



Master Thesis

im Rahmen des

Universitätslehrganges „Geographical Information Science & Systems“
(UNIGIS MSc) am Interfakultären Fachbereich für Geoinformatik (Z_GIS)

der Paris Lodron Universität Salzburg

zum Thema

Archeontology

**Semantische Modellierung georeferenzierter
archäologischer Daten**

vorgelegt von

Stefan Dziwis

104060, UNIGIS MSc Jahrgang 2015

Zur Erlangung des Grades

„Master of Science (Geographical Information Science & Systems) – MSc(GIS)“

Augsburg, 20. Februar 2018

Danksagung

Für die externe Betreuung meiner Thesis herzlichsten Dank an Prof. Dr. Sabine Timpf und Johannes Schwer, sowie dem gesamten UNIGIS Team. Darüber hinaus ganz besonderen Dank für die tatkräftige Unterstützung an Maria Theodoridou vom Institute of Computer Science der Foundation for Research and Technology - Hellas (FORTH)

Zuletzt meiner lieben Frau und meinem kleinen Sohn für die unglaubliche Geduld, den Beistand und die Entbehrungen in den letzten Monaten.

Zusammenfassung

Digitalisierung und Internet verschaffen uns heute eine nahezu unüberschaubare Flut an Informationen. Um an bestimmte Informationen zu gelangen benötigt man spezielle Daten die grundsätzlich in Abhängigkeit von Ansicht, Absicht, Kenntnis und Zweck des Urhebers generiert werden. Im archäologischen Kontext werden in den meisten Fällen projektspezifische und proprietäre Daten generiert und nur innerhalb von Projekten Dokumentationsstandards eingehalten. Daraus resultiert im Bereich des Kulturerbes eine große Anzahl digitaler Datensätze die einen integrierten Zugriff auf heterogene und autonome Informationsquellen erfordern. Die Integration von Daten ist aus Sicht von Archäologen in der Regel mit vielen technischen Hürden und großem Aufwand verbunden. Der hohe Anspruch der technischen Umsetzung ist ein wesentlicher Punkt dem sich die meisten Domänenexperten des kulturellen Erbes gegenüber konfrontiert sehen.

Die Arbeit präsentiert das Vorgehen zur Umsetzung einer semantischen Modellierung der bisher größten archäologischen Datenbank präkolonialer Siedlungen im Amazonasbecken basierend auf der Ontologie des CIDOC CRM. Bisher enthält die Datenbank 948 georeferenzierte und datierte Fundstätten. Das Mapping der Terra Preta Datenbank dient als Grundlage für den standardisierten Austausch von Daten des brasilianischen Kulturerbesystems (Sistema Nacional do Patrimônio Cultural [SNPC]). Der Mappingprozess findet mehrstufig statt und beginnt mit der Normalisierung der relationalen Daten in XML. Die Webapplikation Mapping Memory Manager (3M) des Institute of Computer Science der Foundation for Research and Technology bietet den technischen Rahmen für die Erstellung eines Mappings von relationalen Datenbanken für Archäologen. Zur Abbildung des TerraPreta-Quellschemas ist ein Komplex aus mehreren kompatiblen Erweiterungen des CRM nötig, die als Applikationsontologie dienen. Anschließend folgt für jede Relation die konzeptionelle Festlegung der semantischen Korrespondenzen zwischen der TerraPreta-Datenbank und des kontrollierten Vokabulars des CIDOC CRM.

Aufgrund seiner komplexen Eigenschaften und Vielzahl an Entitäten ist das CIDOC CRM für die spezielle Modellierung der Semantik von Artefakten im archäologischen Kontext hervorragend geeignet. Auch die versteckten und allgemeinen Kontextinformationen, wie sie häufig bei Fachdaten des Kulturerbes auftreten, können mit geeigneten Objekttypen modelliert werden. Die Integration der Geoinformationen,

sowohl in ihrer numerischen als auch kodierten Repräsentation, ist mit den vorhandenen Klassen problemlos möglich.

Abstract

Today, digitisation and the World Wide Web provide an almost unmanageable flood of information. In order to get to certain information you need special data which are basically generated depending on the view, intention, knowledge and purpose of the author. In an archaeological context, in most cases project-specific and proprietary data are generated and are only maintained within projects documentation standards. This results in a large number of digital data sets in the area of cultural heritage that require integrated access to heterogeneous and autonomous sources of information. From the point of view of archaeologists, the integration of data is usually associated with many technical hurdles and great effort. The high standards of technical implementation are a key point facing most domain experts in the cultural heritage.

This Thesis presents the procedure for the implementation of a semantic modelling of the largest archaeological database of precolonial settlements in the Amazon Basin based on the ontology of CIDOC CRM. So far, the database contains 948 georeferenced and dated sites. The mapping of the TerraPreta database serves as basis for the standardized exchange of data of the Brazilian Kulturerbesystems (Sistema Nacional do Patrimônio Cultural [SNPC]). The mapping is multi-level and begins with the normalization of the relational data in XML. The Web application Mapping Memory Manager (3M) of the Institute of Computer Science of the Foundation for Research and Technology provides the technical framework for the creation of a mapping of relational databases for archaeologists. To figure out the TerraPreta source schema, a complex of several compatible enhancements of the CRM is required, which serve as an application ontology. The conceptual definition of the semantic correspondence between the TerraPreta database and the controlled vocabulary of the CIDOC CRM follows for each relation.

Due to its complex properties and multitude of entities, the CIDOC CRM is ideal for the special modeling of the semantics of artifacts in the archaeological context. Also the hidden and general contextual information, which often occurs in cultural heritage data, can be modelled with suitable object types. The integration of geo-information, both in its numerical and encoded representation, is easily possible with the existing classes.

Erklärung der eigenständigen Abfassung der Arbeit

Ich versichere, diese Master Thesis ohne fremde Hilfe und ohne Verwendung anderer als der angeführten Quellen angefertigt zu haben, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt worden ist. Alle Ausführungen der Arbeit, die wörtlich oder sinngemäß übernommen wurden, sind entsprechend gekennzeichnet.

Augsburg, 20. Februar 2018

Stefan Dziwis

Inhalt

Danksagung	I
Zusammenfassung	II
Abstract	IV
Erklärung der eigenständigen Abfassung der Arbeit	V
Inhalt	VI
Abbildungsverzeichnis	VIII
Tabellenverzeichnis	X
Abkürzungsverzeichnis	X
1. Einleitung	1
1.1. Motivation	1
1.2. Ziele	3
1.3. Struktur der Arbeit	4
2. Grundlagen	5
2.1. TerraPreta Datenbank	5
2.1.1. Aufbau und Datenhaltung der Datenbank	7
2.2. Ontologien	12
2.2.1. Dublin Core Element Set – DCES	14
2.2.2. Europeana Data Model – EDM	14
2.2.3. CIDOC Conceptual Reference Model - CIDOC CRM	15
2.3. Auswahl der geeigneten Ontologie	21
3. Methoden und Tools	26
3.1. Schema Ontologie Mapping	26
3.2. RDF - Resource Description Framework	28
3.3. 3M – Memory Mapping Manager	31
3.4. Effective Provider Schema	40
4. Modellierung	41
4.1. XML-Syntax - Normalisierung	42
4.2. Aufbau der TerraPreta Applikationsontologie	43
4.3. Mapping Definition	44
4.3.1. Mapping der Hauptrelation <findings>	46
4.3.2. Mapping der Hilfsrelation <sites>	50
4.3.3. Mapping der Hilfsrelation <archaeologists>	53
4.3.4. Mapping der Hilfsrelationen <traditions> und <cultures>	55
4.3.5. Mapping der Hilfsrelation <bib>	57
4.3.6. URI-Generierung	57
5. Ergebnisse und Diskussion	61
5.1. Ergebnisse	61
5.2. Diskussion	64

6. Konklusion	72
6.1. Ausblick in ein Linked Open Data Ökosystem	72
6.2. Fazit.....	75
Literaturverzeichnis.....	77
Anhang	

Abbildungsverzeichnis

Abbildung 1: Präkoloniale Fundplätze der TerraPreta Datenbank (Printscreen).....	6
Abbildung 2: Quellschema als Entity Relationship Model der TerraPreta Datenbank (Schwer, 2016)	7
Abbildung 3: Objektbrowseransicht des padminIII der TerraPreta Datenbank (Printscreen)	10
Abbildung 4: Kategorisierung von Ontologien (Guarino, 1998, S. 9, Abb. 4)	13
Abbildung 5: CRM Core Ontologie mit Erweiterungen (FORTH ICS, 2007)	16
Abbildung 6: Auswahl an Eigenschaften der Entität E22 Man Made Object (Printscreen)	19
Abbildung 7: Auswahl der range in Abhängigkeit der Auswahl der Eigenschaft (Printscreen).....	19
Abbildung 8: Auswahl der Klassenhierarchie des CIDOC CRM. (Nussbaumer & Haslhofer, 2007, S. 7, Abb. 2)	20
Abbildung 9: Dublin Core Elemente (Kokkelink & Schwänzl, 2002)	22
Abbildung 10: Vergleich DCES und CIDOC CRM (Hohmann & Fichtner, 2015, S. 123, Abb. 6)	23
Abbildung 11: EDM Klassenhierarchie (Olensky, 2010)	24
Abbildung 12: Klassifikationsschema zu database-to-ontology mapping Ansätzen (Ghawi & Cullot, 2007).....	27
Abbildung 13: Struktur der X3ML Definitions Datei (Minadakis, et al., 2015, S. 6, Abb. 2)	31
Abbildung 14: Basis Mapping Schema (Kondylakis, Doerr, & Plexousakis, 2006, S. 1, Abb. 1).....	33
Abbildung 15: Ausschnitt aus der Matching Table des 3M (Printscreen)	34
Abbildung 16: 3M Mapping Workflow verändert nach (Oldman, Theodoridou, & Samaritakis, 2016) ..	36
Abbildung 17 Source und Target Schema Visualizer des 3M. (Printscreen)	37
Abbildung 18: URI Generierungsschema (Spanos, Periklis, & Mitrou, 2012, S. 14, Tab. 3)	39
Abbildung 19: Semantisches ER-Modell der TP-Datenbank (Schwer, 2016)	41
Abbildung 20: Mapping der Hauptrelation <findings> (Printscreen).....	46
Abbildung 21: Mapping der Hilfsrelation <sites> (Printscreen)	50
Abbildung 22: Mapping der Hilfsrelation <archaeologists> (Printscreen)	53
Abbildung 23: Mapping der Hilfsrelation <traditions> (Printscreen).....	55
Abbildung 24: Mapping der Hilfsrelation <cultures> (Printscreen)	55
Abbildung 25: Mapping der Hilfsrelation <bib> (Printscreen)	57
Abbildung 26: URI-Instanzgenerator der <i>domain</i> -Zielentität des Mappings <traditions> (Printscreen) ..	59
Abbildung 27: URI-Instanzgenerator der relationsäquivalenten Entität im Mapping <findings> (Printscreen)	60
Abbildung 28: URI-Instanzgenerator der domain-Zielentität des Mappings <archaeologists> (Printscreen)	61
Abbildung 29: Alternatives Mapping der Hauptrelation <sites> (Printscreen).....	67

Abbildung 30: Raum-Zeit-Volumen Konzept der CRMgeo Erweiterung (Hiebel, Doerr, Hanke, & Masur, 2014, S. 564, Abb. 3).....	68
Abbildung 31: Auszug aus der Fundstättendatenbank des SGPA mit CNAS Identifikatoren für Amazonien (WEB10).....	72
Abbildung 32: Mapping Workflow verändert nach (Doerr, Theodorou, Aspöck, & Masur, 2015, S. 4, Abb. 2)	74
Abbildung 33: Architektur der ARIADNE e-Infrastruktur (Meghini, 2017, S. 10, Abb. 1)	75

Tabellenverzeichnis

Tabelle 1: Konvention zur Benennung von Klassen und Eigenschaften der Kernontologie sowie deren Erweiterungen.....	17
Tabelle 2: Verwendete Ontologien und Versionen.....	43
Tabelle 3: Äquivalenzen zwischen CIDOC-CRM und relationaler Datenbank (Crofts N. , 1999).....	45
Tabelle 4: optimierter Auszug ausgewählter Daten zur Fundstelle Aviso I.....	64
Tabelle 5: Entitäten mit identischen Identifikatoren	71

Abkürzungsverzeichnis

CIDOC	Comité international pour la documentation
CRM	Conceptual Reference Model
CSNA	Cadastro Nacional de Sítios Arqueológicos
SGPA	Sistema de Gerenciamento do Patrimônio Arqueológico
IBGE	Instituto Brasileiro de Geografia e Estatística
IPHAN	Instituto do Patrimônio Histórico e Artístico Nacional
RDF	Resource Description Framework
UML	Unified Modeling Language
OWL	Web Ontology Language
EDM	Europeana Data Model
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
DCES	Dublin Core Element Set
W3C	World Wide Web Consortium
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UUID	Universally Unique Identifier
EWKT	Extended Well Known Text
FRBR	Functional Requirements for Bibliographic Records models
PRONAPA	Programa Nacional de Pesquisas Arqueológicas

1. Einleitung

1.1. Motivation

Digitalisierung und Internet verschaffen uns heute eine nahezu unüberschaubare Flut an Informationen. Um an bestimmte Informationen zu gelangen benötigt man spezielle Daten die grundsätzlich in Abhängigkeit von Ansicht, Absicht, Kenntnis und Zweck des Urhebers generiert werden. Dies thematisiert die Kernproblematik hinsichtlich der Interpretationsmöglichkeiten von Daten die nicht mehr für ihren ursprünglichen Zweck verwendet werden. Gerade im archäologischen Kontext werden in den meisten Fällen projektspezifische und proprietäre Daten generiert und nur innerhalb dieses Projekts bestimmte Dokumentationsstandards eingehalten. In aller Regel werden diese Daten ausgewertet und hinsichtlich einer oder mehreren bestimmten Fragestellungen synthetisiert und abschließend in einer Publikation präsentiert. Deshalb kann bei der Datenerhebung und Dokumentation von Standards abgesehen werden, da der Wert der digitalen Daten nach der Publikation verfällt. Abgesehen von der Publikation ist eine Weiterverwendung der Daten für andere Projekte meist nicht vorgesehen. Dadurch wird der Zugang der digitalen Daten zu anderen Projekten mit ähnlichem Forschungsgegenstand nahezu ausgeschlossen, obwohl die Weiterverwendung relevanter Informationen aus abgeschlossenen Projekten eine gängige Methode der archäologischen Forschung darstellt. Beispielsweise stellt die Literaturanalyse zu einem bestimmten Thema einen existentiellen Teil im Vorfeld jedes neuen Projekts dar, bei der jedoch nur auf die publizierten Ergebnisse und nicht auf die digitalen Daten selbst zugegriffen werden kann. Meist beinhalten diese neben den vorgelegten Resultaten noch mehr Informationen, die aus Mangel an technischen Mitteln oder Schwerpunkt der eigentlichen Forschungsfrage in den Ergebnissen nicht präsentiert werden konnten. Im schlechtesten Fall, und keineswegs ungewöhnlich, werden diese Daten erst zu einem deutlich späteren Zeitpunkt für andere Projekte wieder interessant. Die projektspezifischen Kenntnisse um die Erhebung und Haltung der Daten ist dann aber längst in Vergessenheit geraten (Becker & Lins, 2013).

Eines dieser Projekte stellt die Analyse historischer Siedlungen im Amazonasbecken dar, für das im Zuge einer wissenschaftlichen Arbeit die bisher größte publizierte Sammlung von georeferenzierten präkolonialen Siedlungsorten im Amazonasgebiet erstellt wurde (von Groote-Bidlingmaier, 2016). Diese Daten werden in einer

objektrelationalen Datenbank vorgehalten und können derzeit unter der Webseite TerraPreta WebGIS (WEB1) der geoinformatics group der Universität Augsburg teilweise eingesehen und abgerufen werden. Zudem betreibt das *Instituto do Patrimônio Histórico e Artístico Nacional* (IPHAN) einen Service des Sistema de Gerenciamento do Patrimônio Arqueológico (SGPA) in dem in Form des *Cadastro Nacional de Sítios Arqueológicos* (CNSA) (WEB2) eine weitere Datenbank mit bekannten historischen Siedlungsorten weiter Teile des Amazonasbeckens besteht. Jedoch sind die dort gespeicherten Orte nicht georeferenziert.

Um diese Art von Daten dauerhaft und nutzbar im World Wide Web zur Verfügung zu stellen, gibt es derzeit internationale Bestrebungen, wie z.B. das Ariadne Projekt (Meghini, et al., 2017) oder OpenInfRA (OpenInfRA, 2015), die sich der freien Verfügbarkeit von archäologischen Daten und deren nachhaltige Sicherung im Internet sowie deren Vernetzung zum Ziel gesetzt haben. Zur Lösung dieses Problems wurden auch für die Archäologie in den letzten Jahrzehnten einheitliche Datenmodelle entwickelt. Allen voran das vom *Comité international pour la documentation* (CIDOC) vorgestellte *Conceptual Reference Model* (CRM), das seit 2006 offizieller ISO Standard ist, hat in der Folgezeit durch eine Special Interest Group zunehmend die Integration von georeferenzierten Fachdaten mit Kulturerbeinformationen mit seinen Erweiterungen weiterentwickelt.

Die Nutzarmachung der heterogenen Datenbestände beginnt in aller Regel in ihrer Normalisierung und ihrer Anreicherung mit bedeutungsgleichen semantischen Einheiten oder Begriffen (Becker & Lins, 2013). Die Abbildung des lokalen Datenbestandes, zum Beispiel in einer relationalen Datenbank auf ein vorgegebenes Datenmodell, wird als Mapping bezeichnet (Hohmann, 2008). Mapping gilt bisher als zeitraubender und schwieriger Prozess mit hohem Fehlerpotenzial (Binding, May, & Tudhope, 2008), da die Ontologie im Gegensatz zur Datenbank auch die semantischen Beziehungen der gespeicherten Daten beschreiben kann (Bumanns, 2010). Ausgehend von einer grundsätzlich heterogenen Datenrepräsentation in archäologischen Projekten lassen sich mit Hilfe solcher Modelle verschiedenste Datensätze in eine einheitliche und vergleichbare Struktur integrieren. So wird einerseits eine gewisse Nachhaltigkeit hinsichtlich der Datenhaltung aber auch bei der Wiederverwendung der digitalen Daten erreicht, deren Erhebung grundsätzlich mit Kosten verbunden ist. Gerade deshalb kann eine standardisierte Datenhaltung auch aus finanzieller Sicht sinnvoll sein. Die Integration von Daten ist aus Sicht der

Archäologen in der Regel mit vielen technischen Hürden und großem Aufwand verbunden. Der hohe Anspruch der technischen Umsetzung ist ein wesentlicher Punkt dem sich nicht nur Archäologen gegenüber konfrontiert sehen. So ist es bisher nicht ohne weiteres möglich ohne hohen Aufwand von IT-Experten die Integration und somit eine Interoperabilität der entstandenen heterogenen Datensätze zu verwirklichen.

1.2. Ziele

Das primäre Ziel der vorliegenden Arbeit ist ein semantisches Modell zur Transformation zwischen den objektrelationalen Daten der TerraPreta-Datenbank und der Ontologie des CIDOC CRM zu konzipieren. Mit Hilfe eines auf der Ontologie des CIDOC CRM basierenden kontrollierten Vokabulars soll dieses Modell als Grundlage für den standardisierten Austausch von Daten dienen. Ferner soll dies mit Hilfe einer Webapplikation geschehen, die den technischen Rahmen zur Erstellung eines Mappings für Domänenspezialisten, wie Archäologen, unterstützt. Das Ergebnis der Arbeit stellt ein semantisches Modell der TerraPreta Datenbank vor und eine vollständig ausformulierte RDF-Datei mit den in der TerraPreta Datenbank enthaltenen Daten dar.

Demzufolge fokussiert sich die vorliegende Arbeit auf folgende Fragestellungen:

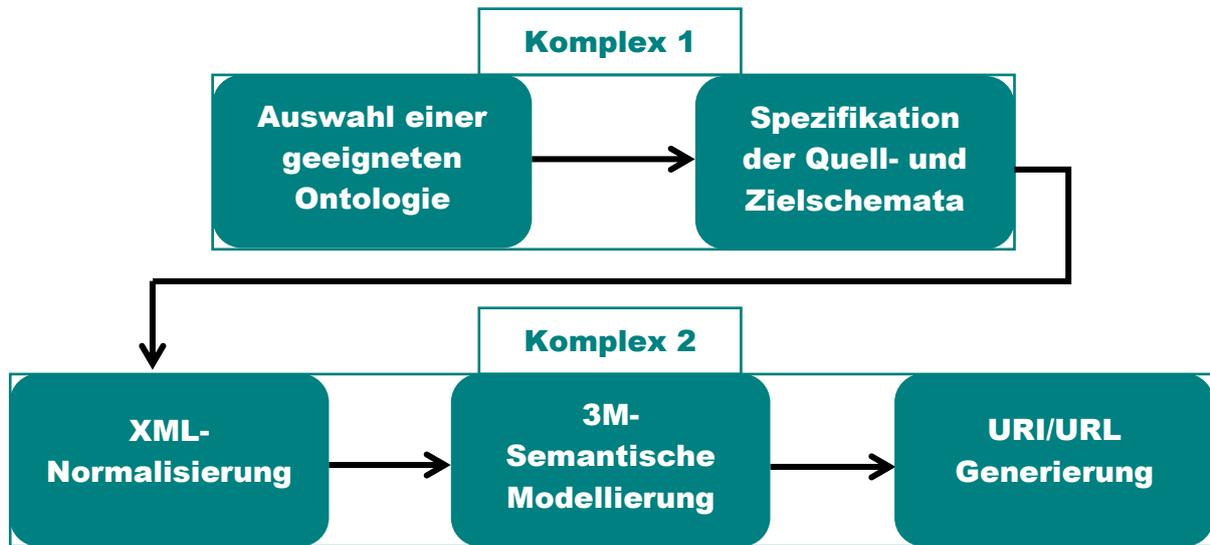
Komplex 1: Welche Ontologie eignet sich zur Modellierung der Semantik der relationalen Daten der TerraPreta Datenbank?

Wie erfolgt die Umsetzung auf technischer Ebene?

Komplex 2: Wie kann die semantische Modellierung umgesetzt werden?

Kann die Semantik der TerraPreta Daten mit dem CIDOC CRM sinnvoll abgebildet werden?

1.3. Struktur der Arbeit



Die vorliegende Arbeit ist in fünf nachfolgende Kapitel aufgebaut.

Im anschließenden Kapitel Zwei wird der aktuelle Stand der TerraPreta Datenbank, sowie die wichtigsten Ontologien für eine kulturerbespezifische Datenmodellierung vorgestellt. Dazu gehört der wissenschaftliche Hintergrund und der grundlegende Aufbau der TerraPreta Datenbank, ebenso wie die Speicherung der Geodaten und der Aufbau und die Verwendung von Ontologien, im speziellen die des CIDOC CRM sowie ein kurzer Vergleich zu anderen Ontologien, die auf Kulturerbeinformationen ausgelegt sind.

Darauf aufbauend werden in Kapitel Drei die methodischen und technischen Grundlagen behandelt die für die Bearbeitung der Thesis notwendig sind. Dies schließt die Repräsentation des CIDOC CRM in Form des Resource Description Frameworks (RDF) und eine Erläuterung zu den technischen Voraussetzungen und Datenformatierung für die Weiterverwendung der Daten mit Hilfe eines Mapping Tools und dessen Verwendung ein.

In Kapitel Vier wird die konkrete Umsetzung der Modellierung beschrieben. Dies erfordert die Beschreibung der notwendigen Bereitstellung der gespeicherten Daten, sowie dem eigentlichen Modellierungsvorgang auf die Ontologie des CIDOC CRM und seinen Erweiterungen.

Das anschließende Kapitel Fünf stellt die wesentlichen Erkenntnisse der bisherigen Arbeit noch einmal kurz und prägnant heraus und diskutiert Anwendungsmöglichkeiten und Einschränkungen, die sich durch die Modellierung der relationalen Daten ergeben. Kapitel Sechs stellt einen Ausblick hinsichtlich der Verwendung der TerraPreta Daten im Semantic Web vor und behandelt im Fazit einige Aspekte als abschließende Betrachtung.

2. Grundlagen¹

2.1. TerraPreta Datenbank

Die TerraPreta-Datenbank liegt in Form des Open-Source-Datenbankmanagementsystems PostgreSQL (WEB3) vor. Durch die Möglichkeit zur PostGIS-Erweiterung (WEB4) können auch Geodaten verarbeitet werden. Mit der Verknüpfung von Daten mit Raumbezügen findet diese Kombination als Geodatenbanksystem als kostengünstige Variante hinsichtlich der Speicherung und Verwaltung von Daten in der Archäologie weite Verbreitung in Projekten. Zudem ist die PostgreSQL/PostGIS-Geodatenbank mit den gängigsten proprietären und quelloffenen Geoinformationssystemen wie beispielsweise ArcGIS oder QGIS problemlos kompatibel. Beide unterstützen das Speichern von geographischen Informationen nach der *Simple Feature Access* Spezifikation der OGC.

Die Datenbank ist die derzeit größte digitale Sammlung georeferenzierter und datierter präkolonialer Fundplätze, die sich über das gesamte Amazonasbecken und darüber hinaus verteilen. Sie ist das Resultat einer Dissertation zur Entwicklung eines daten-gesteuerten Ansatzes zu Betrachtung von funktionalen Siedlungsmuster in einem umweltarchäologischen Kontext (von Groote-Bidlingmaier, 2016).

¹ Hinweis zur Formatierung des Textes:

Für ein besseres Verständnis sind die <blau> hinterlegten Begriffe in spitzen Klammern im folgenden Text direkt mit den Wurzelementen XML-Dateien im Anhang und umgekehrt verlinkt.

