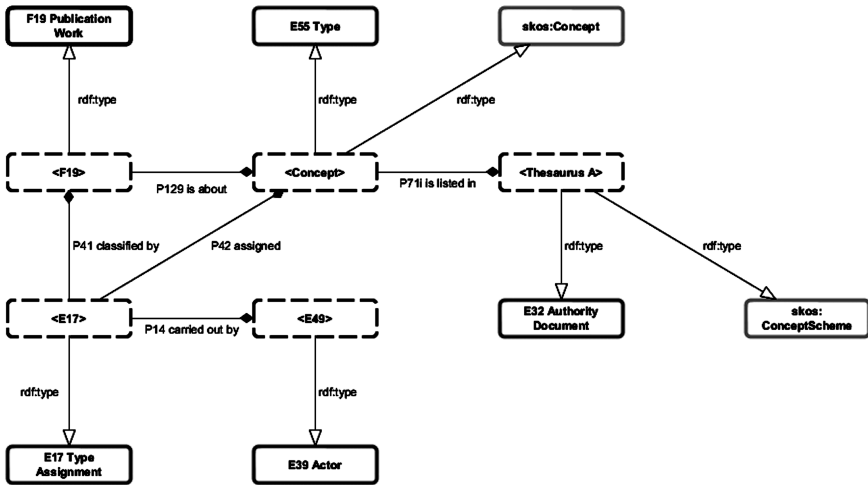


Abbildung 5: Inhaltliche Erschließung einer Publikation.

Über die gemeinsame Verwendung einer Instanz eines *E32 Authority Documents* – welches auch zeitgleich eine Instanz von *skos:Concept* sein kann – lassen sich somit Forschungsobjekte und die damit verbundenen Forschungsdaten mit bibliographischen Daten und Objekten verknüpfen.

Die dritte Variante ist eine eher technische. Hierbei wird davon ausgegangen, dass es zwar eine Verbindung gibt, sich diese aber nicht auf inhaltlicher Ebene darstellt, sondern nur die Metadattendokumente miteinander verbunden sind (vgl. Abbildung 6).

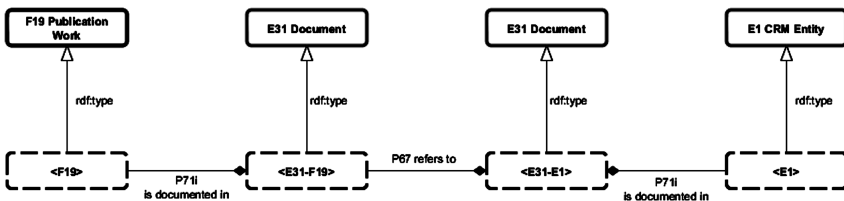


Abbildung 6: Zuordnung von Objekten über ihre Metadattendokumente.

Hierbei werden, wie bei dem Verhältnis zwischen *MODS*-Datensatz zur *Manifestation* (s.u. Indexierung), den Forschungsobjekten beschreibende *E31 Document*-

Objekte zugeordnet und diese dann mit den Metadatendokumenten (in unserem Fall den *MODS*-Datensätzen) verknüpft.

Fazit

Der vorliegende Beitrag hat gezeigt, dass *Linked Open Library Data* im Kontext einer Forschungsdatenumgebung eine zentrale Rolle spielen kann. *Linked Open Library Data* macht es möglich, die im Projekt *ArcheoInf* entwickelte Infrastruktur trotz der (politischen) Probleme bei der Datenakquise fortzuentwickeln und überdies die Inhalte des Informationssystems anzureichern.⁴³

Über die Verwendung des *CIDOC-CRM* und dessen Erweiterung *FRBROo* können die mit Forschungsdaten verknüpften bibliographischen Metadaten der *Linked Data Cloud* bzw. der *Open Data Community* wieder zur Verfügung gestellt werden. Für diese Zwecke ist es unabdingbar, dass die Daten offen im Sinne der *Open Knowledge Definition* zur Verfügung gestellt werden.⁴⁴

Literaturverzeichnis

- About - the Data Hub. <http://thedatahub.org/about> (28.6.2012).
- Apache Lucene - Apache Solr. <http://lucene.apache.org/solr/> (1.7.2012).
- Universitätsbibliothek Bochum; Universitätsbibliothek Dortmund: ArcheoInf Dokumentenrepositorium. <http://archeoinf.ub.rub.de/> (1.7.2012).
- Bayerische Staatsbibliothek: LinkedOpenData-Service des B3Kat: lod.b3kat.de. <http://lod.b3kat.de/doc> (28.6.2012).
- Bayerische Staatsbibliothek: URIs und Datenmodell - Dokumentation des LinkedOpenData-Service lod.b3kat.de. <http://lod.b3kat.de/doc/uris.html> (28.6.2012).
- Bibliographic Data - the Data Hub. <http://thedatahub.org/group/bibliographic> (28.6.2012).
- British Library Data Model - Book. <http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf> (11.10.2012).
- British Library Data Model - Serial. <http://www.bl.uk/bibliographic/pdfs/bldatamodelserial.pdf> (11.10.2012).
- British National Bibliography (BNB) - Linked Open Data / Example resource RDF - ttl/json/html. <http://thedatahub.org/de/dataset/bluk-bnb/resource/13aaf04d-8920-478a-908a-7b0dd858be51> (15.10.2012).

43 Eine Webapplikation zur Suchmaschine des Projektes ist unter Universitätsbibliothek Bochum/Universitätsbibliothek Dortmund: ArcheoInf Dokumentenrepositorium (2012), erreichbar; der Zugang zum Triple Store der FRBROo-RDF-Daten befindet sich unter <http://data.archeoinf.tu-dortmund.de> (29.06.2012).