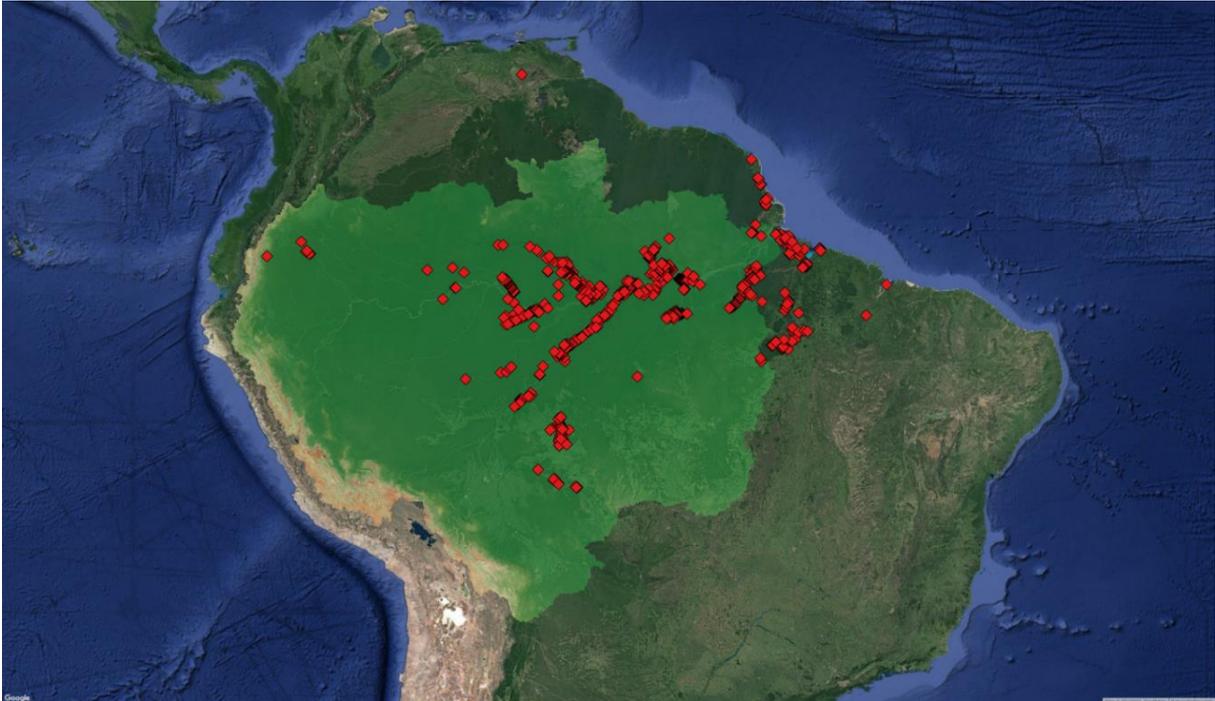


Abbildung 1: Präkoloniale Fundplätze der TerraPreta Datenbank (Printscreen)

Bisher enthält die Datenbank 948 georeferenzierte und datierte Fundstätten. In aufwändiger Einzelarbeit wurden aus zahlreichen wissenschaftlichen Artikeln oder Berichten die Standorte der Siedlungen entweder in Form von Koordinaten oder aus Beschreibungen oder Karten abgeleitet. Unterstützt wurde die Zuweisung der Standorte durch die langjährige Erfahrung der örtlichen Wissenschaftler (von Groote-Bidlingmaier, 2016). Allein das Amazonasbecken - in der Abbildung 1 grün hinterlegt - hat eine Fläche von 6,915 Millionen Quadratkilometern. Die Erforschung des Amazonasbeckens hat eine lange Geschichte die bereits im 16. Jahrhundert ihren Ursprung hat (Hilbert, 1968). Ausgehend von den vielen unterschiedlichen ausgewerteten Quellen liegt die Qualität der Koordinaten deshalb nicht an allen Fundstätten mit der gleichen Präzision vor. Die Datenbank enthält Artefakte die zur chronologischen Einordnung der Fundplätze dienen und kontextbezogene Daten zu den Siedlungen, wie z.B. Koordinaten und bekannte Ausdehnung naturwissenschaftliche und kulturhistorische zeitliche Einordnung.

Für die Anwendung als interaktives WebGIS wurde die Datenbank optimiert (von Groote-Bidlingmaier, Hilbert, Schwer, & Timpf, 2015) und für den Anwendungsfall dieser Arbeit auf fünf Tabellen reduziert.

2.1.1. Aufbau und Datenhaltung der Datenbank

Abbildung 2 zeigt die für den Anwendungsfall reduzierte Version der Datenbank als Quellschema, für das eine semantische Modellierung erzielt werden soll. Deshalb liegt der Fokus auf einer präzisen Beschreibung des Inhalts der Datenbank. Jedoch sind einige wenige Details auf technischer Ebene für das Mapping von Bedeutung. Im Anhang befindet ein aus den TerraPreta Daten erstellter Testdatensatzes aus der TP-Datenbank im XML-Format, der als Testdatensatz für die Migration in das semantische Modell dient.

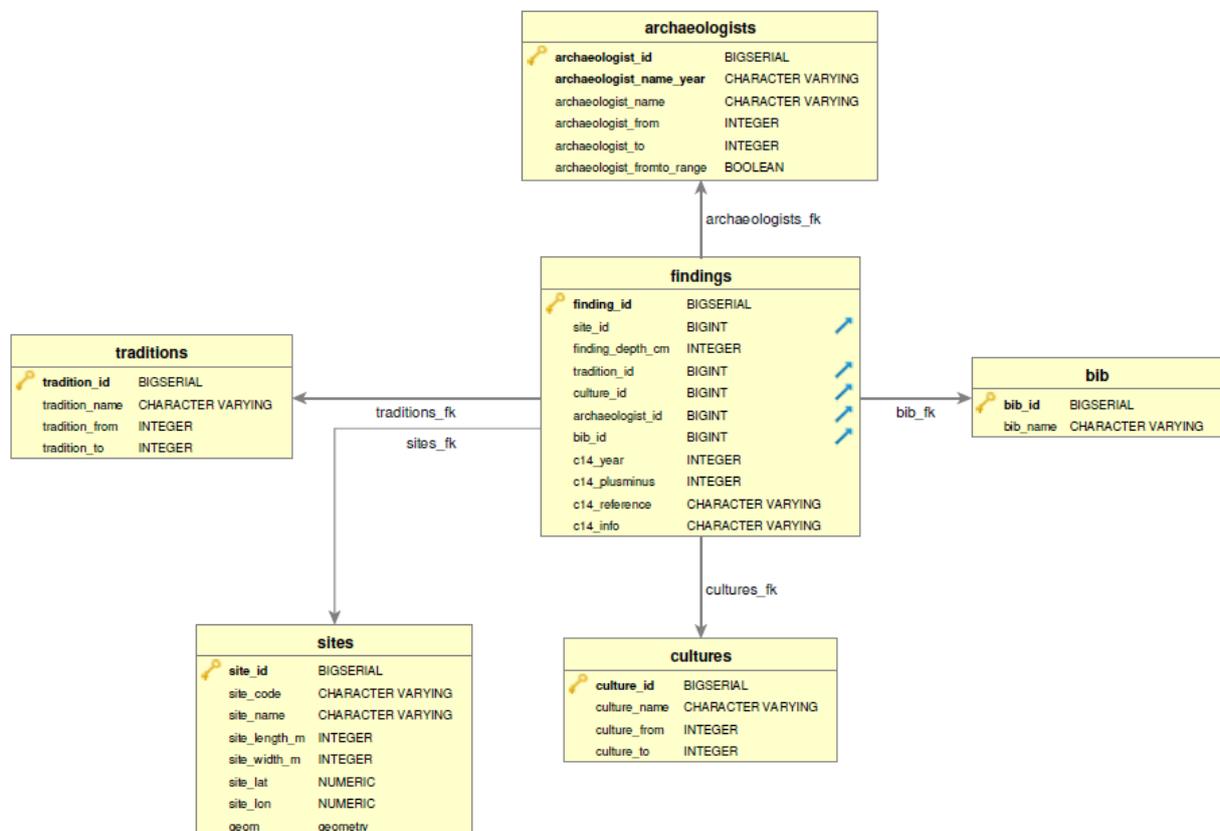


Abbildung 2: Quellschema als Entity Relationship Model der TerraPreta Datenbank (Schwer, 2016)

Die TerraPreta Datenbank besteht demnach aus der Hauptrelation findings und den fünf Hilfsrelationen traditions, sites, archaeologists, cultures und bib. Jeder Zeileneintrag aller Relationen erhält seine eindeutige Identifikation durch einen inkrementellen Primärschlüssel (PK), benannt nach dem englischen Singular der Relation selbst kombiniert mit dem Suffix _id. Die Hauptrelation findings enthält neben ihrem eigenen

Primärschlüssel auch Fremdschlüssel (FK), die auf die Identifikatoren aller Hilfsrelationen verweisen.

In der reduzierten Version der Datenbank repräsentiert die Tabelle [<findings>](#) einzelne Keramikscherben die es ermöglichen die Siedlungsplätze nach Tradition und deren Kulturen zu unterscheiden. Darüber hinaus finden sich in der Tabelle metrische Informationen zur Fundtiefe Artefakte in Zentimetern. Die Dokumentation der Fundtiefe, also der Messung der Entfernung zwischen dem oberen Rand der zum Fundzeitpunkt natürlich gewachsenen Oberfläche und dem exponiertesten Teil des Artefakts, ist als Bestandteil der stratigraphischen Dokumentation einer Grabung gängige Praxis. Darüber hinaus enthält die Tabelle Informationen zu konventionellen unkalibrierten Radiocarbon-Datierungen (von Groote-Bidlingmaier, 2016). Die Informationen sind in den Spalten `c14_year`, `c14_plusminus`, `c14_reference` und `c14_info` enthalten. Die Spalte `c14_year` enthält das Jahr der eigentlichen Datierung, `c14_plusminus` enthält die Streuungsangabe der Messung in Jahren bezogen auf das Datierungsjahr. `c14_reference` enthält die Angaben der Bezugszeit für die Kohlenstoffdatierung `c14_year`. `c14_info` liefert in Form einer Abkürzung weitere Informationen zur C14-Probe selbst und dem Labor in dem die Probe analysiert worden ist (von Groote-Bidlingmaier, 2016).

Die Hilfsrelation [<archaeologists>](#) repräsentiert Archäologen die Ausgrabungen an den jeweiligen Fundstätten durchgeführt bzw. daran teilgenommen haben. Allerdings sind neben dem Namen des Grabungsverantwortlichen `archaeologists_name` nur die Spalten `archaeologists_from` und `archaeologist_to` relevant für die Modellierung, da sie die Information über das Jahr des Beginns und dem Ende der Ausgrabung angeben an der die Person teilgenommen hat.

Die Relation [<bib>](#) enthält rudimentäre Daten über die Publikation zu den Funden und der jeweiligen Siedlungsstätte.

Der Begriff Tradition bezeichnet das übergeordnete Niveau ähnlicher Keramik-Stile, die wiederum in mehrere Kulturen unterteilt ist (von Groote-Bidlingmaier, 2016). Die Beziehung zwischen Kulturen und Traditionen ist hierarchisch, was bedeutet, dass eine Kultur genau eine Tradition hat, eine Tradition hingegen aber in der Regel mehreren Kulturen zugeordnet ist (von Groote-Bidlingmaier, 2016). Es wird darauf hingewiesen, dass die Begriffe Kultur und Phase als Synonyme verwendet werden, obwohl die Bedeutung aus archäologischer Sicht nicht identisch ist (von Groote-Bidlingmaier, 2016). Der heutigen archäologischen Auffassung zufolge umfasst der Begriff Kultur

tradierte, also gemeinsame Verhaltensweisen und Vorstellungen einer Gruppe (Trachsel, 2008). Ein gemeinsames Verhalten wird auch der vorliegenden Chronologie zu Grunde gelegt. Dennoch umfasst das Spektrum des Begriffs weit mehr als nur die Ausprägungen von Keramik, wie überregionale, regionale und lokale Siedlungsmuster, Subsistenzstrategien oder Belegzeiten der Siedlungsplätze um nur einige zu nennen. Mehr als die relativchronologischen Daten der Keramik liegen aber in der Regel zu den meisten Fundplätzen nicht vor (Neves, 2008). Daher ist der Ansatz unabhängig vom archäologischen Diskurs über die Kategorisierung der Funde zu betrachten (von Groote-Bidlingmaier, 2016). Dennoch repräsentieren die verschiedenen Traditionen und Kulturen eine Art künstlerischen Stil der sich durch den historischen Kontext begründen lässt.

Demnach enthält die Tabelle [<traditions>](#) bisher ermittelte absolute Daten zu den relativchronologischen Keramikstilen des Amazonasgebietes (Neves, 2008). Die Bezeichnung des Stils spiegelt sich in der Spalte `tradition_name` wieder. Dazu gehört der in absoluten Daten angegebene Existenzzeitraum der jeweiligen Tradition mit Beginn und Ende, sofern dieser durch Literatur oder Radiocarbonatierung nachgewiesen werden kann.

Aus keramisch chronotypologischer Sicht enthält die Tabelle [<cultures>](#) Metadaten über die Bezeichnung der am Siedlungsplatz vergesellschafteten keramischen Subkategorien in Form von `culture_name`, sowie den in absoluten Daten angegebenen Existenzzeitraum der jeweiligen Kultur mit Beginn und Ende, sofern dieser durch Literatur oder Radiocarbonatierung nachgewiesen werden kann.

Die Hilfsrelation [<sites>](#) enthält Metadaten über die Fundplätze in Form des `site_code`. Ähnlich wie der `site_name` ist diese Signatur abgeleitet von den Bezeichnern der Fundstätten der SGPA bzw. CNAS Datenbank des IPHAN. Informationen zur Ausdehnung der Siedlungsplätze in Metern werden repräsentiert durch `site_length_m` und `site_width_m`. Die geographischen Daten sind in Form von dezimalen WGS84-Koordinaten in der Spalte `site_lat` und `site_lon` gespeichert. In der Spalte `geom` werden die Siedlungsplätze zusätzlich als Geometrie im Extended Well Known Text (EWKT) Format gespeichert, der kein OGC-Standard, sondern ein PostGIS spezifisches Format ist, welches neben den Koordinaten den Bezeichner des räumlichen Bezugssystems enthält (PostGIS Development Group, 2005). Das EWKT Format kann diese Informationen in Form eines hexadezimal kodierten Strings ausdrücken der sowohl die reinen geometrischen Daten (Koordinaten und Geometriety) als auch das

geographische Referenzsystem der Daten kodiert. Damit sind bereits die wichtigsten Metadaten im String inhärent und können nicht verloren gehen oder verwechselt werden. Deshalb ist diese Kodierung für den interoperablen Datentransfer besonders gut geeignet. Bisher enthält die Datenbank ausschließlich Punktdaten. PostGIS unterstützt jedoch unter anderen auch Linien, Polygone, polyedrische- und TIN-Oberflächen. Die Administration einer PostgreSQL/PostGIS-Geodatenbank erfolgt im einfachsten Fall mit der Benutzeroberfläche pgAdminIII. Erhält man einen hexadezimal kodierten String, kann dieser mit einem einfachen SQL-statement in das für Menschen lesbare EWKT-Format umgewandelt und weiterverarbeitet werden.

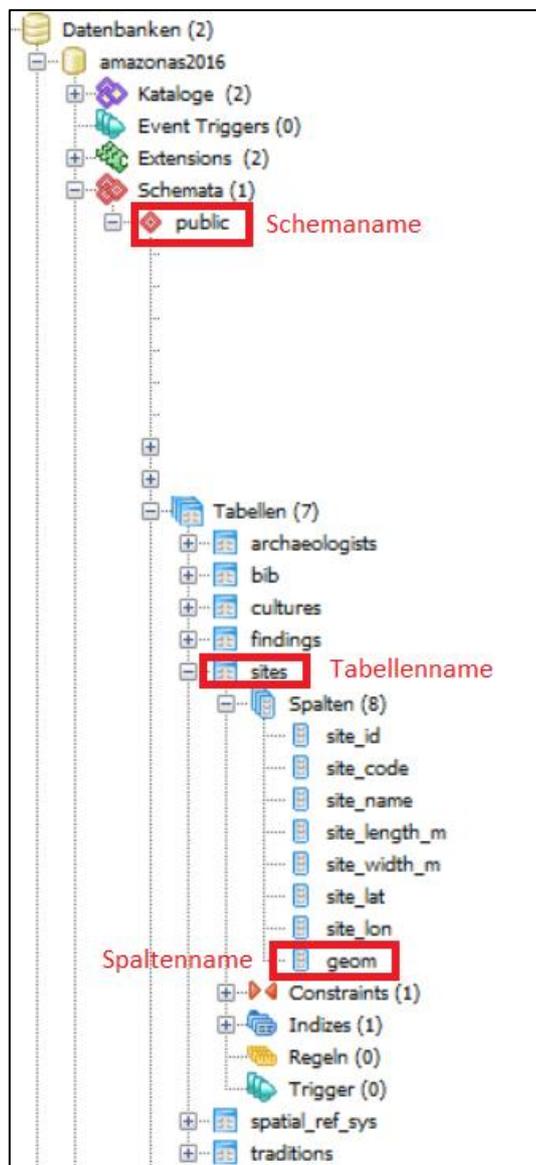


Abbildung 3: Objektbrowseransicht des pgadminIII der TerraPreta Datenbank (Printscreen)

Mit dem Befehl

```
SELECT
    ST_AsEWKT(Tabellenname.Spaltenname)

FROM
    Schemaname.Tabellenname
```

erhält man aus dem kodierten String die wesentlichen geometrischen Informationen zu einer Fundstelle.

Der Befehl im Folgenden wurde exemplarisch für die TerraPreta Datenbank durchgeführt

```
SELECT
    ST_AsEWKT(sites.geom)
FROM
    public.sites
```

und liefert aus der hexadezimal kodierten geometry-Spalte mit folgendem Inhalt

```
0101000020E6100000CAA31B6151E550C0B3446799450806C0
```

das folgende Ergebnis im EWKT-Format

```
SRID=4326;POINT(-67.583092 -2.754039)
```

Um die Transformation und Aggregation der Daten zu erreichen, ist es erforderlich gleichwertige Konzepte und Beziehungen aus einem Quellschema mit einem Zielschema zu verknüpfen. Das Quellschema einer Datenbank kann von den XML-Daten abgeleitet werden, die aus Informationssystemen exportiert werden können (Minadakis, et al., 2015). Das Zielschema lässt sich mit Hilfe einer geeigneten Ontologie definieren.

2.2. Ontologien

Ursprünglich als generischer Begriff der Philosophie wird der Terminus „Ontologie“ seit den 1990er Jahren vermehrt in den Computerwissenschaften behandelt. Dort steht er für eine allgemeingültige und explizit definierte Terminologie, mit der sich die Struktur von Daten beschreiben lässt (Neches, et al., 1991). Aus Sicht der Computerwissenschaften ist die Aufgabe einer Ontologie die Schaffung einer Grundlage für die Strukturierung und Abbildung komplexer Daten. Eigentlich als formal definiertes System von Konzepten und Relationen gedacht welches mit Hilfe von Inferenz- und Integritätsregeln und einem spezifischen Vokabular Bedingungen zur Vollständigkeit und Richtigkeit der Begriffe, sowie Beziehungen zwischen den Begriffen vorgibt (Scheuermann, 2006). Betrachtet man den Begriff im Sinne von Tim Berners-Lee's *Semantic Web* (2001), besteht der maßgebliche Unterschied zu diesen formalen Systemen in der Möglichkeit zur zusätzlichen Abbildung einer Semantik, statt der bloßen Informationen (Hohmann & Fichtner, 2015).

Ziel dieser Ontologien ist es, die Welt in einem formalen inhaltsbezogenen System abzubilden um dadurch das Semantic Web zu verwirklichen. Sie sind das Modell eines spezifischen, für das jeweilige Themengebiet angepassten Ausschnittes der Welt und dienen zur Spezifikation der Abbildung der Realität in einem Datenmodell (Scheuermann, 2006). Ontologien sind jedoch bisher nicht standardisierte Frameworks und lassen sich deshalb inzwischen, gekennzeichnet durch ihre Ausdrucksmächtigkeit, fast allen informatischen Klassifikationssystemen und Konzeptualisierungen, wie Taxonomien, UML, RDF oder OWL zuordnen (Hohmann & Fichtner, 2015; Oberle, Guarino, & Staab, 2009).

Nach Guarino (1998) lassen sich Ontologien aber auch anhand ihres Zwecks oder Bereichs differenzieren, nicht nur nach ihren formalen Spezifika, und werden deshalb in vier Kategorien grob unterteilt.

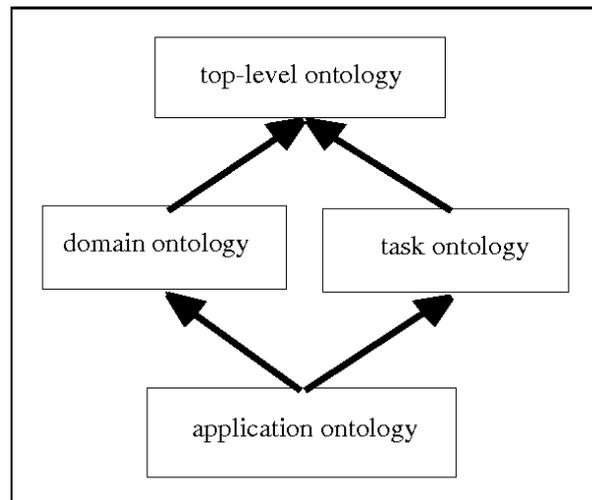


Abbildung 4: Kategorisierung von Ontologien (Guarino, 1998, S. 9, Abb. 4)

Während eine top-level-ontology nur generelle Konzepte unabhängig von einem bestimmten Wissensgebiet beschreibt, fokussieren domain und task ontologies sich auf spezielle Wissens- oder Aufgabenbereiche, wie beispielsweise Informationen zum kulturellen Erbe. Das hierarchische Ende dieser Kategorisierung stellt die application ontology dar, die für sehr spezielle Konzepte entwickelt werden und sich der Konzeption und Terminologie der übergeordneten Ontologien bedient. Grundsätzlich bestehen alle diese Konzeptionen aus drei Kernkomponenten (Hohmann & Fichtner, 2015; Spanos, Periklis, & Mitrou, 2012).

- **Konzepte**, als abstrahierte Repräsentation von Phänomenen der realen Welt die hierarchisch organisiert werden können, auch als Klasse oder Entität bezeichnet
- **Relationen**, die einen Bezug zwischen Konzepten darstellen, denen eine bestimmte Bedeutung innewohnt, auch als Eigenschaften bezeichnet
- **Restriktionen**, als Bedingungen nach denen Relationen eingesetzt werden dürfen, z.B. objektorientierte Vererbung

So haben sich aus verschiedenen Bereichen der Informationswissenschaften² einige Modelle entwickelt, die auf die Integration und Aufbereitung von heterogenen Datenbeständen für das Semantik Web zielen und hierfür den Aufbau eines Linked-

² vornehmlich Gedächtnisinstitutionen wie Bibliotheken, Museen und Archive

Open-Data-Ökosystems zur freien Verfügbarkeit von Daten zu etablieren suchen (Pohl & Danowski, 2013).

2.2.1. Dublin Core Element Set – DCES

Das Dublin Core wurde durch die *Dublin Core Metadata Initiative* (DCMI) von Wissenschaftlern und Bibliothekaren gemeinsam für die Beschreibung von Ressourcen aller Art erarbeitet. Die Benennung des „Dublin Core Element Sets“ (DCES) dienen der Beschreibung von elektronischen Textobjekten, ist leicht verständlich definiert und besteht aus 15 Elementen (Spiridonidou, 2003). Ursprünglich zur standardisierten Beschreibung von Metadaten wurde versucht das DCES als Ontologie zu nutzen (Hohmann & Fichtner, 2015).³ Allerdings wird die Verwendung von Metadaten-Schemata als eine Art Ontologie wissenschaftlich kontrovers diskutiert, da nach Doerr (2009) eine Ontologie dazu dient, eine gemeinsame Konzeptualisierung hinter mehreren Schemata zu beschreiben und deshalb damit nicht gleichgesetzt werden darf. Ferner kann die Reduzierung komplexer Objektbeschreibungen auf wenige Eigenschaften nur zu einer semantischen Überladung der wenigen Konzepte führen, die mit der zuvor diskutierten Definition einer Ontologie zur Abbildung der Realität in einem Datenmodell konfligiert.

Dennoch haben die digitalen Bibliotheken zahlreicher Digitalisierungsprojekte von kulturellen Objekten eine gemeinsame mit Dublin Core Metadaten Elementen als Minimalstandard geschaffene Basis (Doerr, 2009). Deshalb soll das DCES ebenso auf seine Eignung für das Mapping der TerraPreta Datenbank geprüft werden.

2.2.2. Europeana Data Model – EDM

Das Europeana Data Modell (EDM) ist ein weiterer Ansatz zur Modellierung von kulturellespezifischen Daten und stellt ein übergreifendes Modell als interoperablen Schirm über heterogenen Metadatenstandards (Dröge, Iwanowa, Trkulja, Hennicke, & Gradmann, 2013). Europeana nennt sich die zentrale digitale Bibliothek Europas, die im Zuge des EDLnet / EuropeanaNet Projekts geschaffen wurde um ein digitales

³ Unter Verweis auf (Nilsson, Powell, Johnston, & Naeve, 2008)

Museum, ein digitales Archiv, ein digitales audiovisuelles Archiv und eine digitale Bibliothek auf einmal zu sein (Olensky, 2010). Die Besonderheit des EDM liegt in der Unterscheidung des real existierenden kulturellen Objekts und dessen digitaler Repräsentation (Diehr, 2013).

2.2.3. CIDOC Conceptual Reference Model - CIDOC CRM

Das Comité International pour la DOcumentation (CIDOC) ist das internationale Komitee für Dokumentation des internationalen Museumsverbandes (ICOM). Seine Funktion besteht im Setzen von Standards für die Museumsdokumentation und unter anderem auch in der Verantwortlichkeit für das CIDOC CRM. Das Conceptual Reference Modell (CRM) wurde speziell für die semantische Interoperabilität von Kulturerbeinformationen ins Leben gerufen und ist der Nachfolger eines Versuchs zur Entwicklung einer standardisierten Kerndatenstruktur zur Benennung und Speicherung von kulturerbespezifischen Daten, dem CIDOC Relational Model (Doerr, 2003). Aufgrund der Heterogenität der Information und Dokumentation, die für die digitale Erfassung des kulturellen Erbes notwendig ist, konnte sich bisher kein standardisiertes Speicher- und Verwaltungssystem etablieren (Hohmann, 2011). Dementsprechend liegt der Fokus des seit 1996 entwickelten CRMs nicht mehr auf der Vorgabe des zu dokumentierenden Inhalts sowie dessen Speicherung, sondern nur noch auf der Strukturierung und dem Austausch bereits vorhandener Dokumentation (Ermert, et al., 2004). Das als „[...] *domain ontology for cultural heritage information* [...]“ bezeichnete Referenzmodell zielt besonders auf die Beschreibung von historischen und geographischen Kontextinformationen (Crofts, Doerr, & Gill, 2003), zu denen auch archäologische Daten gehören, wie die der TerraPreta Datenbank.

Seit 2006 ist die Domänenontologie des CRM offizieller ISO-Standard. Dieser wurde 2014 durch eine erweiterte Fassung des Referenzmodells (ISO 21127:2014) abgelöst. Zudem trifft sich bereits seit 1999 mittlerweile mehrmals jährlich mit CRM Special Interest Group eine Gruppe von Experten, um über spezielle Anpassungen und Erweiterungen des CIDOC CRM zu diskutieren (Crofts, Doerr, Gill, Stead, & Stiff, 2011).

So ist die Wahl des Zielschemas des CRM hinsichtlich der Modellierung des TerraPreta Quellschemas auf dessen Ontologie nicht nur durch den generischen

Ansatz des Referenzmodells, sondern auch aufgrund seiner mittlerweile hohen internationalen Akzeptanz als besonders nachhaltig zu betrachten.

2.2.3.1. Aufbau und Modell des CIDOC CRM

Das CRM ist eine Domänenontologie und dient zur Modellierung der wesentlichsten Beziehungen zur Beschreibung von Geschehnissen der Vergangenheit, die aus Sicht eines Menschen gelten (Doerr, Theodorou, Aspöck, & Masur, 2015). Die dieser Arbeit zugrundeliegende Version 6.0 besteht aus 90 Klassen (Classes) und 125 Eigenschaften (Properties) (Ore, Doerr, Le Bœuf, & Stead, 2015). Ziel des Modells ist es, den Datenaustausch im Bereich des kulturellen Erbes zu vereinfachen. Neben der physischen Beschreibung von Objekten können auch zeitliche und örtliche Informationen beschrieben werden. Hierfür bietet das CIDOC CRM erweiterbare Ontologien oder Modelle (Extensions), um neben den raumzeitlichen Phänomenen Beschreibungen zu spezielleren Sachverhalten zu ermöglichen, wie beispielsweise Phänomenen im Zusammenhang archäologischen Ausgrabungen repräsentiert durch das CRMarchaeo.

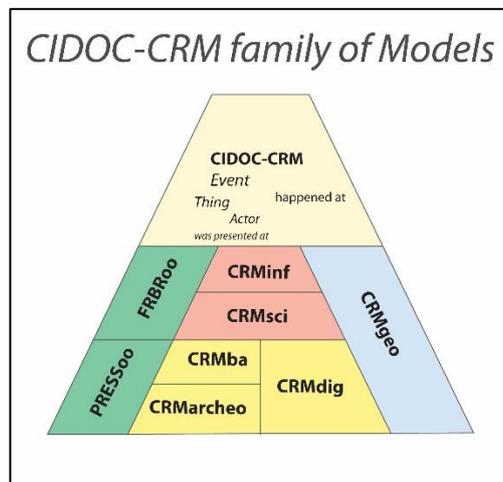


Abbildung 5: CRM Core Ontologie mit Erweiterungen (FORTH ICS, 2007)

In den Erweiterungen finden sich wiederum neue Klassen und Eigenschaften, die den Klassen der Domänenontologie hierarchisch zugeordnet sind. Abbildung 5 zeigt die wichtigsten kompatiblen Erweiterungen, darunter auch jene, die für die Modellierung des relationalen Schemas der TerraPreta Datenbank verwendet worden sind. Zur Vereinfachung wird im folgenden Verlauf der Arbeit bei der Benennung der Klassen

und Eigenschaften auf deutsche Begrifflichkeiten verzichtet, da die deutsche Dokumentation des CRMs (WEB5) lediglich die Version 5.0.1 übersetzt. Daher werden nicht alle in der Arbeit verwendeten Termini mit deutschen Begriffen abgedeckt.

2.2.3.2. Klassen und Eigenschaften

Das CRModell hat einen objektorientierten Aufbau und beschreibt abstrakte Klassen und deren Beziehung zueinander als Eigenschaften.

Klassen der Kernontologie sind mit einer Nummer und dem Buchstaben „E“ gekennzeichnet. Eigenschaften sind mit „P“ und ebenso einer Nummer deklariert. Zur besseren Unterscheidung zwischen Kern- und Erweiterungsontologie variieren die Buchstaben der Klassen und Eigenschaften wie folgt:

Tabelle 1: Konvention zur Benennung von Klassen und Eigenschaften der Kernontologie sowie deren Erweiterungen

Ontologie	Klasse	Eigenschaft
CRM	E	P
CRMsci	S	O
CRMarchaeo	A	AP
CRMgeo	SP	Q

Objektorientierung beinhaltet im Sinne der Informatik ein wesentliches Grundkonzept das für die Klassen und Eigenschaften des CRM sowie die Abbildung eines Datenbankschemas relevant sind. Dazu gehört die so genannte Vererbung, die im CRM sowohl strikt als auch mehrfach sein kann (Nussbaumer & Haslhofer, 2007).

Die oberste hierarchische Einheit der Kernontologie wird repräsentiert durch die Klasse „E1 Entity“, die als Superklasse übergeordnet über allen anderen Klassen mit Ausnahme der „E59 Primitive Value“ und deren Subklassen steht, die für unstrukturierte Daten konzipiert sind.⁴ Durch das Grundkonzept der Vererbung muss die Subklasse die deklarierten Eigenschaften von seinen Superklassen ohne

⁴ Da die Klasse E 59 Primitive Value im weiteren Verlauf keine Rolle spielt, wird an dieser Stelle nicht näher Bezug darauf genommen.

Ausnahme (strikte Vererbung) aufweisen, zusätzlich zu seinen eigenen Eigenschaften (Ore, Doerr, Le Bœuf, & Stead, 2015).

Als Beispiel hat die oberste Klasse „E1 CRM Entity“ die drei Beziehungen

„P1 is identified by“ / „P2 has type“ / P3 has note“

Alle vorhandenen Unterklassen erben diese drei Beziehungen, die somit für jede Subklasse von E1 Entity gelten.

Zudem kann ein Objekt, das abgebildet werden soll, auch mehreren Klassen der gleichen Hierarchieebene zugeordnet sein (mehrfache Vererbung), da sich nicht immer eine monohierarchische Sequenz beschreiben lässt (Ermert, et al., 2004). Klassen und Eigenschaften sind nicht formal, sondern durch ihre Charakteristika natürlichsprachlich in einer Beschreibung (*scope note*) intendiert. Eine Klasse kann keine, eine oder gleich mehrere Eigenschaften haben. Eigenschaften sind gerichtete Beziehungen, die immer zwischen einer festen Ausgangsklasse (*domain*) und einer Zielklasse (*range*) liegen. Ausgehend von einer *domain* beliebiger hierarchischer Ebene erbt die Ausgangsklasse konsequenterweise alle möglichen Eigenschaften ihrer Superklassen und besitzt darüber hinaus auch eigene. Die Bedeutung dieses Sachverhalts für die vorliegende Arbeit soll anhand eines typischen CRM-Statements veranschaulicht werden:

Domain	Property	Range
E22 Man Made Object	O19i was object found by	S19 Encounter Event
„Artefakt“	„wurde gefunden von“	„Begegnungsereignis“

Abbildung 6 zeigt einen Ausschnitt der Klasse „E22 Man Made Object“ und eine Auswahl der Eigenschaften, die diese Klasse haben kann, darunter auch jene der bereits erwähnten Mutterklasse „E1 CRM Entity“.

Domain	<input checked="" type="checkbox"/> E22_Man-Made_Object
Property	Target Relation ↓ Select a value <input type="text"/>
Range	<input type="checkbox"/> CIDOC-CRM P1_is_identified_by P2_has_type P3_has_note P8i_witnessed P12i_was_present_at P13i_was_destroyed_by P15i_influenced

Abbildung 6: Auswahl an Eigenschaften der Entität E22 Man Made Object (Printscreen)

Das CRM definiert bestimmte Eigenschaften (*property*) als notwendig für bestimmte Ausgangsklassen (*domain*) oder als abhängig von deren Zielklassen (*range*). Zur Veranschaulichung zeigt Abbildung 7, dass die Wahl der *range* direkt von der gewählten Eigenschaft abhängt. In diesem Fall lässt die Eigenschaft „O19i was object found by“ keine andere Beziehung von „E22 Man Made Object“ zu als jene zu der Klasse „S19 Encounter Event“.

Domain	<input checked="" type="checkbox"/> E22_Man-Made_Object
Property	Target Relation ↓ O19i_was_object_found_by Add intermediate
Range	<input type="checkbox"/> Target Entity Select a value <input type="text"/>
	CRMsci S19_Encounter_Event

Abbildung 7: Auswahl der range in Abhängigkeit der Auswahl der Eigenschaft (Printscreen)

Darüber hinaus werden durch das vorliegende Beispiel ein weiteres wichtiges strukturelles und ein syntaktisches Paradigma des CIDOC CRM veranschaulicht. Das CRM unterstützt auch inverse Eigenschaften zwischen Klassen. Diese werden wie im vorliegenden Fall, durch die passivierte Phrase der ursprünglichen Eigenschaft

ausgedrückt und zusätzlich mit dem Suffix „i“ gekennzeichnet. So würde das ursprünglich gerichtete CRM-Statement lauten:

S19 Encounter Event **O19 has found object** **E22 Man Made Object**
„Begegnungsereignis“ **„hat gefunden“** **„Artefakt“**

Zusätzlich zum Konzept der Vererbung und der syntaktischen Definition sind weitere Kernkonzepte für das Verständnis zur Modellierung von Daten nötig, die Oldman (2014) als „*Hierarchy of Entity*“ umschreibt in bestimmte Kategorien unterteilt werden und anhand des Beispiels erläutert werden können.

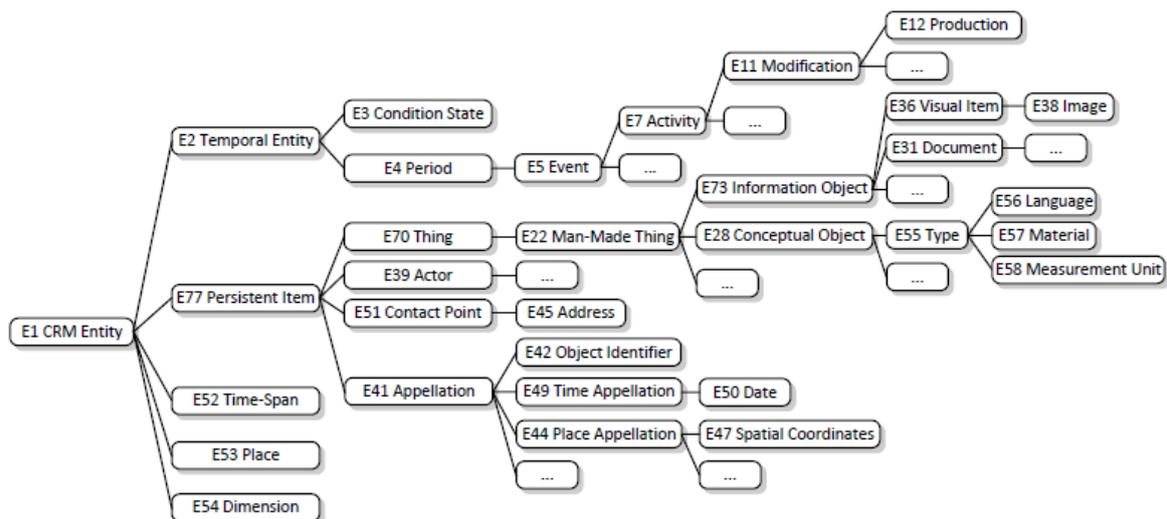


Abbildung 8: Auswahl der Klassenhierarchie des CIDOC CRM. (Nussbaumer & Haslhofer, 2007, S. 7, Abb. 2)

Concept - Persistent Things and Temporary Things – im CRM existieren zwei grundlegende Prinzipien, die sich konzeptionell nicht überschneiden können, aber kohärent sind. Entscheidend ist das im CIDOC CRM die Existenz von Ereignissen als ontologische Kategorie konzipiert ist.

Demnach gibt es zeitliche Phänomene „E2 Temporal Entity“, die nicht andauern, sondern nur über einen bestimmten Zeitraum existieren, wie das Ereignis der Geburt eines Menschen oder historische Perioden. Sie sind aus menschlicher Sicht

immateriell und flüchtig und müssen deshalb als zeitabhängiges oder relatives Ereignis definiert werden.

Zum anderen gibt es persistente Dinge „E77 Persistent Items“, d.h. materielle Objekte wie beispielsweise archäologische Fundobjekte, aber auch Personen, die aufgrund ihrer irdischen Beschaffenheit dazu in der Lage sind Ereignisse auszulösen, zu beeinflussen oder zu bezeugen. Deshalb ist das Finden des materiellen Artefaktes im Beispiel auch abhängig von einem Auslöser, z.B. dem Encounter Event.

„[...] [T]he essence of the CIDOC CRM's event based model [...]“ (Oldman & CRM Labs, 2014, S. 5) ist, dass persistente Dinge in Ereignissen entstehen, überleben, gefunden werden oder enden, die nur durch zeitliche Phänomene beschrieben werden können.

Concept - Use as little as you want – extend if you need to - man muss niemals das gesamte CRM implementieren, sondern nur die Klassen und Eigenschaften die man zur Modellierung benötigt.

2.3. Auswahl der geeigneten Ontologie

Im folgenden Kapitel werden die domänenspezifischen Ontologien, die im weitesten Sinn zur Modellierung des kulturellen Erbes herangezogen werden und auf ihre Tauglichkeit zur Modellierung der TerraPreta Daten, geprüft.

Die Dublin Core Elemente sind fakultativ, beliebig oft wiederholbar und können mit Unterbegriffen genauer beschrieben werden. Abbildung 9 zeigt das sogenannte *hedghog model*. Das zentrale Konzept wird hier nicht näher bestimmt und mittels Relationen direkt mit Datenwerten versehen. Eine Klassenhierarchie, ähnlich zur EDM ist durch beliebige Wiederholungsmöglichkeiten des zentralen Konzepts nicht notwendig.

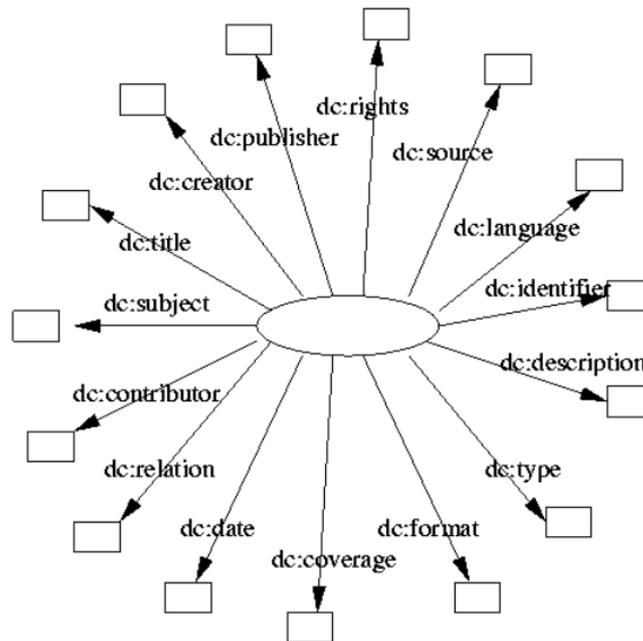


Abbildung 9: Dublin Core Elemente (Kokkelink & Schwänzl, 2002)

Hohmann und Fichtner (2015) beschreiben in ihrer Ausführung ein prägnantes Beispiel der wesentlichen Unterschiede zwischen dem CIDOC CRM und dem DCES, bei dem auch ein wesentlicher Punkt gegen die Modellierung der TerraPreta Daten mit dem Dublin Core Element Set aufgegriffen wird. Demnach verfügt das DCES zur Auszeichnung des Herstellers und des Datums einer Ressource die Relationen *date* und *creator*. Das nicht näher spezifizierte Konzept erhält über diese Relation direkt den entsprechenden Datenwert. Deshalb ist es nicht möglich eine Verbindung zwischen dem Datum und dem Hersteller zu ziehen, weshalb nicht explizit modelliert werden kann, dass sich die Herstellung eines Gegenstandes durch den Hersteller auf eine bestimmte Zeit bezieht.⁵

⁵ Weitere umfangreiche Ausführungen zur Beschreibung von DCES Metadaten durch das CIDOC CRM finden sich in (Doerr, 2000)

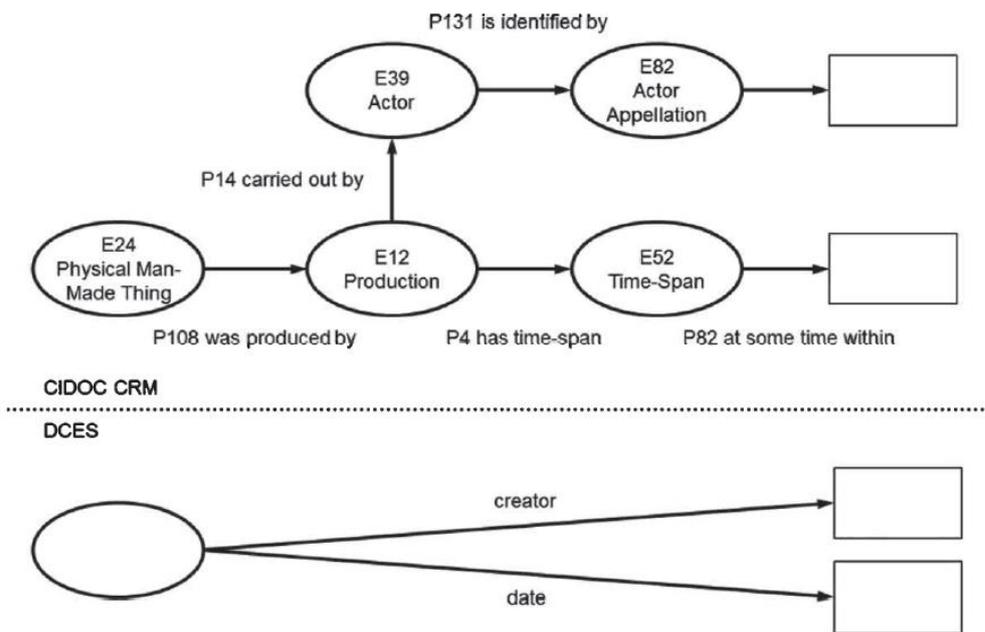


Abbildung 10: Vergleich DCES und CIDOC CRM (Hohmann & Fichtner, 2015, S. 123, Abb. 6)

Jedoch handelt es sich beim zentralen Element der TerraPreta Datenbank eben um solche Herstellungsgegenstände dessen Semantik genau diesen Zusammenhang zwischen Produktionsdatum und deren Produzenten abbilden können muss. Deshalb ist der Einsatz des Dublin Core Element Set nicht für die Modellierung der TerraPreta-Daten geeignet.

Das EDM bietet verschiedene Möglichkeiten zur Integration von Domänenontologien zum Beispiel die Klassenäquivalenzrelation oder die direkte Übernahme von externen Klassen in die EDM-Ontologie (Dröge, Iwanowa, Trkulja, Hennicke, & Gradmann, 2013). Bei erstgenannter wird mit Hilfe einer im EDM bereitgestellten Äquivalenzklasse die Bedeutung von Klassen aus anderen Ontologien gleichgesetzt werden und somit relativ einfach interoperabel. Bei der direkten Übernahme von Klassen kann zum Beispiel das Element der EDM-Ontologie edm:Place mit der Klasse crm:E53 Place der CIDOC CRM verbunden werden. Allerdings ist diese Äquivalenz nachvollziehbar, da für die Konzeption des Europeana Data Model einige Entitäten als Vorlage aus anderen Ontologien, wie dem CIDOC CRM, verwendet werden (Diehr, 2013).

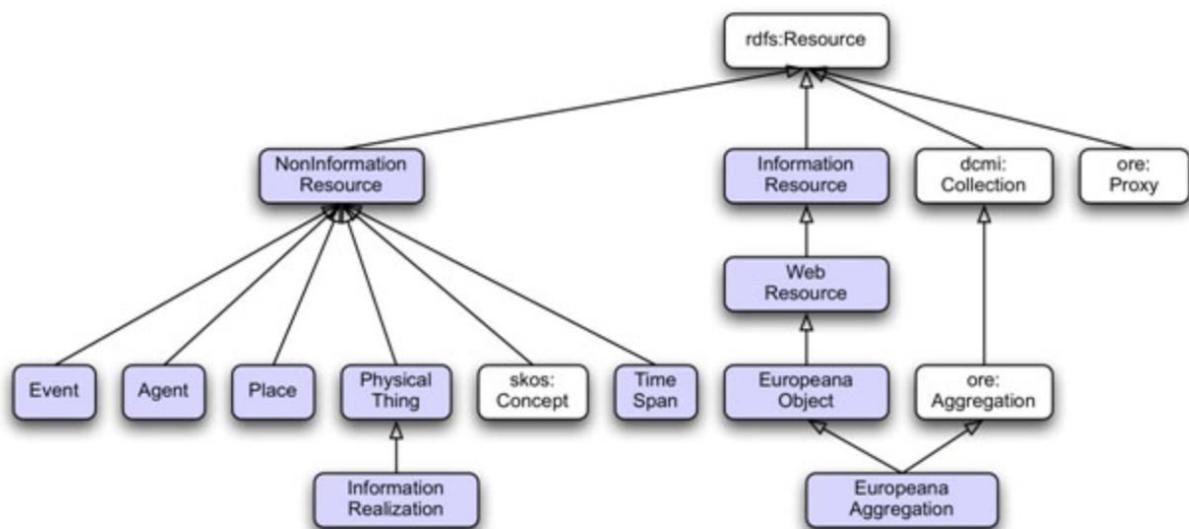


Abbildung 11: EDM Klassenhierarchie (Olensky, 2010)

Abbildung 11 zeigt die gesamte Klassenhierarchie des Europeana Data Models. Jedoch erweist sich das EDM als Top-Level Ontologie für die Modellierung der relationalen Daten der TerraPreta Datenbank zu unspezifisch. Als vereinfachtes Beispiel kennt das EDM zur Auszeichnung eines Platzes und seines Entdeckers die EDM Ressource *place* und *agent*. Jedoch ist die Verbindung zwischen einer Person und dem Platz im EDM durch Relationen nicht vorgesehen und müsste durch die Übernahme externer Klassen und Relationen erst definiert werden.

Grundsätzlich ist die Modellierung durch die Integration externer Klassen zwar möglich, bedeutet aber, dass für die TerraPreta Datenbank zunächst eine Applikationsontologie auf Basis einer geeigneteren Domänenontologie geschaffen werden müsste, welche erst anschließend in die EDM-Ontologie integriert werden kann. Der Einsatz des Europeana Data Models ist deshalb zwar nicht grundsätzlich auszuschließen, zunächst aber für die spezielle Modellierung der TerraPreta-Daten nicht geeignet.

Aus den verschiedenen Bereichen der Informationswissenschaften, vornehmlich gefördert durch Gedächtnisinstitutionen wie Bibliotheken, Museen und Archive, gibt es weitere Kernontologien, die sich der digitalen Dokumentation des kulturellen Erbes angenommen haben und daher aufgrund ihres thematischen Fokus für eine Modellierung in Frage kommen (Doerr, 2009).

Das FRBRoo Model ist eine formale Ontologie zur Erfassung und Abbildung der Semantik bibliografischer Information und soll die Integration, Vermittlung und den Austausch von bibliografischer und musealer Information erleichtern (Le Boeuf, 2012). Es entstand 2008 als objektorientierte Version des *Functional Requirements for Bibliographic Records models* (FRBR) als Versuch, das eventbasierte Paradigma des CIDOC CRM in das bestehende Modell zu integrieren. Allerdings liegt der Fokus neben der Erfassung von bibliographischen Zusammenhängen speziell im Dokumentieren und Bewahren der nicht materiellen Kultur (Doerr, 2009), die dem eigentlichen Zweck der TerraPreta Daten bereits mit der Dokumentation von Artefakten in der Hauptrelation konterkariert.