44 Open Definition der Open Knowledge Foundation: Smith: Open Definition.

- Creative Commons: Creative Commons Deutschland. <http://de.creativecommons.org/> (28.6.2012).
- DCMI Metadata Terms. <http://purl.org/dc/terms/> (28.6.2012).
- Flimm, Oliver: JISC gibt die Britische Nationalbibliographie als Open Data unter CC0 frei - OpenBibBlog. <http://blog.openbib.org/2010/11/18/jisc-gibt-die-britische-nationalbibliographie-als-open-data-unter-cc0-frei/> (28.6.2012).
- Library of Congress: Library of Congress Home. <http://www.loc.gov/index.html> (28.6.2012).
- sMetadaten-Format – OpenBib Wiki. <http://wiki.openbib.org/index.php?title=Metadaten-Format> (28.6.2012).
- OWL Time ontology. <http://www.w3.org/2006/time#> (28.6.2012).
- RDFS. <http://www.w3.org/2000/01/rdf-schema#> (28.6.2012).
- SKOS Simple Knowledge Organization System Namespace Document - HTML Variant, 18 August 2009 Recommendation Edition. <http://www.w3.org/2004/02/skos/core#> (28.6.2012).
- The British Library: Explore the British Library. <http://bnb.bl.uk/> (28.6.2012).
- Universitätsbibliothek Tübingen: Universitätsbibliothek - Home. <http://www.uni-tuebingen.de/nc/einrichtungen/universitaetsbibliothek/home.html> (28.6.2012).
- Universitäts- und Stadtbibliothek Köln: Open Bibliographic Data an der Universitäts- und Stadtbibliothek Köln. <http://opendata.ub.uni-koeln.de/> (28.6.2012).
- XSL Transformations (XSLT). <http://www.w3.org/TR/xslt/> (1.7.2012).
- Bekiari, Chrissoula, Martin Doerr u. Patrick Le Bœuf: Definition of Object-Oriented FRBR. http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0.2.pdf (1.7.2012).
- Crofts, Nick, Martin Doerr u. a.: Definition of the CIDOC Conceptual Reference Model. http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.2.pdf (1.7.2012).
- Crofts, Nick, Martin Doerr u. a.: Definition of the CIDOC Conceptual Reference Model. http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf (2.7.2012).
- Bubke, Karolin u. Maike Lins: Zentrum für archäologische Feldforschungsdaten. das Projekt ArcheoInf. In: Die Bibliothek als Erfolgsfaktor. die Universitätsbibliothek Bochum nach 50 Jahren ; 1962 - 2012. Hrsg. von Erdmute Lapp. Bochum: Schürmann + Klaggas 2012. S. 122–127. <http://repo.ub.rub.de/bibliographie/110791991/Festschrift.pdf> (1.07.2012).
- Deutsche Initiative für Netzwerkinformation e. V.: Positionspapier Forschungsdaten. Arbeitsgruppe «Elektronisches Publizieren». <http://edoc.hu-berlin.de/series/dini-schriften/2009-10/PDF/10.pdf> (28.6.2012).
- Gantert, Klaus u. Rupert Hacker: Bibliothekarisches Grundwissen. 8. Aufl. München: Saur 2008.
- Hammerschmitt, Marcus: Die Uhr läuft. Die NASA hat Probleme mit den Innovationszyklen in der IT-Industrie. <http://www.heise.de/tp/artikel/12/12538/1.html> (29.10.2012).
- Lowery, John: The British National Bibliography. <http://www.bl.uk/bibliographic/natbib.html> (28.6.2012).
- McKinley, Cynthia: The British National Bibliography Exclusions Policy. <http://www.bl.uk/bibliographic/exclude.html> (28.6.2012).
- OpenLibrary: Welcome to Open Library (Open Library). <http://openlibrary.org/> (28.6.2012).
- Oßwald, Achim, Sabine Scheffel u. Heike Neuroth: Langzeitarchivierung von Forschungsdaten. Einführende Überlegungen. In: Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Hrsg. von Heike Neuroth. Boizenburg: vvh Hülsbusch 2012. S. 13–21. <http://nbn-resolving.de/urn:nbn:de:0008-2012031401> (29.10.2012).
- Renfrew, Colin, Paul Bahn u. Helmut Schareika: Basiswissen Archäologie. Theorien, Methoden, Praxis. Darmstadt: WBG (Wiss. Buchges) 2009.

- Schiemann, Bernhard, Martin Oischinger u. a.: Erlangen CRM / OWL - An OWL DL 1.0 implementation of the CIDOC Conceptual Reference Model (CIDOC CRM). <http://erlangen-crm.org/> (29.6.2012).
- Smith, Sam: Open Definition. <http://opendefinition.org/> (1.7.2012).
- The Registry: ISBD Elements. <http://iflstandards.info/ns/isbd/elements/> (28.6.2012).
- Universitätsbibliothek Dortmund: Offene bibliographische Daten an der Universitätsbibliothek Dortmund. <http://data.ub.tu-dortmund.de/projekte/offene-daten/> (11.5.2012).
- Universitäts- und Stadtbibliothek Köln: Universitäts- und Stadtbibliothek Köln. <http://www.ub.uni-koeln.de/> (26.6.2012).
- Wikipedia: XSL Transformation. http://de.wikipedia.org/w/index.php?title=XSL_Transformation&stableid=99612251 (1.7.2012).