Das ABC Model beschreibt den Zusammenhang zeitlicher und materieller Entitäten ebenfalls eventbasiert, ähnlich wie das CIDOC CRM. Allerdings besitzt dieses Modell zusätzlich zum Event das Konzept der *situation*. Lagoze und Hunter (2006) zeigen anhand der angeführten Beispiele, dass sich die Ontologie mit ihren 14 Klassen und 19 Eigenschaften zur Abbildung der Semantik für einen Großteil der TerraPreta Daten eignet. Allerdings verursacht die Darstellung der Objekthistorie durch die Einführung der zusätzlichen Entität bei der Modellierung häufig Informationsredundanzen und Probleme bei der Verbindung zwischen den Historien mehrerer Objekte und Akteure (Doerr, 2009). Letztlich können die wenigen Eigenschaften und Klassen auch hier zur semantischen Überladung führen. Die Konzepte des CIDOC CRM mit seinen zugehörigen Erweiterungen können die benötigten Entitäten der TerraPreta Daten zur semantischen Modellierung besser erfassen.

Als letztes Beispiel wird die *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) vorgestellt (Borgo, Massimiliano, Garbacz, & Vermaas, 2010). Das CRM und DOLCE gleichen sich in vielen Konzepten mit einem entscheidenden konzeptionellen Unterschied. Dort wird im Gegensatz zum CIDOC CRM, Raum und Zeit als abhängige Eigenschaften der Dinge betrachtet, deshalb können sie nicht in einer potenziell leeren Raum-Zeit existieren (Doerr, 2009). Eine raumzeitliche

Zuordnung zu Artefakten ist allerdings aus archäologischer Sicht nicht immer gewährleistet. Deshalb kann auch diese Ontologie für die Modellierung der vorliegenden Daten nicht als Referenzmodell dienen.

3. Methoden und Tools

3.1. Schema Ontologie Mapping

Insbesondere für die Interoperabilität im Semantic Web spielen Ontologien eine wichtige Rolle, da sie eine Lösung der semantischen Heterogenität von Daten durch die Bereitstellung einer gemeinsamen Verständnisbasis für ein bestimmtes Interessensgebiet in Aussicht stellen können. Als Folge der weltweiten Vernetzung von Informationen, wie es die Vision von Tim Berners Lee ist kann die Kommunikation zwischen Informationssystemen nur durch interoperable Daten gewährleistet werden. Der Hauptgrund für die semantische Heterogenität liegt in der Bereitstellung von Wissensbeständen durch verschiedenste Provider (Hanke, Mandl, & Womser-Hacker, 2002). Als Folge erschwert sich die Kommunikation zwischen unterschiedlichen Informationssystemen aufgrund der systeminhärenten semantischen Heterogenität schon bereits innerhalb nur einer bestimmten Domäne. Ontologien werden dazu genutzt, die Interoperabilität zwischen verschiedenen Anwendungen zu unterstützen. Relationale Datenbanken gelten als eine der beliebtesten Speicherlösungen für verschiedene Arten von Daten und sind deshalb ein Schlüsselfaktor für die Vernetzung von großer Datenmengen (Spanos, Periklis, & Mitrou, 2012). Das Semantic Web soll diese Masse an Daten über ihre Beschreibung durch Metadaten, besser strukturieren und maschinenlesbar machen (Berners-Lee, Hendler, & Lassila, 2001). Um mit semantischen Inhalten interpretiert werden zu können, müssen die Daten neben der einheitlichen Strukturierung mit anderen Daten in Beziehung gesetzt werden.

Ein Ansatz zur Überbrückung dieses Kommunikationshindernisses und darüber hinaus zur zusätzlichen Informationsgewinnung kann demnach im Erstellen eines Mappings zwischen relationalen Datenbanken und Ontologien gesehen werden. Sahoo et al. (2009) sprechen diesem Zusammenhang von *domain ontology mapping*, welches angetrieben von realen Anwendungsanforderungen mit Hilfe von Domänenontologien sowohl implizite oder gar nicht enthaltene Bedeutungszusammenhänge in Datenbanken zum Informationsgewinn für das Semantic Web erfassen kann. Ghawi

und Cullot (2007) beschreiben den Prozess, bei dem zwischen einer Datenbank und einer Ontologie semantische Korrespondenzen auf konzeptioneller Ebene festgelegt werden, als *database to ontology mapping*.

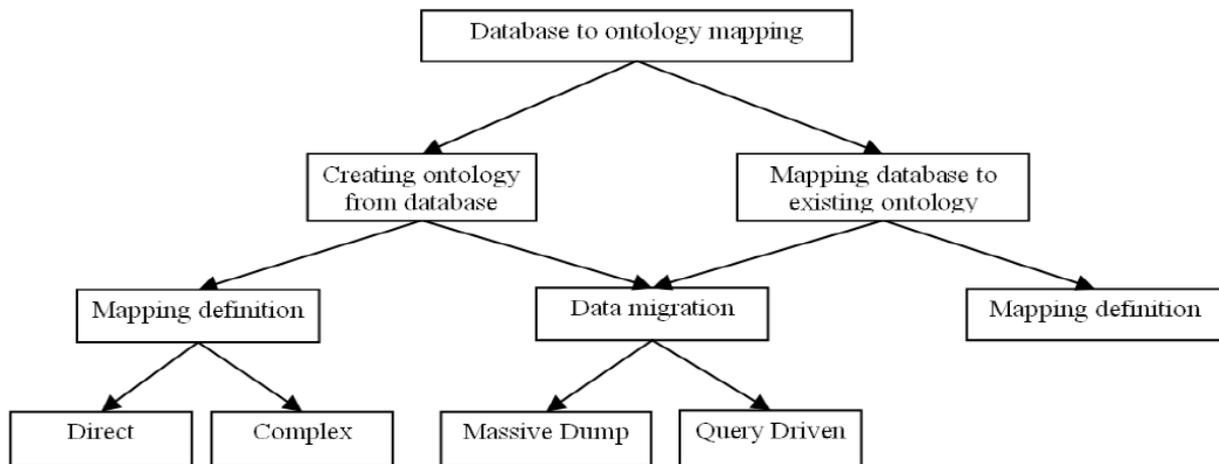


Abbildung 12: Klassifikationsschema zu database-to-ontology mapping Ansätzen (Ghawi & Cullot, 2007)

Demzufolge unterliegt die dieser Arbeit vorangestellten Fragestellung nach der Umsetzung der Modellierung zwei grundlegenden Prozessen, die bei Zuordnung einer Datenbank zu einer vorhandenen Ontologie umgesetzt werden müssen.

Zunächst muss für das Datenbankschema eine Mapping-Definition erstellt werden. Das bedeutet die Transformation des Datenbankschemas in die Ontologiestruktur. Anschließend kann mittels Datenmigration die Überführung der Datenbankinhalte in Ontologie-Instanzen erfolgen. Dieser Schritt wird häufig auch als Ontologiepopulation bezeichnet (Barrasa, Corcho, & Gómez-Pérez, 2003; Ghawi & Cullot, 2007; Spanos, Periklis, & Mitrou, 2012; Sahoo, et al., 2009). Ferner wird als Bedingung für eine diese Art der Ontologiepopulation die Existenz einer Datenbank und einer themenspezifischen Ontologie vorausgesetzt. Die Herausforderung liegt deshalb darin, mit der Mapping Definition semantisch korrekte Bedeutungsäquivalenzen für die unterschiedlichen Konzepte und Relationen zwischen beiden Modellen zu finden. Weil sich die Entwurfskriterien einer Datenbank von denen einer Ontologie unterscheiden, gibt es häufig keine exakte Deckungsgleichheit oder unterschiedliche Überschneidungen zwischen ihren Konzepten (Ghawi & Cullot, 2007), die komplexer umschrieben werden müssen. Deshalb ist für die Mapping Definition und

Datenmigration eine themenspezifische Ontologie unerlässlich, deren Konzepte so nah wie möglich an denen der Datenbank liegen.

Nach Doerr (2009) kann kulturelles Erbe als materieller Beweis für menschliche Tätigkeiten von sozialer Relevanz in der Vergangenheit gesehen werden. Dies lässt sich im weiteren Sinn auch auf die in der TerraPreta-Datenbank gespeicherten Information übertragen. Deshalb kommen für die Modellierung der Daten ausschließlich Ontologien in Frage, die sich mit der Semantik von kulturellem Erbe befassen, welche der vorliegenden Datenstruktur der Datenbank entspricht. Das CIDOC CRM bietet diese elementare Eigenschaft in seiner Bereitstellung in Form des Resource Description Framework Schema (RDFS) an.

3.2. RDF - Resource Description Framework

Nach Hohmann (2011) wird *„[i]n der Vision des Semantic Web [...] die ursprüngliche Datenquelle selbst zum Bestandteil des Netzes und muss nicht durch eine andere Form repräsentiert werden.“* Als Grundlage für diese Art von Datenrepräsentation dient das Resource Description Framework (RDF), welches bereits seit 1999 vom World Wide Web Consortium (W3C) entwickelt und standardisiert wird (Allemang & Hendler, 2011). Bei RDF handelt es sich um ein graphenbasiertes Modell, mit dem man Aussagen in Form von Tripeln, als Informationen über Ressourcen im World Wide Web, darstellen kann. Diese Triple bestehen immer aus einer einfachen Grammatik in Form eines Subjekts (S), eines Prädikats (P) und einem Objekt (O), das wiederum als Subjekt weiterer Aussagen dienen kann (Decker, Prasenjit, & Melnik, 2000). Als Analogie zur Subjekt-Prädikat-Objekt-Syntax des RDF ist die entity-property-entity bzw. domain-property-range-Sequenz des CIDOC CRM zu sehen.

Ein Tripel hat die Gestalt (Severiens, et al., 2006):

`((Subject) URI/UUID), [(Predicate) URI(optional)], [(Object) URI/UUID])`

oder

`((Subject) URI/UUID), [(Predicate) URI(optional)], [Literal])`

Das Subjekt ist immer repräsentiert durch eine Ressource über das eine Aussage getroffen werden kann und das durch einen Uniform Resource Identifier (URI) oder UUID mit dem Attribut `rdf:about` referenziert wird.

```
<crm:E54_Dimension rdf:about="uuid:AABE">
```

Eine Ressource ist eine Entität, auf die eine URI im Web zugreifen kann. Analog zur XML-Syntax werden sie dazu verwendet, um Tagnamen zu unterscheiden.

Severiens et al. (2006) veranschaulichen zur RDF/XML-Syntax einige wesentliche grundlegende Regeln. Demnach kann man die benannten Ressourcen eines RDF-Triples aufteilen in die referenzierten und die im Objekt neu benannten Ressourcen. Es ist nicht möglich, in einem RDF Triple einen identischen Bezeichner für verschiedene Dinge zu verwenden.

Das Prädikat ermöglicht uns, Informationen an Ressourcen anzubinden, um Beschreibungen für sie anzubieten. Auf Objekte, die Literale sind, wird der Wert innerhalb des Tags, für das das Prädikat steht instanziiert.

```
<crm:E54_Dimension rdf:about="uuid:AAAP">
  <crm:P90_has_value>90</crm:P90_has_value>
</crm:E54_Dimension>
```

Alternativ können Werte auch durch das `rdfs:label` repräsentiert sein. Diese stehen für einen Namen bzw. eine Bezeichnung einer Resource in menschenlesbarer Version.

```
<crm:E41_Appellation rdf:about="uuid:AAAT">
  <rdfs:label>Steere</rdfs:label>
</crm:E41_Appellation>
```

Für alle Objekte, die einmal als Ressource mit einer URI bzw. UUID ausgezeichnet worden sind, wird mit dem Attribut `rdf:resource` auf diese verwiesen.

```
<crm:E54_Dimension rdf:about="uuid:AABE">
  <crm:P2_has_type>
    <crm:E55_Type rdf:about="http://ConceptURI/depth"/>
  </crm:P2_has_type>
</crm:E54_Dimension
```

Das gilt insbesondere bei der Wiederholung einer bereits benannten Ressource, die als Konstante fungiert. In diesem Fall trägt das Prädikat stets das Attribut samt URI seiner Ressource.

```
<crm:E54_Dimension rdf:about="uuid:AAAP">  
  <crm:P90_has_value>90</crm:P90_has_value>  
  <crm:P2_has_type rdf:resource=ConceptURI/depth"/>  
</crm:E54_Dimension>
```

Die URI kann in verschiedenen Formen vorliegen (Pohl & Danowski, 2013). Im Hinblick auf die Nutzung als Linked-Open-Data müssen die einzelnen Ressourcen im Internet global eindeutig adressierbar sein. Deshalb muss für jede einzelne Ressource ein eindeutiger Uniform Resource Locator (URL) vergeben werden (Bizer & Cyganiak, 2006). Wenn also an zwei verschiedenen Stellen die gleiche URL verwendet wird, ist eine identische Ressource adressiert. Die Objektstelle eines Tripels kann aber auch durch ein Literal repräsentiert sein, das zudem um die Angabe der Sprache erweitert sein kann, in der das Literal verfasst ist. Darüber hinaus kann das Literal auch durch die Angabe eines Datentyps, wie etwa eine Ganzzahl oder das Datum, charakterisiert werden (Pohl & Danowski, 2013).

Mit Hilfe dieser Grammatik lassen sich Aussagen zu beliebig komplexen Sachverhalten kombinieren. Allerdings werden durch das RDF ausschließlich die grammatikalischen Grundlagen vorgegeben. Deshalb benötigt die Grammatik ein allgemeines oder domänenspezifisches Vokabular oder Schema in Form einer Ontologie, welche die Klassen und deren untereinander definierten und geregelten Eigenschaften vorgibt (Hohmann & Fichtner, 2015).

In RDF formulierte Aussagen lassen sich in verschiedenerlei Syntax darstellen und zwischen Maschinen austauschen. Dazu gehören RDF-Turtle, N-Triples sowie XML/RDF. Für die Serialisierung und maschinelle Verarbeitung sind diese Repräsentationen des RDF-Graphen sehr weit verbreitet (Spanos, Periklis, & Mitrou, 2012).

Das RDF ist also ein Gerüst, mit dem sich Aussagen über Ressourcen in standardisierter Weise treffen lassen. Um Daten aus unterschiedlichen Quellen zu vernetzen, ist es möglich auf ein gemeinsames Vokabular, wie Ontologien, zurückzugreifen, das als Regelwerk dafür zu sehen ist, wie diese Aussagen zu treffen.

Als eine solche Quelle ist auch die objektrelationale TerraPreta Datenbank zu betrachten. Im Gegensatz zum graphenbasierten Modell des RDF sind dort alle Daten in unterschiedlichen Tabellen gespeichert, zwischen denen keine direkte Verbindung besteht (Bruschke & Wacker, 2014).

3.3. 3M – Memory Mapping Manager

Der 3M ist ein Web-Tool für das Erstellen von Mapping Definitionen (WEB6). Es bietet eine Kombination aus Softwarekomponenten, die zusammen einen Rahmen für eine effektive und effiziente Datenintegration mit dem CIDOC CRM bilden (Marketakis, et al., 2017).

Der 3M unterstützt diesen Vorgang mit der X3ML Engine, die während des gesamten Mappingprozesses auf dem User Interface des 3M-Web-Tools komplett automatisiert eine X3ML-Mapping Definitions Datei in Form der X3ML Mapping Definition Language erstellt (Minadakis, et al., 2015). Deshalb werden im Folgenden nur die Aspekte beleuchtet, die für das Verständnis des Mappingprozess besonders relevant sind. Die Definitions-Datei ist ebenfalls in XML repräsentiert. Abbildung 13 zeigt das Basisschema der X3ML-Mapping Definitions Datei.

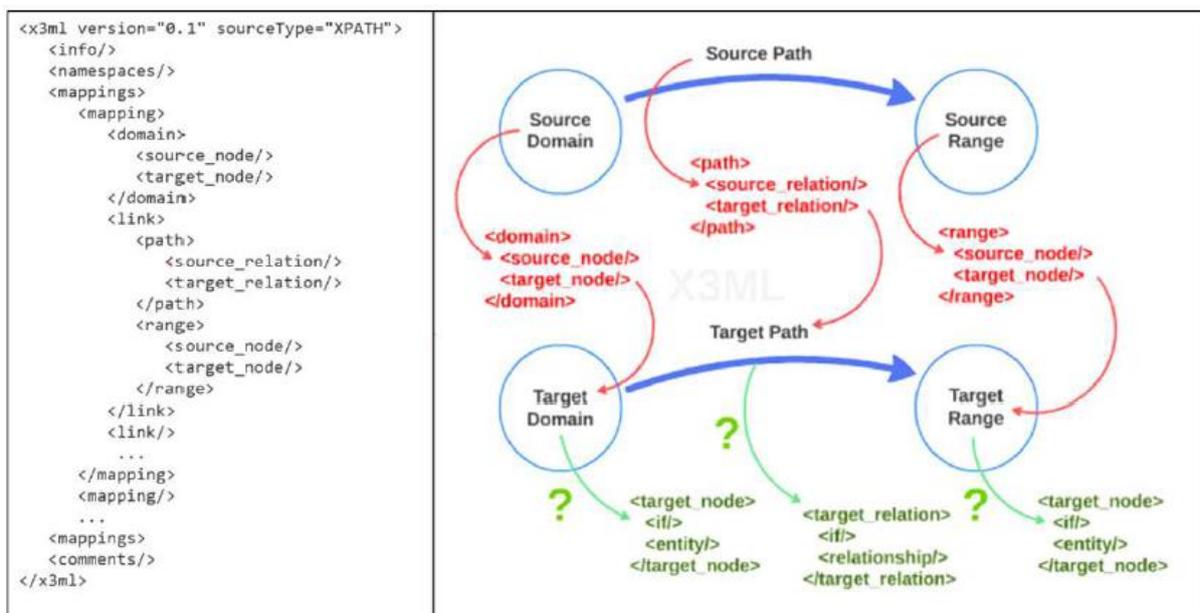


Abbildung 13: Struktur der X3ML Definitions Datei (Minadakis, et al., 2015, S. 6, Abb. 2)

Bedingt durch die Struktur der XML-Quelldatei benötigt die X3ML-Engine für die Pfadnavigation, z.B. zu Quellentitäten einen XPath-Ausdruck (Marketakis, et al., 2017).

Innerhalb des *domain*-Elements erfolgt die Zuordnung zwischen der *source node* oder *domain*-Quellentität `//root/sites` und der ihr semantisch entsprechenden *target node* oder *domain*-Zielentität `crm:E27 Sites`. Innerhalb des Link-Elements kann es ein weiteres Element geben, das *path*-Element. Es ermöglicht die Zuordnung einer CRM-Eigenschaft zwischen der gewählten *domain*-Quellentität und der *domain*-Zielentität. Die Spalte `site_code` repräsentiert in vorliegendem Fall die *source relation* und muss durch die CRM-Eigenschaft `crm:P1 is identified by` als *target relation* mit der *domain*-Zielentität in Beziehung gesetzt werden. Häufig liegt ein wesentlicher Teil der Semantik neben der Klasse, die von den Tabellen/Relationen einer Datenbank auf das semantische Modell mit abgebildet werden soll, in diesen formulierten Beziehungen. Auf das *path*-Element muss in der X3ML-Definition ein *range*-Element folgen. Im folgenden Beispiel wird es verwendet, um die Quellbeziehung `site_code` der *range*-Zielentität `crm:E42 Identifier` zuzuordnen, die dem gültigen Zielbereich der *domain*-Zielentität im *domain*-Knoten entsprechen muss.

Um im späteren Verlauf eine korrekte RDF/XML-Syntax zu erhalten müssen für die CRM-Zielentitäten UUIDs oder URIs in der X3ML-Mapping Definitions Datei definiert werden.

```

...
<mappings>
  <mapping>
    <domain>
      <source_node>//root/sites</source_node>
      <target_node>
        <entity>
          <type>crm:E27_Site</type>
          <instance_generator name="SiteURI">
            <arg name="term"
              type="xpath">site_id/text()</arg>
          </entity>
        </target_node>
      </domain>
    <link>
      <path>
        <source_relation>site_code</source_relation>
        <target_relation>crm:P1_is_identified_by</target_relation>
      </path>
    </link>
    <range>
      <source_node>site_code</source_node>
      <target_node>
        <entity>
          <type>crm:E42_Identifier</type>
          <instance_generator name="UUID"/>
        </entity>
      </target_node>
    </range>
    ...
  </mapping>
</mappings>
...

```

Im Idealfall kann eine Quellentität mit oder ohne Quellbeziehungen direkt durch eine entsprechende Zielentität und deren Eigenschaft einem einfachen *one-to-one mapping* ausgedrückt werden (Doerr, Theodorou, Aspöck, & Masur, 2015). Abbildung 14 zeigt schematisch das grundlegende Vorgehen beim Mapping zwischen Quell- und Zielpfad.

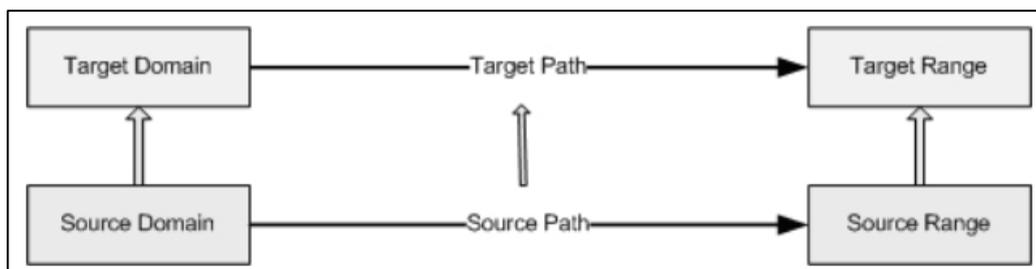


Abbildung 14: Basis Mapping Schema (Kondylakis, Doerr, & Plexousakis, 2006, S. 1, Abb. 1)

In den meisten Fällen besitzt eine Quellentität jedoch gleich mehrere oder sehr spezielle Eigenschaften die durch komplexere Ausdrücke interpretiert werden müssen. Deshalb unterstützt die X3ML Mapping Definition Language auch verschiedene 1:N Operationen mit denen auch komplexe Pfadäquivalenzen ausgedrückt werden können, welche für das Mapping von relationalen Datenbanken, wie die der TerraPreta-Datenbank von Bedeutung sind (Minadakis, et al., 2015).

#	SOURCE	TARGET	CONSTANT EXPRESSION
D	<input type="checkbox"/> //findings/row	<input type="checkbox"/> crm:E22_Man-Made_Object	
P	tradition_id == tradition_id	↓ crm:P108i_was_produced_by crm:E12_Production [p1]	
R	<input type="checkbox"/> //traditions/row	↓ crm:P10_falls_within <input type="checkbox"/> crm:E4_Period	
P	culture_id == culture_id	↓ crm:P108i_was_produced_by crm:E12_Production [p1]	
R	<input checked="" type="checkbox"/> //cultures/row	↓ crm:P10_falls_within <input type="checkbox"/> crm:E4_Period	
P	↓ finding_depth_cm	↓ crm:P53_has_former_or_current_location crm:E53_Place ↓ crm:P39i_was_measured_by crm:E16_Measurement ↓ crmsci:O12_has_dimension crm:E54_Dimension	[crm:P2_has_type] [crm:P91_has_unit]
R	<input type="checkbox"/> finding_depth_cm	↓ crm:P90_has_value <input type="checkbox"/> http://www.w3.org/2000/01/rdf-schema#Literal	[crm:E55_Type = "depth"] [crm:E58_Measurement_Unit = "cm"]

Abbildung 15 zeigt eine Matching Table mit roten Pfeilen, die die Zuordnungen zwischen den Spalten verdeutlichen. Ein vertikaler Pfeil links markiert die 'Quellpfade', ein vertikaler Pfeil rechts die 'Zielpfade'. Ein zentraler Pfeil markiert 'Intermediate Nodes'. Ein Pfeil zeigt auf die 'CONSTANT EXPRESSION' Spalte. Ein Pfeil zeigt auf die '1:N' Notation.

Abbildung 15: Ausschnitt aus der Matching Table des 3M (Printscreen)

Deshalb kann die Entität *crm:E22 Man-Made Object* also nicht nur einer, sondern mehreren Entitäten parallel, zugeordnet werden.

Die sogenannten *Intermediate Nodes* sollen die Zuordnung eines einfachen Quellpfads auf einen komplexen Zielpfad (eine Sequenz von Entity-Property-Entity-Ausdrücken zwischen *domain* und *range*) darstellen. Beispielsweise kann die einfache Beziehung von *finding* zu *finding_depth_cm* im Quellpfad mit dem Ausdruck „finding hat Tiefe in Zentimetern“ paraphrasiert werden. Diese Phrasen können helfen, die Semantik der zu mappenden Elemente besser interpretieren zu können. Dies kann in Form der *source schema interpretation* bzw. Quellschemainterpretation erreicht werden (Doerr, Theodorou, Aspöck, & Masur, 2015). In Zeile 1.3 der Abbildung 15 befindet sich der semantisch äquivalente Zielpfad zu diesem einfachen Quellpfad. Dieser ist durch den eventzentrierten Charakter des CIDOC CRM etwas komplexer und muss daher mit zusätzlichen Klassen umschrieben werden.

Mit *Constant Nodes* können ein oder mehrere konstante Attribute an eine Entität zugewiesen werden. Da die Tiefenangabe in Zentimetern erfolgt und grundsätzlich für alle Entitätsinstanzen gilt, kann die Entität E54 Dimension mit weiteren Konstanten, wie Einheit (*crm: E58 Measurement Unit*) und Tiefe (*crm: E55 Type*), näher umschrieben werden.

Mit *“same as”-Variablen* werden bestimmte Entitäten identifiziert, die im Mappingprozess als identisch betrachtet werden. Beispielsweise sind beide Intermediate Nodes *crm: E12 Production* in Zeile 1.1 und 1.2 durch die Variable „p1“ gekennzeichnet. Damit erhält das Intermediate Node *crm:E12 Production* beim Mapping eine identische UUID bzw. URI. und wird daher nur einmal generiert.

Mit dem *Join Operator* „==“ können Werte aus verschiedenen Tabellen kombiniert werden. Er findet innerhalb des Link-Elements Anwendung. Das bedeutet, dass durch den entsprechenden XPath-Ausdruck gleiche Werte des Quellpfades durch verschiedene Zielentitäten ausgedrückt werden können. Deshalb weist der X-Path-Ausdruck in Zeile 1.1. *//traditions/row* auf den identischen id-Wert in der Tabelle *traditions* hin. Beim Mapping der TerraPreta Daten im 3M erfolgt die Umsetzung eines Joins zweier Tabellen im Quellpfad zunächst mit der Zuordnung der Eigenschaft *tradition_id == tradition_id* zur Quellentität. Mit der Zielentität bzw. *range* erfolgt der Verweis in Form des *xpath*-Ausdrucks auf die entsprechende Tabelle *//traditions/row*. Analog zur Abbildung 15 muss der Zielpfad entsprechend der Semantik zwischen seiner CRM-Ausgangsklasse und seiner CRM-Zielklasse im Mapping der Hauptrelation ausformuliert werden. Darüber hinaus muss bei einem Join die *range* der CRM-Zielklasse im Zielpfad des Mappings der Hauptrelation mit der *domain* der CRM-Ausgangsklasse beim Mapping der Hilfsrelation identisch sein. Wenn Informationen durch einen Join auf die Hauptrelation referenziert sind, dann werden diese unter Verwendung der RDF/XML Syntax in die S-P-O Sequenz der Quellentität der Hauptrelation integriert. Erfolgt bei der Datenmigration keine Fremdschlüsselreferenzierung, dann werden die Daten in der RDF/XML als eigene CRM-Entität in der CRM-Ausgangsklasse ausformuliert.

Nach Doerr et al. (2015) muss bei den Informationen grundsätzlich zwischen kategorischen und faktischen Informationen unterschieden werden. Demnach versteht man unter kategorischen Informationen alle historischen Daten, die für die gesamte Fundgattung relevant sind, wie zum Beispiel die relativchronologische Einordnung durch historische Perioden. Üblicherweise werden solche Informationen in normalisierten

relationalen Datenbanken aufgrund ihrer N:1 Beziehung in Hilfsrelationen vorgehalten. Faktische Informationen zeichnen sich durch Informationen aus, die nur zu einem einzigen Fund gehören, wie Fundposition oder Gewicht. In der TerraPreta Datenbank gehören diese Informationen zur Hauptrelation.

Ähnlich zum *Jena and R2RML*-Konzept (Cioara, et al., 2015) erstellt der 3M für die Transformation der XML-Quelldatei in ein gültiges RDF-Dokument, die Beschreibung der Zuordnungen zwischen Quell- und Zielschema in drei Schritten.



Abbildung 16: 3M Mapping Workflow verändert nach (Oldman, Theodoridou, & Samaritakis, 2016)

1. Auswahl der Quell- und Zielschemata
2. Mapping Definition
3. Erstellen eines Regelwerks für die Uniform Resource Identifier (*URI Generation Policy*)

Zur Auswahl des Quellschemas bietet der 3M zwei verschiedene Importvarianten an. Zum einen kann eine XML Schema Definition oder direkt die XML-Quelldatei nach der Struktur des effective provider schemas importiert werden. Darüber hinaus unterstützt der 3M den Import der Zielschemata in Form von rdf-Dateien. Die Dateien wurden für das Mapping entsprechend der Schemata der TerraPreta-Applikationsontologie importiert.

Durch die als Source Schema Visualizer implementierte Softwarekomponente unterstützt der 3M den Mappingprozess durch das Anzeigen aller im Quelldokument vorhandenen XML-Knoten als XPath-Quellpfade die als Auswahl für die Zuordnung zu Domain, Property oder Range zu Verfügung stehen.

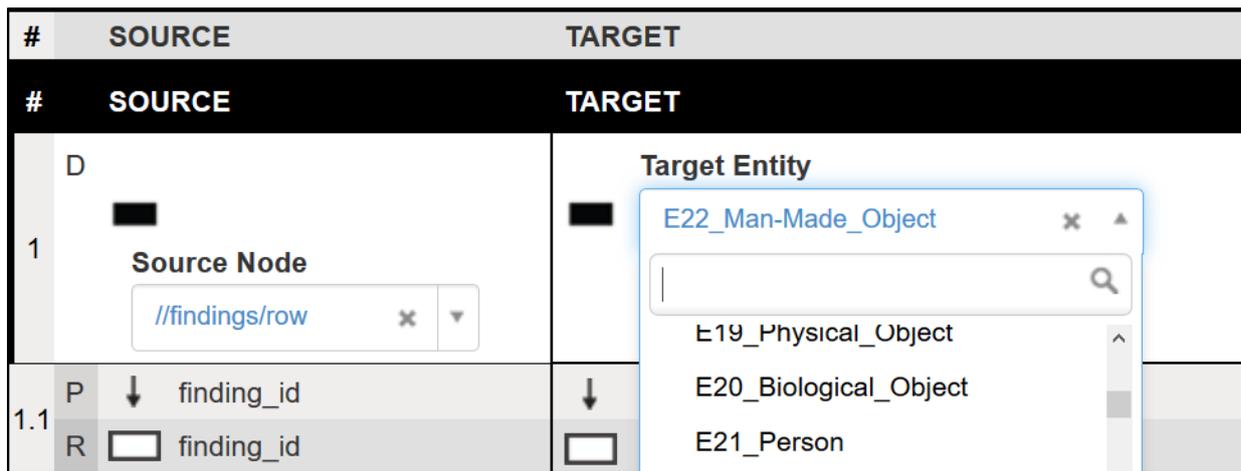


Abbildung 17 Source und Target Schema Visualizer des 3M. (Printscreen)

Analog zu den vorgeschlagenen Quellpfaden kann der Benutzer auch gültige Optionen der Zielschemata aus einem Auswahlfeld auswählen (Marketakis, et al., 2017). Auf diese Weise wird das eigentliche Mapping entscheidend gefördert, da die Komponenten nach Vorgabe der importierten Schemata ausschließlich valide Auswahlmöglichkeiten zulassen und somit eine maximale Fehlervermeidung garantiert. Mit Hilfe dieses einfachen Prinzips erfolgt das gesamte Matching zwischen Quell- und Zielschema im 3M. Wenn keine Ziel-Schemata importiert wurden, ist eine manuelle Eingabe mit freiem Text möglich.

Im letzten Schritt muss für jede Entität des gesamten Matchings ein eindeutiger Identifikationswert generiert werden. Grundsätzlich können diese in Form von URIs oder Literalen sowohl Subjekten als auch Objekten zugewiesen werden (Minadakis, et al., 2015). Darüber hinaus können Objekten zusätzlich Werte in Form von Labels oder Literalen zugewiesen werden.

Durch den Instance Generator, ebenfalls eine Subkomponente der X3ML Engine, können diese Werte erzeugt werden. Dabei muss jede Entität einen Instanz-Generator besitzen und kann optional eine beliebige Zahl an Label-Generatoren besitzen (Minadakis, et al., 2015).

Jede Mapping Definition verfügt über ein vordefiniertes Set an Instanz- und Label Generatoren für UUIDs, URIs, Literale und konstante Werte (Minadakis, et al., 2015). Zudem können weitere URI-Generierungsfunktionen mittels Import einer XML-

basierten *URI Generation Policy File* im 3M implementiert werden und so an die Mapping Definition angepasst werden.

Der folgende Block zeigt das Default-Template zur Definition eines selbst zu definierenden Instanzgenerators (Marketakis, et al., 2017).

```
<instance_generator name="[gen-name]">
  <arg name="[arg-name]" type="[arg-type]">[arg-value]</arg>
  ...
</instance_generator>

<label_generator name="[gen-name]">
  <arg name="[arg-name]" type="[arg-type]">[arg-value]</arg>
  <arg name="language" type="constant">[language-code]</arg>
  ...
</label_generator>
```

Der Argumentname muss entsprechend seiner Bezeichnung immer mit angegeben werden. Der Argumenttyp *type* ermöglicht die Auswahl zwischen den Argumenten *xpath*, *constant* und *position* (Minadakis, et al., 2015). Das Argument *xpath* erlaubt den Zugriff auf Elementinhalte der XML-Quelldatei. Soll also der *arg-value* auf einen Wert aus der XML-Quelldatei zugreifen muss der entsprechende *xpath*-Ausdruck formuliert werden und anschließend die Werteingabe *text()* erfolgen damit der Instanzgenerator auf den entsprechenden Elementinhalt zugreifen kann. Das *position*-Argument ermöglicht dem Generator Zugriff auf die Indexposition des Quellknotens innerhalb seines Kontexts. Auf diese Weise ist es möglich den Stellenwert des migrierten Tuples in die URIs einzubinden. Mit dem Argumentwert *constant* können unveränderbare Ausdrücke in URIs eingebunden werden, um URI Generierungs Schemata umzusetzen. Dazu muss als *value* die entsprechende Werteingabe erfolgen, die in der URI platziert werden soll. Analog zum Vorgehen beim Instanzgenerator kann der Label-Generator verwendet werden.

Spanos et. al (2012) schlagen für den Export von relationalen Datenbanken auf RDFS-Ontologien ein strukturiertes *URI generation scheme* vor, die eine komplexe Linked Data Einbindung noch nicht unterstützt, aber als Basisvorlage für die Entwicklung der URI-Adressierung im TerraPreta-Umfeld geeignet ist.

Database Element	URI Template ^a	Example
Database	$\{base_URI\}/\{db\}/$	http://www.example.org/univ/
Relation	$\{base_URI\}/\{db\}/\{rel\}$	http://www.example.org/univ/staff/
Attribute	$\{base_URI\}/\{db\}/\{rel\}\#\{attr\}$	http://www.example.org/univ/staff#address
Tuple	$\{base_URI\}/\{db\}/\{rel\}/\{pk\} = \{pkval\}$	http://www.example.org/univ/staff/id=278

Abbildung 18: URI Generierungsschema (Spanos, Periklis, & Mitrou, 2012, S. 14, Tab. 3)

Da es sich bei der gesamten Modellierung um ein eher theoretisches Konzept zur semantischen Modellierung von relationalen Daten handelt, ist bei der Erstellung der konzeptionelle Charakter der URIs zu beachten. Die verwendeten URIs und UUIDs zielen daher nur auf grundlegende Funktionalität der RDF-Syntax. Ein erfolgreiches Zusammenspiel mit der zugrundeliegenden Datenbank oder einem existierenden Trip-Store erfolgt nicht im Rahmen dieser Arbeit.

Demnach besitzt das TerraPreta-Datenbankelement folgendes Ausdrucksschema für das URI-Basistemplate:

Datenbank	http://www.tp.de/database/
Relation	http://www.tp.de/database/relation/#pkvalue#
Konzepte	http://www.tp.de/database/Concept/Bezeichner/...

Zur Vereinfachung werden zunächst nur für die relationsäquivalenten Domains URIs generiert. Dabei sollten die lokalen Identifikatoren aus den Relationen nur dann gemappt werden, wenn ihre Bezeichner auch in anderen Dokumenten oder Datenbanken Verwendung finden (Doerr, Theodorou, Aspöck, & Masur, 2015). Alternativ werden diese nur zum Generieren der Uniform Resource Identifier für die *domains* der CRM-Entitäten verwendet.

Die Daten der Relationen werden mit Hilfe von RDF-Labels migriert, wenn die target range keine direkte Instanziierung mittels Literal zulässt.

Unter Konzepten werden im Sinne der vorliegenden Arbeit konstante Einheits- und Formatangaben verstanden, die für alle Werte innerhalb einer Spalte gelten. Nach der Angabe des englischen Begriffs für Konzept wird in der URI zunächst die Art der Bezeichnung definiert auf die wiederum die entsprechende Einheit oder das vorliegende Format bezeichnet wird. Auf diese Weise können mittels URI-Adresse Metainformationen zu den vorliegenden Daten gegeben werden.

3.4. Effective Provider Schema

Im Bereitstellungsprozess der TerraPreta Daten ist die Normalisierung in Form einer korrekten XML-Syntax für die Definition der gleichwertigen Verknüpfungen und demzufolge für eine Verwendung im 3M erforderlich. Marketakis et al. (2017) schlagen als Grundlage das *effective provider schema* als strukturiertes Quellschema für die Zusammenführung mit dem Zielschema in Form der Applikationsontologie vor. Mit der pgadminIII Benutzeroberfläche ist es möglich einen einfachen XML-Export der PostgreSQL/PostGIS-Geodatenbank in mehreren Varianten durchzuführen (PostgreSQL Global Development Group, 2017). Um das *effective provider schema* zu generieren kann die folgende SQL-Funktion Relationen und deren Inhalt einer strukturierten XML-Syntax zuweisen.

```
table_to_xml(tbl regclass, nulls boolean, tableforest boolean,  
            targetns text)
```

Der `regclass`-Typ akzeptiert Zeichenfolgen, die Tabellen anhand ihres Tabellennamens identifiziert. Der zweite Parameter `nulls` bestimmt mittels einer Booleschen Variable, ob Null-Werte zugelassen werden sollen oder nicht. Mit dem vierten Parameter `targetns` kann bei Bedarf ein Zielnamensraum festgelegt werden, der durch das Setzen eines Leerzeichens ausgelassen werden kann.

Mit Hilfe des Parameters `tableforest` wird die Struktur der XML Ausgabe beeinflusst. Ist dieser Parameter auf den Wert „false“ gesetzt, weist die generierte XML die im Anhang präsentierte Struktur auf, die der des *effective provider schema* entspricht. Infolgedessen wird der Tabellename nur einmal als Wurzelement ausgegeben und jeder Datensatz als Kindelement der Tabelle mit der Bezeichnung `<row>`. Jedes Attribut wird als Kindelement der Bezeichnung `<row>` angeschlossen. Jede Tabelle muss einzeln exportiert und anschließend zu einer gesamten Datei zusammengeführt werden. Auf diese Weise erhält man den strukturierten Inhalt einer Sequenz aus Eltern- und Kindelementen wie er sich im Anhang befindet.

4. Modellierung

Das Mapping der TerraPreta Datenbank beginnt in der Zerlegung der Tabellen in Beziehungen zwischen Entitäten. Ziel ist es dabei eine möglichst genaue Abbildung der Daten zu gewährleisten, ohne Verlust von Informationen zu verursachen. Gerade im Kontext archäologischer Projekte enthalten auch Tabellennamen oder Feldnamen wichtige implizite Informationen, die ebenso zur Semantik den gespeicherten Daten gehören und bei der Modellierung berücksichtigt werden müssen (Ermert, et al., 2004).

Das Mapping orientiert sich dabei weitestgehend an Standardmappings, die von den Entwicklern des CRM veröffentlicht wurden. Bereits auf das CIDOC CRM übertragene Modellierungen von archäologischen Daten werden dabei berücksichtigt und deren Ansätze gegebenenfalls auf das in dieser Arbeit vorgeschlagene Modell übertragen.

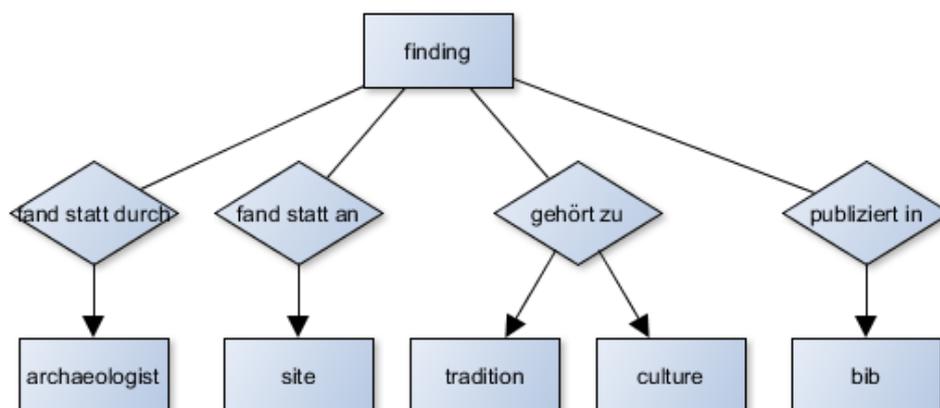


Abbildung 19: Semantisches ER-Modell der TP-Datenbank (Schwer, 2016)

Zur Förderung des Verständnisses, in welchem Verhältnis die Tabellen zu einander intendiert sind, zeigt Abbildung 19 das semantische Entity-Relationship-Modell, welches dem reduzierten Entwurf der TerraPreta Datenbank zugrunde lag.

Zur Umsetzung der Abbildung einer sinnvollen Semantik wurde zunächst eine neue Spalte mit dem Namen `concat_c14_year` und mit dem Datentyp `text` mit folgendem Statement erstellt:

```
ALTER TABLE findings ADD COLUMN concat_c14_year text;
```

Anschließend können mit folgendem Statement alle Spalten miteinander verkettet werden, die im Modell in einen Ausdruck zusammen migriert werden müssen. Dabei ist zu beachten, dass die Zeichenfolge '+/-' noch zusätzlich zwischen dem Datierungsjahr und der Streuungsangabe eingefügt werden muss.

```
UPDATE findings SET concat_c14_year = concat_ws (' ', c14_year, '+/-' , c14_plusminus, c14_reference) WHERE c14_year IS NOT NULL;
```

Der verkettete Ausdruck der drei Spalten kann anschließend korrekt in RDF/XML transformiert werden.

4.1. XML-Syntax - Normalisierung

Das folgende SQL-Statement des pgadminIII-XML-Exports kann eine für den 3M verwendbaren XML-Datei in Form des *effective provider schema* erzeugen.

Hier, exemplarisch für die Relation <findings> angewendet, gilt das Statement auch für alle anderen Tabellen der TerraPreta Datenbank zur Erzeugung einer normalisierten XML-Quelldatei.

```
table_to_xml(`findings`, false, false, ` `)
```

Die Instanz jedes Elements kann nun in der XML-Struktur über einen bestimmten Pfad adressiert werden. Im folgenden Beispiel für das Kind-Element `finding_id`.

```
//root/findings/row/finding_id
```

Anschließend kann die Interpretation des Quellschemas als semantisches Modell oder Graph in Form von Nodes und Links erfolgen. Dabei muss jedes Element einem gleichwertigen CIDOC-CRM-Pfad zugeordnet sein, so dass jede Instanz eines Elements des Quellschemas in eine gültige Instanz des CIDOC-CRM mit derselben Bedeutung umgewandelt werden kann.

Dieser Vorgang wird mit Hilfe des Memory Mapping Manager (3M) unterstützt.

4.2. Aufbau der TerraPreta Applikationsontologie

Zur Abbildung des TerraPreta-Quellschemas ist ein Komplex aus mehreren kompatiblen Domänenontologien nötig, die nach Hohmann (2008) als Grundlage für eine lokale Applikationsontologie dient, um die Modellierung möglichst ohne Datenverluste zu gewährleisten. Die Applikationsontologie enthält im Gegensatz zur Domänenontologie eine minimalistische Struktur, die genau auf die Bedürfnisse der speziellen Anwendung zugeschnitten ist (Menzel, 2003) und so auch dem Minimalitätsprinzip von Oldman (2014) entspricht. So werden ausschließlich die Klassen und Eigenschaften verwendet, die zur Modellierung der Applikationsontologie benötigt werden. Da die Versionen des CIDOC CRM und dessen Erweiterung mittlerweile sehr umfangreich sind, finden sich in Tabelle 2 alle zur semantischen Modellierung benötigten Ontologien des Targetschemas mit deren Versionsstand, Schemarepräsentation und Referenz.

Tabelle 2: Verwendete Ontologien und Versionen

Schema	Repräsentation	Version	Referenz [STAND 29.10.2017]
CIDOC-CRM	RDFS	6.0	http://www.cidoc-crm.org/sites/default/files2/cidoc_crm_v6.0-draft-2015January.rdfs.xml
CRMsci	RDFS	1.2.2	http://www.ics.forth.gr/isl/CRMext/CRMsci_v1.2.2.rdfs
CRMarchaeo	RDFS	1.4	http://www.ics.forth.gr/isl/CRMext/CRMarchaeo_v1.4.rdfs
CRMgeo	RDFS	1.2	http://www.ics.forth.gr/isl/CRMext/CRMgeo_v1.2.rdfs
CRMext4SKOS andLabel	RDFS	1.2	https://www.w3.org/2009/08/skos-reference/skos.rdf

Das CRMsci stellt ein Schema zur Integration von Metadaten über wissenschaftliche Beobachtungen, Messungen und verarbeitete Daten, die in beschreibenden und empirischen Wissenschaften verwendet werden sollen (Doerr, Kritsotaki, Rousakis, Hiebel, & Theodoridou, 2014).

Das CRMgeo stellt ein Schema zur Integration raumzeitlicher Eigenschaften und zeitlicher Entitäten, sowie der Einbindung von Geodaten mithilfe von Konzeptualisierungen, formalen Definitionen, Encoding-Standards und der durch das Open Spatial Consortium (OGC) definierten Beziehungen. Es verbindet das CIDOC CRM mit dem OGC-Standard von GeoSPARQL (Hiebel, Doerr, Eide, & Theodoridou, 2015).

Das CRMarchaeo stellt ein Schema zur Beschreibung von Sachverhalten rund um archäologische Ausgrabungen, sowie allen Einrichtungen und Aktivitäten in diesem Zusammenhang (Doerr, et al., 2016).

Das Simple Knowledge Organisation System (SKOS) ist ein elementarer Standard für die Funktionalität von Linked Open Data. Mit der Nutzung von SKOS können Verknüpfungen zu anderen Wissensorganisationssystemen, wie Thesauri, Taxonomien, Klassifizierungen oder anderen Schemata zum Beispiel über das Web hergestellt werden, da es kompatibel mit XML, RDF und OWL Standards ist (Olensky, 2010).

Alle Entitäten der TerraPreta Datenbank sollen als eine Sequenz in Form von entity-property-entity-Tripeln ausgedrückt werden können, mit denen die Objekte der Datenbank syntaktisch als *domain-property-range*-Abfolge beschrieben werden können.

4.3. Mapping Definition

Zu den wichtigsten Schritten der Bereitstellung von Daten in einem Linked Open Data Ökosystem gehört das Mapping. Beim Mapping wird eine Beziehung zwischen einem Datenbankschema und dem vorgegebenen Vokabular der Ontologie hergestellt. Dabei werden einander semantisch entsprechende Entitäten und Klassen zwischen einer Quelldatei und einem Zielschema zugeordnet (Spanos, Periklis, & Mitrou, 2012). Die Zuordnung der Klassen und Eigenschaften des CIDOC CRM liegen im RDF Format vor und die Elemente der TerraPreta Daten in Form einer XML-Quelldatei. Deshalb ist es erforderlich zu definieren welche Instanz eines Elements des Quellschemas in welche Instanz des CIDOC-CRM mit derselben Bedeutung überführt werden soll, eine Mapping Definition.

Die Interpretation des Quell-Schemas als semantisches Modell erfolgt in Form von Nodes und Links durch die Zuordnung jedes einzelnen Elementes zu einem gleichwertigen CIDOC-CRM-Pfad der der gleichen Bedeutung entsprechen muss. Tabelle 3 zeigt eine vereinfachte Äquivalenztabelle, wie die Bestandteile einer relationalen Datenbank den Komponenten der Ontologie des CIDOC CRM gleichgesetzt werden kann.

Tabelle 3: Äquivalenzen zwischen CIDOC-CRM und relationaler Datenbank (Crofts N. , 1999)

CIDOC-CRM	Relational DB
class	entity, table
instance	row, record, tuple
attribute	Field, column
method	stored procedure or function

Jede Tabelle wird auf diese Weise repräsentiert durch eine CRM-Klasse die gleichzeitig auch der CRM-Domain entspricht.

Die Spalten der Tabellen werden in Sequenzen aus mehreren Entity-Relation-Entity- bzw. Domain-Property-Range Pfaden ausgedrückt und müssen in Beziehung zur CRM-Domain gesetzt werden. Im Mapping Memory Manger wird eine Datenbank-Tabelle in Form von mehreren Domain-Property-Range-Sequenzen ausformuliert.

Setzt man das Schema der TerraPreta-Datenbank mit der RDF-Syntax gleich, kann folgendes vereinfacht formuliert werden. Jedes Tupel einer Tabelle entspricht einer gesamten CRM-Entität (Doerr, Theodorou, Aspöck, & Masur, 2015). Das bedeutet, dass für jedes Tupel in einer Tabelle eine gesamte CRM-Entität oder domain-property-range-Sequenz mit allen dazu benötigten Klassen und Instanzen in RDF/XML formuliert wird. Die *domain-property-range*-Sequenz wird stellvertretend für alle Tupel einer Tabelle nur einmal im 3M ausformuliert. Die Spalten einer Tabelle werden als Eigenschaften in Form der CRM-Property in Beziehung zur CRM-Domain gesetzt. Der Schnittpunkt aus Tupel und Spalte, fortan Spaltenwert, wird als Entitätsinstanz durch eine weitere CRM-Klasse beschrieben, der CRM-Range. Sie können durch ein Literal oder ein Label in RDF/XML instanziiert werden.

4.3.1. Mapping der Hauptrelation <findings>

#	SOURCE	TARGET
2	D <input checked="" type="checkbox"/> //findings/row	<input checked="" type="checkbox"/> crm:E22_Man-Made_Object
2.1	P ↓ finding_id	↓ crm:P1_is_identified_by
	R <input type="checkbox"/> finding_id	<input type="checkbox"/> crm:E42_Identifier
2.2	P ↓ site_id == site_id	↓ crmsci:O19i_was_object_found_by ↓ crmsci:S19_Encounter_Event [enev1]
	R <input type="checkbox"/> //sites/row	↓ crm:P117_occurs_during ↓ :A9_Archaeological_Excavation [ae1] ↓ crm:P8_took_place_on_or_within <input type="checkbox"/> crm:E27_Site
2.3	P ↓ tradition_id == tradition_id	↓ crm:P108i_was_produced_by ↓ crm:E12_Production [p1]
	R <input type="checkbox"/> //traditions/row	↓ crm:P10_falls_within <input type="checkbox"/> crm:E4_Period
2.4	P ↓ culture_id == culture_id	↓ crm:P108i_was_produced_by ↓ crm:E12_Production [p1]
	R <input type="checkbox"/> //cultures/row	↓ crm:P10_falls_within <input type="checkbox"/> crm:E4_Period
2.5	P ↓ archaeologist_id == archaeologist_id	↓ crmsci:O19i_was_object_found_by ↓ crmsci:S19_Encounter_Event [enev1]
	R <input type="checkbox"/> //archaeologists/row	↓ crm:P9i_forms_part_of <input type="checkbox"/> crm:E7_Activity
2.6	P ↓ bib_id == bib_id	↓ crm:P70j_is_documented_in
	R <input type="checkbox"/> //bib/row	<input type="checkbox"/> crm:E31_Document
2.7	P ↓ c14_year	↓ crm:P108i_was_produced_by ↓ crm:E12_Production [p1]
	R <input type="checkbox"/> c14_year	↓ crm:P4_has_time-span ↓ crmgeo:SP10_Declarative_Time-Span ↓ crm:P82_at_some_time_within <input type="checkbox"/> http://www.w3.org/2000/01/rdf-schema#Literal
2.8	P ↓ c14_info	↓ crm:P3_has_note
	R <input type="checkbox"/> c14_info	<input type="checkbox"/> :http://www.w3.org/2000/01/rdf-schema#Literal
2.9	P ↓ finding_depth_cm	↓ crm:P53_has_former_or_current_location ↓ crm:E53_Place ↓ crm:P39i_was_measured_by ↓ crm:E16_Measurement ↓ crmsci:O12_has_dimension ↓ crm:E54_Dimension [crm:P2_has_type] [crm:P91_has_unit] [crm:E55_Type = "depth"] [crm:E58_Measurement_Unit = "cm"]
	R <input type="checkbox"/> finding_depth_cm	↓ crm:P90_has_value <input type="checkbox"/> http://www.w3.org/2000/01/rdf-schema#Literal

Abbildung 20: Mapping der Hauptrelation <findings> (Printscreen)

Abbildung 20 zeigt das gesamte Mapping der Tabelle. Die Hauptrelation <findings> lässt sich in einem einfachen und mehreren komplexen semantischen Pfaden abbilden. Die Relation, also die Quellentität *finding* wird dabei repräsentiert durch die Zielentität *E22 Man Made Object*, da sie konkrete Gegenstände bzw. Artefakte umfasst, die in ihrem Erschaffungsprozess eine bewusste menschliche Aktivität voraussetzen. Somit muss sie als domain-Quellentität im 3M-Mappingprozess deklariert werden. Jedes einzelne in der Datenbank gespeicherte Artefakt wird mit der domain der Klasse E22 gemappt. Insgesamt weist die domain in ihrer Quellschemainterpretation neun

Eigenschaften bzw. Quellbeziehungen auf die auf den Zielpfad übertragen werden müssen.

- Das/ein finding hat eine ID
- Das/ein finding fand statt an *site*
- Das/ein finding gehört zu Chronologie *tradition*
- Das/ein finding gehört zu Chronologie *culture*
- Das/ein finding wurde gefunden durch *archaeologist*
- Das/ein finding ist publiziert in *bib*
- Das/ein finding wurde produziert innerhalb eines Zeitraums der Radiocarbonatierung *c14 year, c14 plusminus, c14 reference*
- Das/ein finding hat eine Radiocarbonatierung mit Information *c14 info*
- Das/ein finding wurde gefunden in Tiefe *finding depth cm*

Übertragen auf das Zielschema ergeben sich die in Abbildung 20 gezeigten Ausdruckssequenzen.

Die ID eines Artefaktes wird hier nur exemplarisch beim Mappingprozess umgesetzt und über die Eigenschaft *P1 is identified by* der range *E42 Identifier* zugeordnet, da die lokalen Bezeichner im Fall der TerraPreta Datenbank nur für die URI-Generierung verwendet werden.

Wie bereits erwähnt, muss bei Joins der Zielpfad entsprechend seiner semantischen Interpretation zwischen seiner CRM-Ausgangsklasse und seiner CRM-Zielklasse ausformuliert werden.⁶

Folglich erfolgt der Join der Hilfstabelle *sites* auf die Haupttabelle *findings* mit dem Mapping in einer komplexen Pfadsequenz, die zunächst durch das „Finden“ selbst in einer Art unsichtbarem Event als *Intermediate Node* umschrieben werden muss. Aus archäologischer Sicht führt das Auffinden des Artefakts zur chronologischen Einordnung des Fundplatzes, wird somit als potentiell relevant für eine Gemeinschaft betrachtet (Doerr, et al., 2016) und kann deshalb im Zielpfad mit Hilfe der Klasse *S19 Encounter Event* und seiner inversen Eigenschaft *O19i was object found by* umschrieben werden. Darüber hinaus enthält die Beziehung zwischen Artefakt und

⁶ In der PostgreSQL-Datenbank entspräche der hier übertragene Join folgendem SQL-statement:
SELECT findings.site_id, sites.site_id FROM public.findings, public.sites WHERE findings.site_id = sites.site_id;

Fundplatz auch implizite Informationen darüber, dass das Auffinden während (*P117 occurs during*) der speziellen Aktivität *A9 Archaeological Excavation* stattgefunden hat, die wiederum am Fundplatz *E27 Site* lokalisiert ist (*P8 took place on or within*).

Das Mapping der beiden relativchronologischen Zuordnungen *traditions* und *cultures* zu den Artefakten erfolgt im Mappingprozess identisch. Ausgehend von der *domain E22 Man Made Object* erfolgt die Einführung des Intermediate Node *E12 Production* aufgrund der konzeptionellen Kohärenz zwischen dem Artefakt als persistentem Ding und seiner Entstehung als zeitlichem Phänomen. Die Produktion eines Artefakts fand statt (*P10 falls within*) innerhalb eines bestimmten relativchronologischen Zeitraums *E4 Period*, repräsentiert durch Tradition und Kultur. Typischerweise wird diese Klasse zur Repräsentation prähistorischer oder historischer Zeiträume verwendet (Ore, Doerr, Le Bœuf, & Stead, 2015). Allerdings sind damit keine festen zeitlichen Grenzen verbunden. Deshalb können sich *culture*-Perioden als chronologische Differenzierung einer *tradition*-Periode in unterschiedlichen Zeiträumen überschneiden und zeitlich koexistieren. Unabhängig von der chronotypologischen Differenzierung ist der Produktionszeitpunkt eines Artefaktes als identischer Event „p1“ zu betrachten.

Durch den Join auf die Hilfsrelation *archaeologists* wird im Quellpfad eigentlich auf den Akteur, der für das Finden verantwortlich ist, verwiesen. Jedoch ist nicht wirklich bekannt, ob der Archäologe an der gesamten Ausgrabung oder nur teilweise partizipiert hat. Deshalb muss die *range* des Zielpfades als *E7 Activity* und nicht als Person interpretiert werden, da diese als absichtliche Handlung von Instanzen der Klasse *E39 Akteur* ausgeführt werden und zu Zustandsänderungen kultureller, sozialer oder physischer Systeme führen (Ore, Doerr, Le Bœuf, & Stead, 2015). Als Teil der Aktivität (*P9i forms part of*) interpretiert werden kann wiederum das Finden selbst als *S19 Encounter Event*. Der Event ist ganz ähnlich zur Produktion des Gegenstandes, aus Sicht des Gegenstandes gleich und muss deshalb als identischer Event „enev1“ deklariert werden.

Durch den Join wird auf die in der Hilfstabelle *bib* gespeicherten Angaben zu Literatur verwiesen. Die *domain* steht deshalb in diesem Zusammenhang in einer Beziehung zu *E31 Document*. Die Klasse umfasst identifizierbare immaterielle Gegenstände, die Vorschläge über die Realität in Form von Texten, Bildern oder Grafiken machen (Ore, Doerr, Le Bœuf, & Stead, 2015). Die gespeicherten Angaben verweisen auf eben solche Gegenstände. Die Artefakte müssen demnach über die inverse Eigenschaft *P70i is documented* in Beziehung zueinander gesetzt werden.

In Kapitel 2.1 wurde bereits auf die Datenhaltung der Radiocarbonatierung in Form von vier verschiedenen Attributen hingewiesen. Diese Art der Speicherung erweist sich im ursprünglichen Fall zur Berechnung von Einzugsgebieten als vorteilhaft (von Groote-Bidlingmaier, Timpf, & Hilbert, 2014), allerdings kann dieser Zusammenhang nicht sinnvoll semantisch umgesetzt werden. Deshalb wird bei der Modellierung ein für C14-Datierungen übliches Format vorausgesetzt (Jahr +/- Jahr und Referenzangabe), das in der Datenbank einer Verkettung der drei Spalten `c14_year`, `c14_plusminus` und `c14_reference` entspricht.

Infolge dessen wird zur Beschreibung der Modellierung weiterhin dieses fiktive Attribut mit den Namen `c14_year` verwendet. Es muss durch die direkte Zuweisung zu *E22 Man Made Objects* als *domain* in diesem Zusammenhang ebenfalls den identischen Produktionszeitpunkt „p1“ besitzen, der innerhalb der hier ermittelten Zeitspanne liegt (*P82 at some time within*). Allerdings muss die Instanz, also das Ergebnis der Radiocarbonatierung, zunächst durch den Intermediate Node der Unterklasse *SP 10 Declarative Time Span* gefasst werden. Die *SP 10 Declarative Time Span* ist definiert durch einen Ausdruck, der einen bestimmten zeitlichen Umfang beschreibt (Hiebel, Doerr, Eide, & Theodoridou, 2015). Sie können zur Beschreibung verwendet werden, um ungefähre tatsächliche Zeitspannen von zeitlichen Einheiten, wie die Produktion von Artefakten zu charakterisieren, die beispielsweise in Form von C14 Radiocarbonatierungen vorliegen (Hiebel, Doerr, Hanke, & Masur, 2014). Deklarative Zeitspannen hängen von der Dokumentation und der Mess-Methode ab. Diese Art von Informationen dienen im Besonderen zur nicht formalen Charakterisierung des Gegenstands und werden deshalb in einer weiteren Beziehung von den Man Made Objects über die Eigenschaft *P3 has note* zu einem RDF-Literal beschrieben, welches die in der Datenbank gespeicherte Information instanziiert.

4.3.2. Mapping der Hilfsrelation <sites>

#	SOURCE	TARGET
6	D  //sites/row	 crm:E27_Site
6.1	P  site_code	 crm:P1_is_identified_by
	R  site_code	 crm:E42_Identifier
6.2	P  site_name	 crm:P1_is_identified_by
	R  site_name	 crm:E41_Appellation
6.3	P  site_length_m	 crm:P43_has_dimension
		 crm:E54_Dimension [crm:P91_has_unit] [crm:E58_Measurement_Unit = "m"] [crm:P2_has_type] [crm:E55_Type = "length"]
	R  site_length_m	 :http://www.w3.org/2000/01/rdf-schema#Literal
6.4	P  site_width_m	 crm:P43_has_dimension
		 crm:E54_Dimension [crm:P91_has_unit] [crm:E58_Measurement_Unit = "m"] [crm:P2_has_type] [crm:E55_Type = "width"]
	R  site_width_m	 :http://www.w3.org/2000/01/rdf-schema#Literal
6.5	P  site_lat	 crm:P1_is_identified_by
	R  site_lat	 crm:E47_Spatial_Coordinates [crm:P2_has_type] [crm:E55_Type = "latitude"]
6.6	P  site_lon	 crm:P1_is_identified_by
	R  site_lon	 crm:E47_Spatial_Coordinates [crm:P2_has_type] [crm:E55_Type = "longitude"]
6.7	P  geom	 crmgeo:Q11i_is_approximated_by
	R  geom	 crmgeo:SP6_Declarative_Place

Abbildung 21: Mapping der Hilfsrelation <sites> (Printscreen)

Die Hilfsrelation sites lässt sich in fünf einfachen und zwei komplexen semantischen Pfaden abbilden. Die Relation, also die *domain*-Quellentität *sites* muss aufgrund der Referenzierung zur Haupttabelle mit der gleichen die *domain*-Zielentität *E27 Site* gemappt werden. Mit dieser Klasse können Landstücke beschrieben werden, die sich aus verhältnismäßig unbeweglichen materiellen Gegenständen und spezifischen Merkmalen an einer bestimmten Lokalität zusammensetzen (Ore, Doerr, Le Bœuf, & Stead, 2015). Die Beschreibung der Entität entspricht der Semantik der Relation und muss deshalb als *domain*-Zielentität im 3M-Mappingprozess deklariert werden. Jeder einzelne Fundplatz wird als *domain* der Klasse *E27 Site* gemappt. Insgesamt weist die *domain*-Quellentität in ihrer Quellschemainterpretation sieben Eigenschaften bzw. Quellbeziehungen auf, die auf das Zielschema übertragen werden müssen.

- Der/ein site hat eine ID
- Der/ein site hat einen site code
- Der/ein site hat einen site name
- Der/ein site hat eine Länge
- Der/ein site hat eine Breite
- Der/ein site hat eine Koordinate des Breitengrades
- Der/ein site hat eine Koordinate Längengrades
- Der/ein site hat eine geometrische Repräsentation

Im Gegensatz zur Mappingtabelle der findings werden die sites nicht mehr mit deren datenbankinternen IDs gemappt. Da die Tabelle sites als Hilfsrelation der Haupttabelle angelegt ist, müssen folglich keine Joins mehr in den Hilfsrelationen umschrieben werden.

Da das Attribut *side_code* nicht nur als lokaler Identifikator der Fundstätten fungiert, sondern zum Teil auch mit Bezeichnern und Akronymen des *Sistema de Gerenciamento do Patrimônio Arqueológico (SGPA)* referenziert, muss die Instanz im Hinblick auf ein mögliches Linked Data Ökosystem als *E42 Identifier* gemappt werden. Von ähnlicher Bedeutung hinsichtlich der Vernetzung der Daten ist der Name der entsprechenden Fundstätte, der mit *E41 Appellation* instanziiert werden kann und mit hoher Wahrscheinlichkeit auch in der GeoNames-Datenbank (WEB7) referenziert ist. Anschließend werden die ermittelten Ausmaße *site length m* und *site width m* der Fundstätten gemappt. Ausgehend von der *domain* ist es dafür erforderlich einen Intermediate Node der Klasse *E54 Dimension* einzuführen, welche quantifizierbare Eigenschaften unabhängig von ihrer numerischen Annäherung instanziiert (Ore, Doerr, Le Bœuf, & Stead, 2015). Wiederum handelt es sich bei der Benennung der Spalten um wichtige implizite Informationen. Deshalb benötigt die Klasse aufgrund der Unabhängigkeit ihrer numerischen Annäherung zusätzlich eine Spezialisierung in Form von zwei konstanten Typuzuweisungen. Die *E54 Dimension* wird dementsprechend über *P91 has unit* zur Klasse *E58 Measurement Unit* in Beziehung gesetzt, die in Metern vorliegt. Die zweidimensionale Ausdehnung der Fläche der Fundstätte in horizontaler Richtung ist lediglich mit zwei Werten angegeben. Analog zu einer Ellipse beschreibt der größere Wert die Hauptachse als Länge und wird als *E55 Type* dem konstanten Wert „length“ zugewiesen. Der kleinere Wert beschreibt die

Nebenachse als Breite mit dem konstanten Wert „width“. Die Konstanten werden mit Hilfe von Konzepten aus dem URI-Basistemplate adressiert.

Die gespeicherten Raumkoordinaten können als numerische Information direkt mit der Klasse *E47 Spatial Coordinates* zur Lokalisation der Instanz *E27 Site* abgebildet werden. Jedes reguläre räumliche Bezugssystem kann dazu verwendet werden. Im vorliegenden Fall handelt es sich um Längen- und Breitengrade des WGS84-Koordinatensystems. Durch die Klasse *E55 Type* werden die Koordinatenangaben durch die Konstanten „latitude“ und „longitude“ typisiert und mit Hilfe von Konzepten aus dem URI-Basistemplate im Kapitel 3.3 adressiert.

Das Attribut *geom* der Relation *sites* gibt Auskunft über Position der Fundstätten und das vorliegende Referenzsystem der numerischen Information. Mit der Klasse *SP6 Declarative Place* kann die Position eines Ortes definiert werden (Hiebel, Doerr, Eide, & Theodoridou, 2015). Das reale Ausmaß der Fundstellen ist wie der geometrische Ausdruck in Form eines Punktes nur eine Abstrahierung oder Annäherung an die reale Ausdehnung und Position der Fundstelle. Deshalb wird die Klasse *SP6 Declarative Place* über die Eigenschaft *Q12 approximates* zu *E27 Site* in Beziehung gesetzt. Durch die Klasse *E55 Type* kann die vorliegende Kodierung durch die Konstante „hexEWKT“ typisiert und mit Hilfe von Konzepten aus dem URI-Basistemplate im Kapitel 3.3 adressiert.

4.3.3. Mapping der Hilfsrelation <archaeologists>

#	SOURCE	TARGET	CONSTANT
3	D  //archaeologists/row	 crm:E7_Activity	
3.1	P ↓ archaeologist_name	↓ crm:P14_carried_out_by  crm:E39_Actor  [act1]	
	R  archaeologist_name	↓ crm:P1_is_identified_by  crm:E41_Appellation	
3.2	P ↓ archaeologist_from	↓ crm:P11_had_participant  crm:E39_Actor  [act1] ↓ crm:P11i_participated_in  crmarchaeo:A9_Archaeological_Excavation  [ae1]	
	R  archaeologist_from	↓ crm:P4_has_time-span  crm:E52_Time-Span ↓ crm:P82a_begin_of_the_begin  http://www.w3.org/2001/XMLSchema#dateTime	
3.3	P ↓ archaeologist_to	↓ crm:P11_had_participant  crm:E39_Actor  [act1] ↓ crm:P11i_participated_in  crmarchaeo:A9_Archaeological_Excavation  [ae1]	
	R  archaeologist_to	↓ crm:P4_has_time-span  crm:E52_Time-Span ↓ crm:P82b_end_of_the_end  http://www.w3.org/2001/XMLSchema#dateTime	

Abbildung 22: Mapping der Hilfsrelation <archaeologists> (Printscreen)

Abbildung 22 zeigt das gesamte Mapping der Tabelle <archaeologists>. Die Hilfsrelation lässt sich in drei komplexen semantischen Pfaden abbilden. Die Relation, also die Quellentität *archaeologists* muss dabei analog zum Mapping der Hauptrelation wiederum durch die gleiche Klasse E7 Activity repräsentiert sein. Insgesamt weist die domain in ihrer Quellschemainterpretation drei Eigenschaften bzw. Quellbeziehungen auf die auf den Zielpfad übertragen werden müssen.

- archaeologist name hat das Artefakt gefunden
- Der Archäologe nimmt an der Aktivität/Grabung teil im Jahr archaeologist from
- Der Archäologe hört auf mit der Aktivität/Grabung im Jahr archaeologist from

Die Art und Weise wie in der Datenbank die Verknüpfung der Daten zwischen Artefakt, Archäologe und Fundstelle angelegt ist, setzt einige implizite Informationen voraus. Deshalb kann das Artefakt nicht direkt mit dem Namen seines Finders *archaeologist name* in Beziehung gesetzt werden. Vielmehr muss der gesamte Fundablauf semantisch umgesetzt werden, der in den Daten inhärent ist. Wie bereits erwähnt ist selbst das Finden eines Artefaktes bereits eine bestimmte Aktivität, die ontologisch abgebildet werden muss. Da die Vorstellung von der Klasse *E7 Activity* auch komplexe, zusammengesetzte und lang andauernde Aktionen umfasst (Ore, Doerr, Le Bœuf, & Stead, 2015) aber in diesem Zusammenhang eine reale Person voraussetzt, die diese Aktion ausführt, kann sowohl das Finden als auch die Teilnahme an einer Ausgrabung als allgemeine abstrahierte Aktion sinnvoll in Zusammenhang mit seinem Finder *E39 Actor* in Kontext gesetzt werden. Erst die Person, die den Namen trägt, kann daraufhin mit der Klasse *E41 Appellation* identifiziert werden.

Die folgenden Attribute *archaeologist from* und *archaeologist to* beinhalten zeitliche Informationen über die Teilnahme der genannten Person an einer Grabung. Auch in diesem Fall werden Entitäten als implizit vorausgesetzt. Der Beginn der Teilnahme einer Person an einer Grabung setzt zum einen die Person und zum anderen die Grabung selbst in einem semantisch sinnvollen Zusammenhang voraus. Deshalb müssen beide Entitäten zunächst mit Hilfe der beiden Intermediate Nodes *E39 Actor* und *A9 Archaeological Excavation* in Zusammenhang gebracht werden, die für die Angabe der Teilnahmezeit unabdingbar ist. Jede Aktivität einer CRM-Entität wird von der gleichen Person ausgeführt. Dieser Zusammenhang kann mit der Zuweisung „act1“ Entität *E39 Actor* als „same as“-Variable umgesetzt werden. Ebenso setzt das Finden eines Artefakts die Teilnahme der ausführenden Person an der identischen Grabung voraus. Deshalb muss die Entität im *findings*-Mapping und im *archaeologists*-Mapping mit der Variable „ae1“ gleichgesetzt werden. Im Anschluss daran kann der Teilnahmezeitraum an der Grabung mit einem RDF-Literal instanziiert werden.

4.3.4. Mapping der Hilfsrelationen <traditions> und <cultures>

#	SOURCE	TARGET	CONSTANT
7	D  //traditions/row	 crm:E4_Period	
7.1	P ↓ tradition_name	↓ crm:P1_is_identified_by	
	R  tradition_name	 crm:E41_Appellation de	
7.2	P ↓ tradition_from	↓ crm:P4_has_time-span	
		 crm:E52_Time-Span  [ts1]	
	R  tradition_from	↓ crm:P82a_begin_of_the_begin	 http://www.w3.org/2001/XMLSchema#dateTime
7.3	P ↓ tradition_to	↓ crm:P4_has_time-span	
		 crm:E52_Time-Span  [ts1]	
	R  tradition_to	↓ crm:P82b_end_of_the_end	 http://www.w3.org/2001/XMLSchema#dateTime

Abbildung 23: Mapping der Hilfsrelation <traditions> (Printscreen)

#	SOURCE	TARGET	CONSTANT
6	D  //cultures/row	 crm:E4_Period	
6.1	P ↓ culture_name	↓ crm:P1_is_identified_by	
	R  culture_name	 crm:E41_Appellation de	
6.2	P ↓ culture_from	↓ crm:P4_has_time-span	
		 crm:E52_Time-Span  [ts2]	
	R  culture_from	↓ crm:P82a_begin_of_the_begin	 http://www.w3.org/2001/XMLSchema#dateTime
6.3	P ↓ culture_to	↓ crm:P4_has_time-span	
		 crm:E52_Time-Span  [ts2]	
	R  culture_to	↓ crm:P82b_end_of_the_end	 http://www.w3.org/2001/XMLSchema#dateTime

Abbildung 24: Mapping der Hilfsrelation <cultures> (Printscreen)

Abbildung 23 und Abbildung 24 zeigen das gesamte Mapping der Tabellen <traditions> und <cultures>. Beide Hilfsrelationen lassen sich identisch in einem einfachen und zwei komplexen semantischen Pfaden abbilden. Die Relationen, also die Quellentitäten *traditions* und *cultures* müssen dabei analog zum Mapping der Hauptrelation *findings* wiederum durch die gleiche Klasse *E4 Period* repräsentiert sein. Insgesamt

weist die domain in ihrer Quellschemainterpretation drei Eigenschaften bzw. Quellbeziehungen auf, die auf den Zielpfad übertragen werden müssen.

- Die tradition/culture hat einen tradition name
- Die tradition/culture hat einen Anfangszeitpunkt
- Die tradition/culture hat einen Endzeitpunkt

Aus archäologischer Sicht werden historische Perioden häufig unterschiedlich interpretiert oder definiert. Deshalb existieren in vielen Fällen zu den Bezeichnern der historischen Perioden mehrere verschiedene Zeiträume mit unterschiedlichen Anfangs- und Endzeitpunkten. Auch im Hinblick auf die vorliegenden historischen Perioden liegt keine allgemein anerkannten Grenzen der Anfangs- und Endzeitpunkte vor. Diese sind stattdessen nach wie vor Diskussionsgegenstand aktueller Forschung (Neves, 2008). Deshalb wird zur Abbildung einer sinnvollen Semantik die Klasse *E4 Period* verwendet (Doerr, Theodorou, Aspöck, & Masur, 2015). Das Attribut *tradition/culture name* kann ausgehend von der domain über die Beziehung *P1 is identified by* zur Klasse *E41 Appellation* instanziiert werden.

Das Attribut *tradition/culture from* gibt den Anfangszeitraum an, der durch die Klasse *E52 Time Span* abgebildet werden kann. Analog dazu gibt das Attribut *tradition/culture to* den Zeitraum an, ab dem der jeweilige Keramikstil nicht mehr in Grabungszusammenhängen vergesellschaftet ist. Sowohl Beginn als auch Endzeitraum beziehen sich mit ihren Angaben auf die gleiche Periode. Deshalb muss die Entität *E52 Time Span*, die jeweils den Anfang und das Ende einer historischen Periode angeben, mit der „same as“-Variable „ts1“ bzw. „ts2“ gekennzeichnet sein.

4.3.5. Mapping der Hilfsrelation <bib>

#	SOURCE	TARGET
3	D  //bib/row	 crm:E31_Document
3.1	P  bib_name	 crm:P48_has_preferred_identifizier
	R  bib_name	 crm:E42_Identifizier

Abbildung 25: Mapping der Hilfsrelation <bib> (Printscreen)

Abbildung 25 zeigt das gesamte Mapping der Tabelle <bib>. Die Hilfsrelation lässt sich durch ein einfaches *one-to-one* Mapping umsetzen. Die Relation, also die Quellentität *bib* muss dabei analog zum Mapping der Hauptrelation wiederum durch die gleiche Klasse *E31 Document* abgebildet werden.

Jede bibliographische Angabe wird als ein Objekt der Klasse *crm:E31 Document* gemappt. Jede Instanz dieser Klasse stellt ein Dokument dar, in dem das entsprechende Artefakt beschrieben wird. Allerdings handelt es sich in der Datenbank nur um die Verweise als Abkürzungen. So enthält das Attribut *bib_name* ausschließlich bibliographische Akronyme bestehend aus Autor(-schaft) und Publikationsjahr und muss dementsprechend über die Beziehung *P48 has preferred identifizier* zu *E42:Identifizier* zugeordnet werden.

4.3.6. URI-Generierung

Auf die Schema-Mapping-Definition folgt die Definition der URI-Generation Policy. Das Ergebnis der beiden Schritte ist eine vollständige X3ML Mapping-Definitionsdatei, die von der X3ML-Engine für die Transformation der Daten in RDF verarbeitet wird (Minadakis, et al., 2015). Wie bereits erwähnt ist ein Satz vordefinierter URI-Generatoren (UUIDs, Literale) und Vorlagen im 3M bereits verfügbar. Zusätzlich können eigene URI-generierende Funktionen mit der URI-Generation Policy File implementiert werden. In der X3ML-Definition erhält die *target domain* und die *target range* Funktionen, die URIs, Literale oder beides generieren können.

Entsprechend des Default-Templates für die Instanzgeneratoren und dem URI-Basistemplate aus den vorigen Kapiteln wurde für das Mapping eine eigene *URI*

Generation Policy File entworfen und in den 3M importiert.⁷ Mit Ausnahme des Mappings der Hilfsrelation <archaeologists> wurden für alle für *domain*-Zielentitäten eigene URIs nach folgendem Schema erzeugt.

```
<generator_policy>
  <generator name="TabellennameURI" prefix="tp">
    <pattern>Tabellenname/{term}</pattern>
  </generator>
  ...
</generator_policy>
```

Das Präfix "tp" steht stellvertretend für <http://www.tp.de/database/>, einen selbst definierten Namespace, der nicht durch Quell- oder Zielschemata vorgegeben ist, aber in der X3ML-Definitionsdatei noch vor dem Mapping deklariert werden muss.⁸

Das für die TerraPreta Daten konzipierte URI Basistemplate wurde bereits in Kapitel 3.3 erläutert, deshalb wird an dieser Stelle nur noch einmal auf die beiden wesentlichen Punkte für die Konzeption der URIs im 3M eingegangen.

1. nur für die *domains* bzw. relationsäquivalenten Entitäten werden URIs generiert
2. die Spaltenwerte werden nur dann mittels RDF-Label migriert, wenn die *target range* keine direkte Instanziierung des Wertes durch RDF-Literal zulässt.

Das folgende Beispiel zeigt eine gesamte CRM-Entität mit den migrierten Daten des Tupels mit dem Primärschlüsselwert „2“ der Spalte *tradition_id* aus der Tabelle <traditions> im RDF/XML-Format.

⁷ Mit freundlicher Unterstützung von Maria Theodoridou vom FORTH ICS

⁸ Kann im 3M einfach über die Benutzeroberfläche eingegeben werden

```

<crm:E4_Period rdf:about="http://www.tp.de/database/Tradition/2">
  <crm:P4_has_time-span>
    <crm:E52_Time-Span rdf:about="uuid:AABT">
      <crm:P82b_end_of_the_end>3000</crm:P82b_end_of_the_end>
      <crm:P82a_begin_of_the_begin>7000</crm:P82a_begin_of_the_begin>
    </crm:E52_Time-Span>
  </crm:P4_has_time-span>
  <crm:P1_is_identified_by>
    <crm:E41_Appellation rdf:about="uuid:AABS">
      <rdfs:label>Früh-Keramisch</rdfs:label>
    </crm:E41_Appellation>
  </crm:P1_is_identified_by>
</crm:E4_Period>

```

Bei der Instanziierung mittels RDF-Literal werden die Spaltenwerte direkt als Werte von CRM-Property-Elementen gemappt. Die Entität `crm:E41 Appellation` erhält den Spaltenwert „Früh-Keramisch“ als RDF-Label. Entsprechend der konzeptionellen Definition ist der migrierte Primärschlüsselwert aus der Tabelle `<traditions>` in der URI der *domain* `crm:E4 Period` zu finden.

Zur Veranschaulichung zeigt Abbildung 26 einen Ausschnitt der Benutzeroberfläche des 3M Generators.

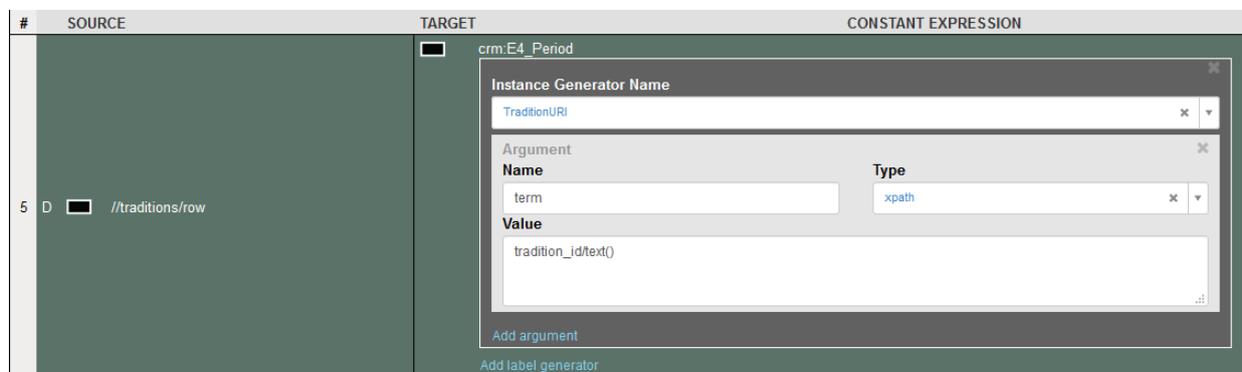


Abbildung 26: URI-Instanzgenerator der *domain*-Zielentität des Mappings `<traditions>` (Printscreen)

Durch die Wahl des Argumenttyps `xpath` kann der Primärschlüsselwert aus dem XML-Kindelement `tradition_id` der XML-Quelldatei in die URI migriert werden, dessen gültiger `xpath`-Pfad im `value`-Bereich gesetzt werden muss. Auf diese Weise wird die URI mit dem Präfix `rdf:about` als XML-Attribut für die domain-Zielentität `crm:E4 Period` in RDF/XML erzeugt.

Da in der Tabelle <findings> auch die Hilfsrelationen referenziert sind, müssen auch im Mapping <findings> alle *range*-Zielentitäten mit diesem URI-Schema versehen werden, die im Quellpfad auf eine andere Tabelle verweisen.

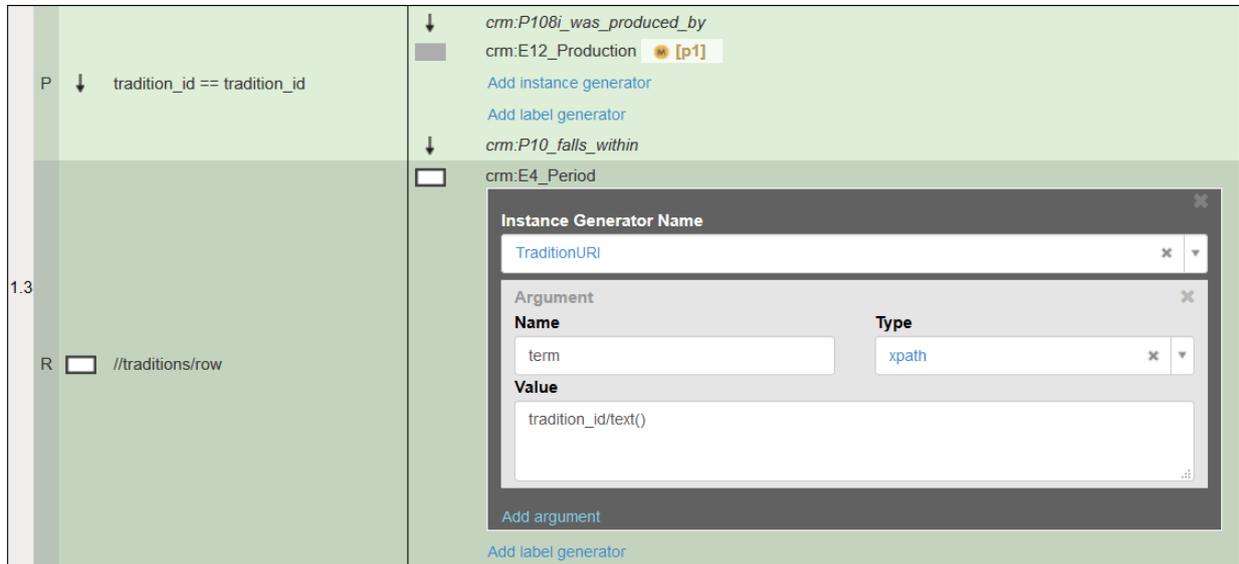


Abbildung 27: URI-Instanzgenerator der relationsäquivalenten Entität im Mapping <findings> (Printscreen)

Abbildung 27 zeigt die URI Generierung im Mapping <findings> für die Entität <crm:E4 Period>, die bei Fremdschlüsselreferenzierung in die CRM-Entität der Hauptrelation integriert wird.

Lediglich für die URI Generierung der Entität <crm:E7 Activity> wird ein anderes Schema gewählt. Da die Tabelle <archaeologists> in dieser Konstellation als Aktivität semantisiert werden muss, wird die URI auch zunächst mit dem englischen Begriff *Activity* fortgeführt. Da es theoretisch vorkommen kann, dass die gleiche Person mehrfach an der gleichen Grabung teilnimmt, wird die Anzahl der XML-Quellelemente mitgezählt und an zweiter Stelle in URI eingefügt, um redundante Ressourcenadressierung zu vermeiden. Zuletzt wird der Name des Teilnehmenden angehängt.

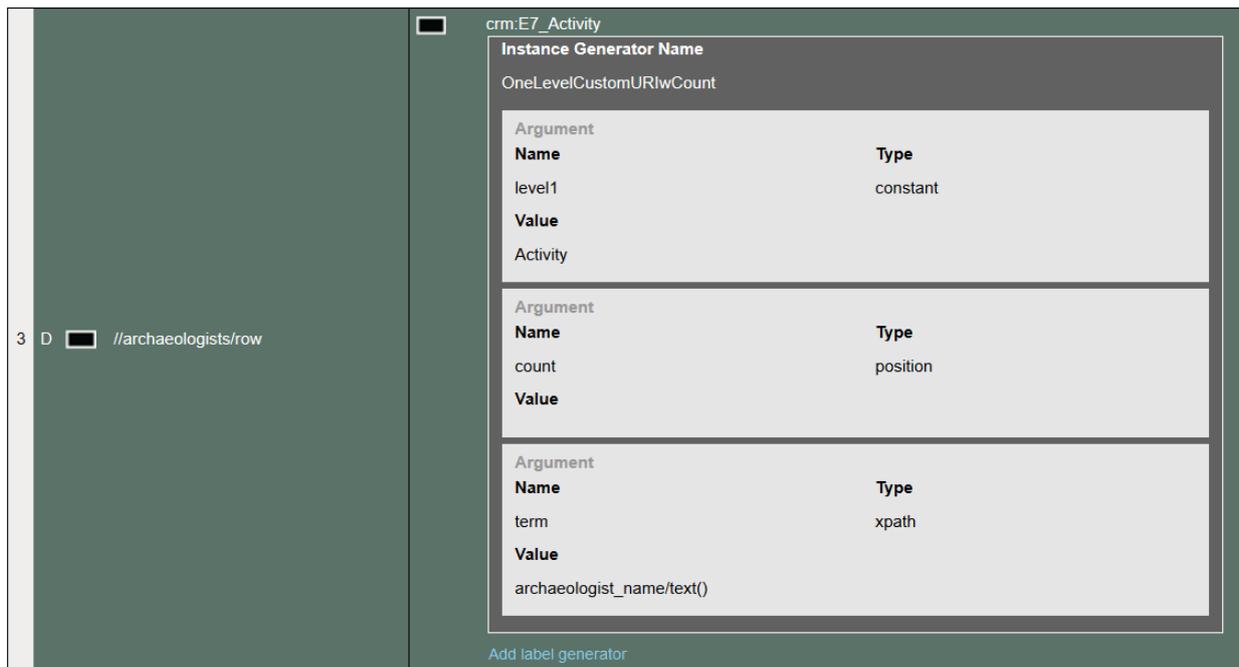


Abbildung 28: URI-Instanzgenerator der domain-Zielentität des Mappings <archaeologists> (Printscreen)

Nach Abschluss des letzten Schritts kann die Transformation der TerraPreta-Daten in ein gültiges RDF-Format erfolgen, das im Anhang mit dem angegebenen Testdatensatz realisiert worden ist. Der RDF-Graph im Anhang wurde mit Hilfe des RDF-Validators des W3C (WEB8) erstellt und bei der Erstellung auf seine Gültigkeit überprüft.

5. Ergebnisse und Diskussion

5.1. Ergebnisse

Im Hinblick auf die vorangestellten Fragestellungen des ersten Komplexes konnte gezeigt werden, dass mit Hilfe des angeführten Vergleichs der verschiedenen Ontologien das CIDOC CRM nicht ausschließlich, jedoch am besten zur Abbildung der Semantik der TerraPreta Daten geeignet ist. Keine andere Ontologie bietet eine derartige Klassenvielfalt, welche für die notwendige exakte semantische Modellierung der archäologischen TerraPreta Daten unabdingbar ist. Der Vorteil der eventbasierten Konzeption des CRM zeigt sich besonders bei der Verknüpfung zwischen temporären und persistenten Klassen, die für die Modellierung von zeitlich unabhängigen Entitäten allerdings

nicht zwingend ist. Dennoch muss festgehalten werden, dass eine anschließende Integration des vorliegenden Modells mit dem EDM als Top-Level Ontologie durchaus in Betracht gezogen werden kann.

Ferner dient das *database to ontology mapping* von Ghawi und Cullot (2007) als methodischer Ansatz beim notwendigen Prozess der konzeptionellen Festlegung der semantischen Korrespondenzen zwischen den Entitäten der TerraPreta-Datenbank und den Klassen der CIDOC CRM Ontologie. Das mittel- bis langfristig übergeordnete Ziel der Integration der TerraPreta-Daten ins Semantic Web erfordert die Übertragung der Informationen aus der Datenbank in das Resource Description Framework. Auf technischer Ebene muss vor der Mapping Definition mit dem 3M zunächst eine normalisierte XML-Datei auf Basis des *effective provider schemas* nach Minadakis et al. (2015) erstellt werden. Der Mappingprozess findet mehrstufig auf dem Web-Interface des 3M statt und beinhaltet zunächst die Auswahl des Quellschemas und das Zusammenfügen der benötigten RDFS-Schemata bzw. CRM-Erweiterungen in Form einer Applikationsontologie als Zielschema. Die beiden nachfolgenden Schritte sind Mapping Definition und URI-Generierung. Die Transformation und Migration der Daten aus der XML-Quelldatei, sowie die Zuordnung der URIs und die anschließende Überführung in das Resource Description Framework findet durch die X3ML-Engine des 3M in Form der X3ML-Definitionsdatei vollautomatisiert statt. Dies erweist sich als besonderer Vorteil, da die Mapping Definition nach wie vor einen zeitintensiven Prozess darstellt. Gleichzeitig dient die Automatisierung weitestgehend zur Vermeidung von Flüchtigkeits- oder Syntaxfehlern bei der Semantisierung von Daten, die wiederum für sich häufig zu weiteren zeitintensiven Arbeitsanteilen führen können.

Bedingt durch die methodisch-technischen Voraussetzungen, die im ersten Komplex erarbeitet worden sind, konnte herausgestellt werden, dass die Erstellung des *effective provider schemas* durch eine bestimmte Art des XML-Exports der PostgreSQL-Datenbank unterstützt wird und einfach umsetzbar ist. Für die exemplarische Migration und Transformation der Daten in eine RDF-Datei wurde eigens ein Testdatensatz aus der Datenbank zusammengestellt. Die einfache Umsetzung des XML-Exports erscheint im Hinblick für die Wahl von Projektdatenbanken archäologischer Unternehmen ein relevanter Faktor, insofern es beabsichtigt ist zumindest mittel- bzw. langfristig einer Silohaltung von Daten entgegenzuwirken und so die Vernetzung von archäologischen Projektdaten im Semantic Web voranzutreiben. Zur Umsetzung einer

sinnvollen Semantik ist es jedoch erforderlich das zu Grunde liegende Datenbankdesign zuvor durch Verkettung der Spalten mit den C14-Daten zu ändern.

Zur Abbildung des TerraPreta-Quellschemas ist ein Komplex aus mehreren kompatiblen Domänenontologien des CRM nötig, die als lokale Applikationsontologie dienen. Die für das Mapping benötigte Applikationsontologie setzt sich zum Großteil aus Klassen des CRM selbst und vier weiteren thematischen Erweiterungen der Ontologie zusammen. Für jede Tabelle der Datenbank muss eigens eine Mapping Definition erstellt werden in der die Bedeutungszusammenhänge der Informationen untereinander mit Hilfe von Klassen und Eigenschaften beschrieben werden. Besonders wichtig ist dabei auch die Umsetzung von relationalen Joins, deren Zusammenhang ebenso als semantischer Pfad formuliert wird. Die anschließende URI-Generierung ist für die Validität der RDF Datei unabdingbar, auf konzeptioneller Ebene genügt dazu allerdings die Verwendung von UUIDs, die bereits im Basis-Template des 3M vorhanden sind. Auf Grundlage des Default-Templates nach Marketakis et al. (2017) wurde eine URI Generation Policy File entworfen, deren Konzeption allerdings nur als rudimentäre Vorüberlegung hinsichtlich der Vernetzung mit weiteren Datenbanken zu betrachten ist. Vielmehr steht die funktionale Umsetzung im Fokus der Arbeit. Aufgrund seiner komplexen Eigenschaften und Vielzahl an Entitäten ist das CIDOC CRM als Domänen-Ontologie für die spezielle Modellierung der Semantik von Artefakten im archäologischen Kontext hervorragend geeignet. Auch die versteckten Informationen und allgemeinen Kontextinformationen der TerraPreta-Daten können mit geeigneten Objekttypen modelliert werden. Die Integration der Geoinformationen, sowohl in ihrer numerischen als auch kodierten Repräsentation, ist mit den vorhandenen Klassen problemlos möglich. Die abschließend durch den 3M generierte RDF-Datei wurde mit Hilfe des RDF-Validators des W3C auf Gültigkeit überprüft und visualisiert.

Das zu Beginn in Ausblick gestellte Ziel der vorliegenden Arbeit in Form einer vollständig ausformulierten und validierten RDF Datei konnte auf Basis des erstellten Testdatensatzes erreicht werden. Eine Transformation der Gesamtheit aller Daten der TerraPreta Datenbank in RDF konnte jedoch aufgrund von Performanceschwierigkeiten der Webapplikation nicht erreicht werden, kann aber durch Teilen des Datensatzes dennoch umgangen werden.

5.2. Diskussion

Darüber hinaus ergeben sich bei näherer Betrachtung der Daten bestimmte Aspekte, die mit der Semantik des vorgeschlagenen Modells konfliktieren und deshalb im Folgenden zur Diskussion stehen.

Wie bereits erwähnt repräsentiert in der angepassten und reduzierten Version die Tabelle <findings> einzelne Funde, welche die an den Siedlungsplätzen vorgefundenen Keramikscherben nach Tradition und deren Kulturen unterscheidet und kann deshalb folgerichtig mit der Entität *E52 Man Made Object* semantisiert werden. Allerdings unterstützt die Angabe der Fundtiefe der Artefakte die vorgeschlagene Semantik nicht immer eindeutig. Als eines von mehreren Beispielen zeigt Tabelle 4 einen Datenauszug von fünf verschiedenen Keramikscherben, die alle an der gleichen Fundstätte Aviso I gefunden wurden. Die gezeigten Funde sind mit der exakt gleichen Fundtiefe von 120 cm dokumentiert. Bei einer ungestörten Horizontstratigraphie ist aufgrund des zeitlichen Abstands der Entstehung bei dem Artefakt mit der Nummer 791 im Vergleich zu den älteren ein geringerer Wert zu erwarten.

Tabelle 4: optimierter Auszug ausgewählter Daten zur Fundstelle Aviso I

finding id	site name	finding depth cm	trad name	trad from	trad to	cult name	cult from	cult to	c14 year
791	Aviso I	120	Rit-zung-Punk-tierung	1000	500	Konduri	800	400	550 +/- 40 BC
792	Aviso I	120	Poly-chrom-Früh	3000	1500	Pocó	2500	1500	2210 +/- 50 BC
793	Aviso I	120	Poly-chrom-Früh	3000	1500	Pocó	2500	1500	2020 +/- 60 BC
794	Aviso I	120	Poly-chrom-Früh	3000	1500	Pocó	2500	1500	2930 +/- 50 BC
795	Aviso I	120	Rit-zung-Punk-tierung	1000	500	Tapajós			

Allerdings kann die natürliche regionale Bodenentwicklung ein solches Bild rechtfertigen, da diese in den Tropen sogar mehrere Jahrtausende umfassen kann und überdies durch bio- und anthropogene Einflüsse gestört werden kann (Thiemeyer, 2008).

Darüber hinaus enthält die Tabelle Informationen zu konventionellen unkalibrierten Radiocarbon-Datierungen. Als logische Folge sollte die durch die C14 Proben ermittelte Datierung innerhalb der relativchronologischen Zeiträume liegen und muss daher auf Gültigkeit überprüft werden.

Ferner wurden die vorhandenen C14-Daten verwendet, um die Dauer der jeweiligen historischen Intervalle näher zu spezifizieren (von Groote-Bidlingmaier, 2016).

Im vorliegenden Fall sind die Ergebnisse der Datierung als faktische Informationen speziell den Keramikfunden zugeordnet. Die Zuweisung der C14-Datierung zu einem einzelnen Artefakt ist allerdings nur in den folgenden drei Fällen semantisch korrekt, da nur in diesen Fällen die Probe für die C14 Datierung direkt vom Artefakt abhängt (Radiokohlenstoff Datierung von Keramik, (o.D.)):

1. wenn die Probe zur Datierung von verkohlten Speiserückständen von der inneren Keramikoberfläche genau dieses Artefakts stammt.
2. wenn die Datierung von einer Probe, die aus der Tonscherbe gewonnen Biomasse erfolgt, die weder vom Brennen oder organischen Stoffen, die durch Aufbewahrung von Speisen und Flüssigkeiten absorbiert wurden, beeinflusst wurde.
3. wenn die Datierung aufgrund einer Probe der extrahierbaren Brennmaterialien bzw. Magerungsmitteln aus dem Artefakt selbst erfolgt.

In allen anderen Fällen muss die Zuordnung der C14 Daten zur Fundstätte selbst in Betracht gezogen werden, da die Datierung dann nicht mehr explizit zum Artefakt selbst, aber immernoch mit dem Fundplatz in Beziehung gesetzt werden kann. Bei einer Ergänzung um C14 Daten von Proben die nicht direkt vom Artefakt selbst stammen muss dieser Sachverhalt in der Umsetzung einer Erweiterung des semantischen Modells verwirklicht werden.

Eine weitere Anomalie ergibt sich aus der Tatsache, dass in der Datenbank heterogene Datensätze gespeichert sind. Ein Indiz zur Modellierung eines weiteren Ansatzes liefert folgende Aussage:

„*For now the database contains 665 excavation sites with 750 findings whereas a finding is not defined by single artifacts, (which might at first be intuitive) but rather by distinct cultures.*“ (von Groote-Bidlingmaier, 2016, S. 52). Demnach befinden sich in der Relation <findings> nicht nur einzelne Artefakte, sondern teilweise auch andere Fundzusammenhänge, wie Schichtpakete oder Fundorte. Allerdings ist eine Rekonstruktion ohne die erneute Sichtung der Originaldaten und Publikationen nicht möglich, deshalb wurde dieser Sachverhalt in den vorangehenden Teilen der Arbeit ignoriert und zunächst ein Fall ausführlich diskutiert.

Als Folge der heterogenen Semantik der Datenbank ergibt sich für die gesamte Modellierung der Entwurf weiterer Mappings, die der Semantik der gespeicherten Datensätze entsprechen. Wurde eine Dateneinheit beispielsweise ursprünglich als Schichtpaket gespeichert, muss das semantische Modell der TerraPreta Datenbank ein weiteres Mapping beinhalten, beginnend bei der Hauptrelation <findings> mit der Ausgangsklasse *S22 Segment of Matter* semantisch abgebildet werden. Da die Ausgangsklassen beider Versionen des Mappings als Subklassen der CRM-Klasse *E18 Physical Thing* definiert sind, ändern sich ihre Vererbungseigenschaften nicht und es gibt demzufolge keine Auswirkung auf die Mappings der Hilfsrelationen. Darüber hinaus liegen alle bisher verwendeten Klassen der ebenso innerhalb der *range* der Ausgangsklasse der zweiten Mappingversion.

#	SOURCE	TARGET
1	D <input type="checkbox"/> //findings/row	<input type="checkbox"/> crmsci:S22_Segment_of_Matter
1.1	P ↓ finding_id	↓ crm:P1_is_identified_by
	R <input type="checkbox"/> finding_id	<input type="checkbox"/> crm:E42_Identifier
1.2	P ↓ site_id == site_id	↓ crmsci:O19i_was_object_found_by crmgeo:S19_Encounter_Event [enev1] ↓ crm:P117_occurs_during crmarchaeo:A9_Archaeological_Excavation [ae1] ↓ crm:P8_took_place_on_or_within
	R <input type="checkbox"/> //sites/row	<input type="checkbox"/> crm:E27_Site
1.3	P ↓ tradition_id == tradition_id	↓ crmgeo:Q3_has_temporal_projection crmgeo:SP13_Phenomenal_Time-Span [p1] ↓ crm:P4i_is_time-span_of
	R <input type="checkbox"/> //traditions/row	<input type="checkbox"/> crm:E4_Period
1.4	P ↓ culture_id == culture_id	↓ crmgeo:Q3_has_temporal_projection crmgeo:SP13_Phenomenal_Time-Span [p1] ↓ crm:P4i_is_time-span_of
	R <input type="checkbox"/> //cultures/row	<input type="checkbox"/> crm:E4_Period
1.5	P ↓ archaeologist_id == archaeologist_id	↓ crmsci:O19i_was_object_found_by crmgeo:S19_Encounter_Event [enev1] ↓ crm:P9i_forms_part_of
	R <input type="checkbox"/> //archaeologists/row	<input type="checkbox"/> crm:E7_Activity
1.6	P ↓ bib_id == bib_id	↓ crm:P70i_is_documented_in
	R <input type="checkbox"/> //bib/row	<input type="checkbox"/> crm:E31_Document
1.7	P ↓ c14_year	↓ crmgeo:Q3_has_temporal_projection crmgeo:SP13_Phenomenal_Time-Span [p1] ↓ crm:P86_falls_within crmgeo:SP10_Declarative_Time-Span ↓ crm:P82_at_some_time_within
	R <input type="checkbox"/> c14_year	<input type="checkbox"/> http://www.w3.org/2000/01/rdf-schema#Literal
1.8	P ↓ c14_info	↓ crm:P3_has_note
	R <input type="checkbox"/> c14_info	<input type="checkbox"/> http://www.w3.org/2000/01/rdf-schema#Literal
1.9	P ↓ finding_depth_cm	↓ crm:P53_has_former_or_current_location crm:E53_Place ↓ crm:P39i_was_measured_by crm:E16_Measurement ↓ crmsci:O12_has_dimension crm:E54_Dimension ↓ crm:P90_has_value
	R <input type="checkbox"/> finding_depth_cm	<input type="checkbox"/> http://www.w3.org/2000/01/rdf-schema#Literal

Abbildung 29: Alternatives Mapping der Hauptrelation <sites> (Printscreen)

Anhand dieser Beispiele wird ein wesentlicher Punkt bei der semantischen Modellierung deutlich. Durch den Zwang zur eindeutigen Beschreibung für ungenaue und variable Daten, wie sie im Umgang mit dem kulturellen Erbe oft auftreten, ergeben sich bei der Modellierung Probleme mit der exakten Semantisierung zur Erhaltung der ursprünglichen Bedeutung der Informationen. Deshalb sollte bei der Modellierung grundsätzlich der Urheber der Daten involviert sein. Das erweist sich jedoch besonders bei der Erhebung kultureller Daten als Herausforderung, da diese für neue Projekte mitunter auch aus mehreren unterschiedlichen Quellen zusammengezogen werden oder der Urheber nicht mehr lebt.

2014). Mit dem Ziel räumliche Abfragen mittels GEOSPARQL zu ermöglichen, können die zeitlichen Eigenschaften der CRM-Klassen mit der CRMgeo-Erweiterung integriert werden. Allerdings ist die derzeitige geometrische Datengrundlage der vorliegenden Datenbank bisher nicht ausreichend um räumliche Abfragen mittels GEOSPARQL zu ermöglichen.

Nach Hiebel et al. (2014) ist es im Genre der historischen und archäologischen Wissenschaften aufgrund von Faktoren wie unscharfen Grenzen von Phänomenen und Messfehlern jedoch unmöglich exakte raumzeitliche Eigenschaften von Entitäten wie E4 Periods oder E18 Physical Things zu kennen und diese Inhalte in ihrer Gesamtheit zu beschreiben. Zumindest ist es möglich sich der räumlichen historischen Realität im Zuge hypothetischer Diskurse anzunähern. Die Unterscheidung von Phänomenen zwischen beiden Welten beschränkt sich allerdings bisher nur auf zeitliche und geometrische Dimensionen. Deshalb schlagen van Ruymbeke, Hallot und Billen (2017) folgerichtig die Einführung des *Multiple Interpretation Data Model* vor. Es ermöglicht die Einhaltung einer exakten Semantik durch die grundsätzliche Unterscheidung zwischen realen Phänomenen und Informationen die sie umschreiben. Auf diese Weise ist es möglich durch die Darstellung verschiedener Interpretationen (*Interpretative Sequences*) gleichzeitig einen Diskurs zu realen historischen Phänomenen (*Historical Objects*), wie zum Beispiel Artefakten, sinnvoll semantisch abzubilden.

Im Hinblick auf die in der TerraPreta-Datenbank gespeicherten Informationen muss jedoch Folgendes festgehalten werden. Für die Ermittlung der Einzugsgebiete wurde sich bisher nur auf das Klassifikationsschema von Keramikartefakten beschränkt, welches 1977 im Zuge des PRONAPA-Projekts mit dem Ziel zur Definition von Kultursequenzen anhand von Keramikanalysen und -datierung ins Leben gerufen wurde (von Groote-Bidlingmaier, 2016). Das reduzierte Datenbankdesign kann aber auch zur Speicherung anderer Fundgattungen oder Perioden herangezogen werden und damit die Modellierung einen solchen Diskurs ermöglichen.

Bei der automatisierten Erstellung der RDF/XML Datei wird die Generierung eines identischen Identifikators der mit *“same as”-Variablen* gekennzeichneten Entitäten in einigen Fällen funktional nicht übernommen. Bisher sind alle Entitäten mit dieser Kennzeichnung mit unterschiedlichen UUIDs versehen, die innerhalb einer in RDF/XML ausformulierten CRM-Entität entsprechend ihrer Kennzeichnung eigentlich

identische Identifikatoren haben müssten. Folgender Abschnitt zeigt einen Ausschnitt der im Anhang beigefügten RDF/XML die vom Memory Mapping Manager erstellt wurde.

...

```
<crm:E7 Activity rdf:about="http://www.tp.de/database/Activity/1/Steere">
  <crm:P11 had_participant>
    <crm:E39 Actor rdf:about="uuid:AAAS">
      <crm:P11i participated_in>
        <crmarchaeo:A9 Archaeological_Excavation rdf:about="uuid:AAAU">
          <crm:P4 has_time-span>
            <crm:E52 Time-Span rdf:about="uuid:AAAW">
              <crm:P82b_end_of_the_end>1871</crm:P82b_end_of_the_end>
            </crm:E52 Time-Span>
          </crm:P4 has_time-span>
          <crm:P4 has_time-span>
            <crm:E52 Time-Span rdf:about="uuid:AAAV">
              <crm:P82a_begin_of_the_begin>1871</crm:P82a_begin_of_the_begin>
            </crm:E52 Time-Span>
          </crm:P4 has_time-span>
        </crmarchaeo:A9 Archaeological_Excavation>
      </crm:P11i participated_in>
    <crm:P1 is_identified_by>
      <crm:E41 Appellation rdf:about="uuid:AAAT">
        <rdfs:label>Steere</rdfs:label>
      </crm:E41 Appellation>
    </crm:P1 is_identified_by>
  </crm:E39 Actor>
</crm:P11 had_participant>
<crm:P14 carried_out_by rdf:resource="uuid:AAAS"/>
</crm:E7 Activity>
</crm:P9i forms_part_of>
<crm:P117 occurs_during>
  <crmarchaeo:A9 Archaeological_Excavation rdf:about="uuid:AAAC">
    <crm:P8 took_place_on_or_within>
      <crm:E27 Site rdf:about="http://www.tp.de/database/Site/1">
        <crmgeo:Q11i is_approximated_by>
          <crmgeo:SP6 Declarative_Place rdf:about="uuid:AABI">

```

Durch die Kennzeichnung mit der *“same as”-Variable „ae1“* in der Matching Tabelle des 3M müsste der in Anführungszeichen stehende Ausdruck identisch sein. Das **rdf:about**-Attribut, dessen UUID die Ressource des Subjekts **crmarchaeo:A9_Archaeological_Excavation** referenziert, ist allerdings nicht gleich. Im Gegensatz dazu sind die UUIDs der Ressource mit der Bezeichnung **uuid:AAAS** korrekt. Denn dabei handelt es sich in diesem Fall um die gleiche Person. Tabelle 5 zeigt alle Entitäten für die die manuelle Korrektur relevant ist.

Tabelle 5: Entitäten mit identischen Identifikatoren

Entität	Mapping	Zeile
crmsci:S19 Encounter Event "enev 1"	findings	2.2
		2.5
crmarchaeo:A9 Archaeological Excavation "ae1"	findings	2.2
	archaeologists	3.2
		3.3
crm:E12 Production "p1"	findings	2.3
		2.4
		2.7
crm:E52 Time Span "ts2"	cultures	6.2
		6.3
crm:E52 Time Span "ts1"	traditions	7.2
		7.3

In Bezug auf die angeführte Fragestellung steht jedoch zunächst die Validität bzw. korrekte Syntax der RDF/XML im Vordergrund, deshalb wurde bei der Ausarbeitung auf eine Korrektur verzichtet. Es bedeutet aber, dass bei Einsatz von URIs im Hinblick auf ein Linked Open Data Ökosystem für alle diese Entitäten eine manuelle Nachbesserung des **rdf:about**-Attributs nach der Transformation der Daten mit dem 3M vorgenommen werden muss.

Als abschließenden Punkt stehen die in der Datenbank verwendeten Identifikatoren im Fokus. Diese sind nicht eindeutig und müssten bei der Verwendung in einem Linked Open Data Ökosystem eigentlich auf den CNAS Code des *Sistema de Gerenciamento do Patrimônio Arqueológico* (SGPA) referenzieren. Abbildung 31 zeigt einen Auszug aus dem SGPA mit einigen Fundstätten aus Amazonien (AM) mit Bezirksangabe.

Sistema de Gerenciamento do Patrimônio Arqueológico (SGPA)

AM00098	AM-MC-17: Povoação	Manicoré	AM
AM00099	AM-MC-18: Uruapiára	Manicoré	AM
AM00100	AM-MC-19: Acará	Manicoré	AM
AM00101	AM-MC-20: Bom Intento	Manicoré	AM
AM00102	AM-MC-21: Monense	Manicoré	AM
AM00103	AM-MC-22: Piraiba	Manicoré	AM
AM00104	AM-MC-23: Segundo	Manicoré	AM
AM00105	Tambos 1,3,4,7,10 e 12	São Gabriel da Cachoeira	AM
AM00106	Lago Parime	São Gabriel da Cachoeira	AM
AM00107	Clareira 16	Coari	AM
AM00108	Clareira 10	Coari	AM
AM00109	Clareira 14	Coari	AM
AM00110	Clareira 15	Coari	AM

Abbildung 31: Auszug aus der Fundstättendatenbank des SGPA mit CNAS Identifikatoren für Amazonien (WEB10)

Dort sind die Identifikatoren des CNAS links und die in der TerraPreta-Datenbank verwendeten Bezeichner in der zweiten Spalte von links zu finden. Die Daten der TP Datenbank haben andere Bezeichner und Akronyme, die sich nur teilweise mit denen aus dem SGPA decken. Der Code des *Cadastro Nacional de Sítios Arqueológicos*, der für jede Fundstätte einen eindeutigen Identifikator besitzt, ist in der TP Datenbank gar nicht vorhanden. Denn für eine Vernetzung zwischen der TerraPreta Datenbank und der CNAS müssen alle TP-Datensätze zusätzlich um die relevanten Identifikatoren der CNAS Datenbank erweitert werden.

Denn ein weiterer wesentlicher Bestandteil neben der semantischen Modellierung ist die Bereitstellung im Internet und die Verknüpfung eben dieser Daten mit anderen Wissensbeständen, da die Daten immer wertvoller werden, je stärker sie vernetzt und interpretierbar sind.

6. Konklusion

6.1. Ausblick in ein Linked Open Data Ökosystem

Bevor eine Bereitstellung der Informationen im Semantic Web als Linked Open Data stattfinden kann, ist es deshalb erforderlich, zunächst eine e-Infrastruktur bereitzustellen, die alle isolierten Systeme mit relevanten archäologischen Daten in ein eigenes Informationsnetz oder einem so genannten Linked Open Data Ökosystem zu integrieren. Das übergeordnete Ziel dieses Informationsnetzes ist es, der

Fragmentierung archäologischer Datenbestände entgegenzuwirken durch Bereitstellung einer e-Infrastruktur, welche archäologische Informationen und Datensätzen integriert, die von Rechenzentren in Institutionen oder anderen Archiven genutzt werden und in ganz Brasilien verteilt sind. Die Infrastruktur dient dazu, den Austausch und die gemeinsame Nutzung der verfügbaren Daten über organisatorische Grenzen hinweg zu gewährleisten. Mit ARIDANE gibt es auf europäischer Ebene bereits seit einiger Zeit solche Bestrebungen.

Deshalb sollen im folgenden Kapitel kurz die bereits existierenden isolierten Systeme benannt und die notwendigen Bestandteile einer Infrastruktur skizziert werden, die dazu dienen die Quellen zu vernetzen aus denen in Zusammenhang mit den TerraPreta Daten ein Mehrwert generiert werden kann.

Das Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN) ist eine mit dem Kulturministerium verbundene föderale Behörde Brasiliens, mit dem Ziel zur Bewahrung und Schutz des brasilianischen Kulturerbes und kulturellen Vermögens. Als fünftes Ziel des nationalen Kultur Plans zielt das nationale Kulturerbe-System (Sistema Nacional do Patrimônio Cultural [SNPC]) darauf ab, eine geteilte Verwaltung des brasilianischen Kulturerbes durchzuführen, um die personellen und finanziellen Ressourcen für seinen wirksamen Schutz zu optimieren (WEB9).

Dessen Architektur wurde deshalb in verschiedenen Modulen umgesetzt und anschließend von verschiedenen Behörden verwaltet. Darunter befinden sich das nationale Register der archäologischen Stätten (CNSA), eine Bild-Datenbank, das nationale Inventar der archäologischen Sammlungen (INCA), sowie Projekte und archäologische Forschungsberichte (PPA/RPA) (WEB10).

Diese Module bilden die wesentlichen Bestandteile, die für eine Vernetzung der archäologischen Daten in Form eines Linked Open Data Ökosystems in Frage kommen. Wie zu Beginn erwähnt, enthalten die Datensätze der archäologischen Stätten des CNSA keine Information über die georeferenzierte Lokation. Unabhängig von finanziellen Aspekten liegt bereits mit der Überführung der Daten zwischen der TerraPreta Datenbank und dem CNSA in ein Linked Open Data Ökosystem ein Informationsgewinn und damit ein Mehrwert zugrunde.

Im Hinblick auf dieses Ökosystem soll das Mapping der TerraPreta Datenbank exemplarisch für weitere Mappings innerhalb dieses Systems dienen.

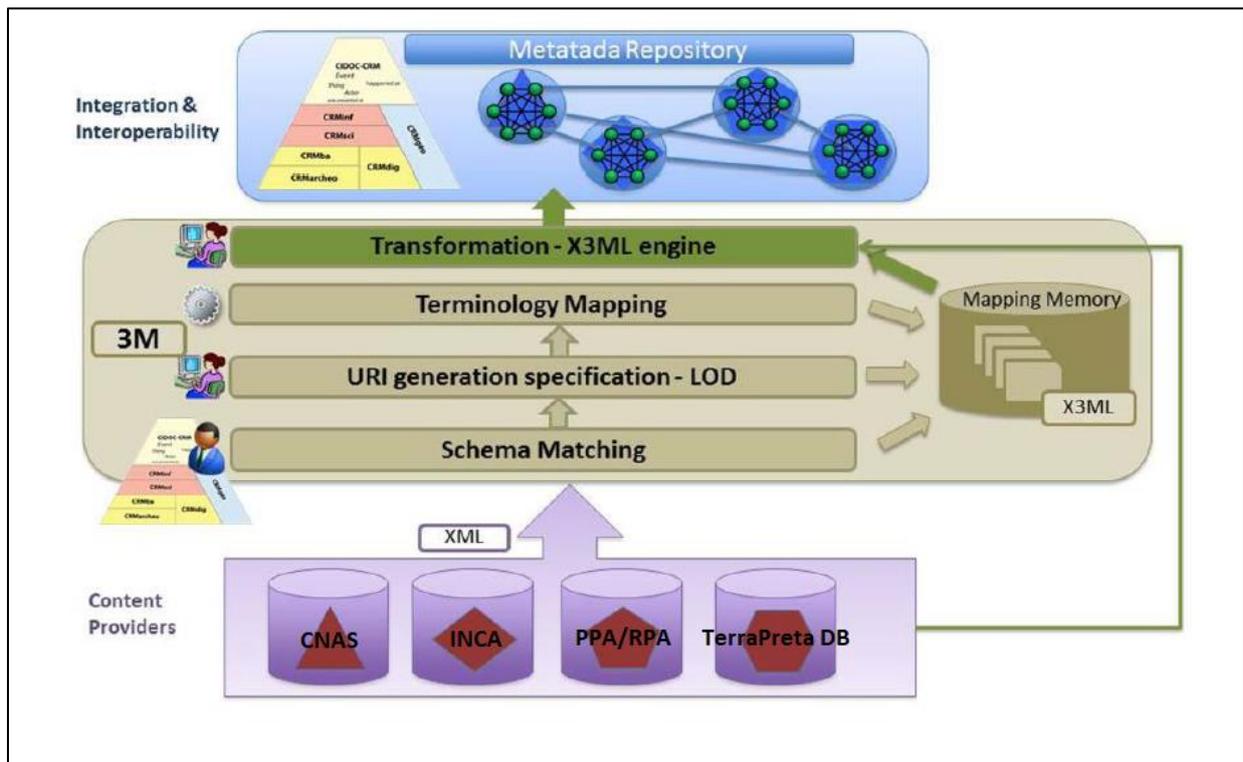


Abbildung 32: Mapping Workflow verändert nach (Doerr, Theodorou, Aspöck, & Masur, 2015, S. 4, Abb. 2)

Allerdings sind die Anforderungen an die semantische Interoperabilität beträchtlich. Denn für den nächsten Schritt ist es erforderlich ein semantisches Modell bzw. *Schema Matching* für jede Datenbank des SNPC zu erstellen. Dadurch entstehen weitere konzeptionelle Mappings auf Basis realer Daten, die es ermöglichen die Darstellung einer zusammenhängenden Semantik in RDF zu einem immer komplexeren Diagramm zu kombinieren. Als großes Vorbild für das Ökosystem kann die Architektur der ARIADNE Infrastruktur gesehen werden, als Teil dessen auch der Mapping Memory Manager entwickelt worden ist (Meghini, et al., 2017).

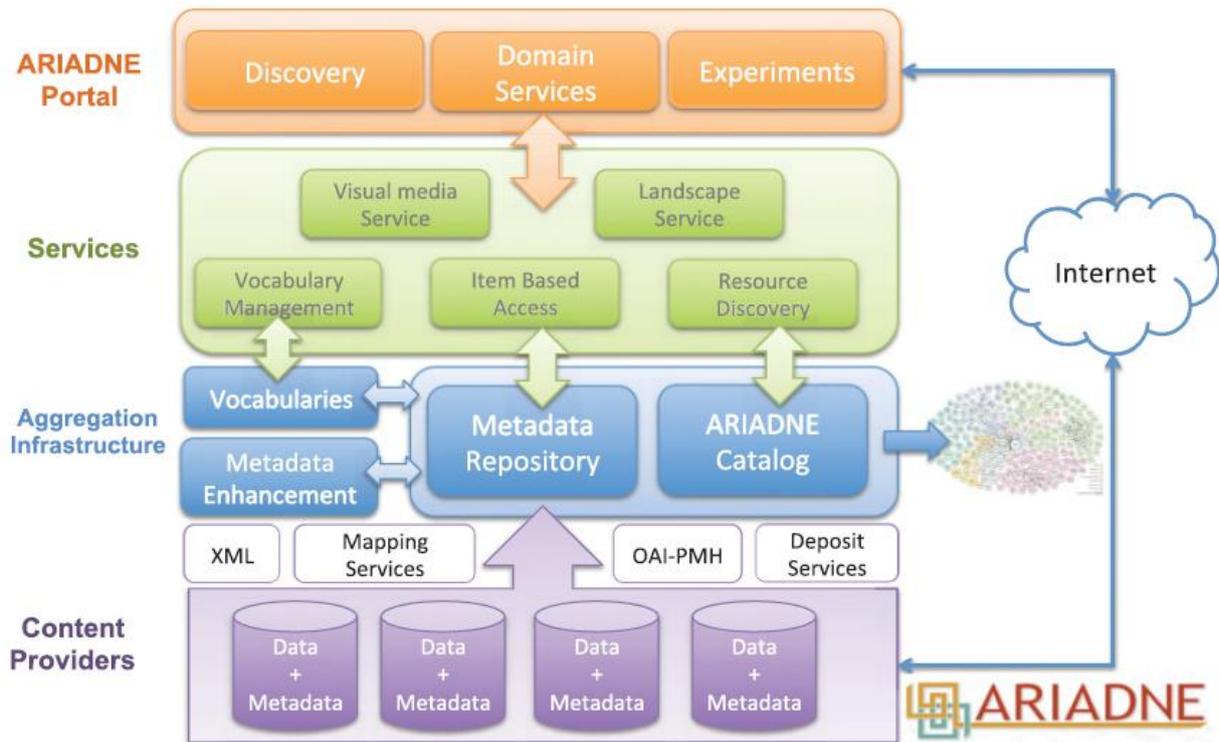


Abbildung 33: Architektur der ARIADNE e-Infrastruktur (Meghini, et al., 2017, S. 10, Abb. 1)

Zweifelsohne gilt es jedoch für die Verwirklichung eines solchen Mammut-Projektes noch unzählige Hürden zu überwinden.

6.2. Fazit

Im Falle von Datensätzen archäologischer Projekte, die in verschiedenen digitalen Archiven gespeichert sind, müssen die Metadaten in Resource Description Framework (RDF) umgewandelt werden und Begriffe eines gemeinsamen Vokabulars enthalten, die auch als verknüpfte Daten verfügbar sein müssen. Die Ontologie des CIDOC CRM dient dazu Begriffe eines gemeinsamen Vokabulars bereitzustellen um die Daten so zu beschreiben, dass sie interoperabel werden. Die Umsetzung kann wie auch im vorliegenden Fall mit dem 3M umgesetzt werden. Der hohe Detailgrad des CRM bedingt allerdings einen hohen Grad an Komplexität. Vor dem Mapping ist daher eine zeitaufwendige und mit einer steilen Lernkurve versehene Einarbeitung notwendig (Hohmann & Fichtner, 2015). Ein besonderer Vorteil zeigt sich jedoch im starken Abstraktionsgrad des CRM mit dem man auch archäologische Daten sinnvoll beschreiben kann.

In der Vielzahl von semantischen Interpretationsmöglichkeiten liegt jedoch die besondere Schwierigkeit des Mappings, denn je nach Kenntnis über die Zusammenhänge von Daten sind unterschiedliche Ergebnisse zu erwarten. Was bereits für andere Projekten festgestellt wurde, lässt sich an dieser Stelle nicht nur für die Terra Preta Daten festhalten. Denn eine der wichtigsten Schlussfolgerungen ist, dass Datenbanken genau dokumentiert werden müssen, da dies den Interpretationsspielraum gerade beim Versuch zur Abbildung der Realität stark einschränkt (Carver, 2012). Dem Verlust von impliziter und expliziter Information von Daten aus archäologischen Projekten kann nur durch präventive Planung und genaue Beschreibung Einhalt geboten werden. Das CIDOC CRM kann im Vorfeld als Leitfaden für eine Datenstruktur dienen, die sich anschließend leicht semantisieren lässt.

Die vorliegende Arbeit platziert sich damit im Kontext von Linked Open Data als grundlegendster Baustein und soll dazu dienen weitere archäologische Mapping-Projekte auf Basis des CIDOC CRM anzustossen und so die Vision der Linked Open Data Cloud voranzutreiben.

Literaturverzeichnis

- Allemang, D., & Hendler, J. (2011). *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. Elsevier.
- Barrasa, J., Corcho, O., & Gómez-Pérez, A. (2003). Fund Finder: A case study of database-to-ontology mapping. *Semantic Integration Workshop (SI-2003)*, S. 9-14. URL: <http://oa.upm.es/5670/1/Workshop11.SemInt2003.pdf> [abgerufen: 31.10.17].
- Becker, H.-G., & Lins, M. (2013). Open Data und Linked Data in einem Informationssystem für die Archäologie. In P. Danowski (Hrsg.), *(Open) Linked Data in Bibliotheken*. Berlin [u.a.], S. 201-223.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284 (5), S. 34-43.
- Binding, C., May, K., & Tudhope, D. (2008). Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC CRM. In C. D. Christensen-Dalsgaard B. (Hrsg.), *Research and Advanced Technology for Digital Libraries. ECDL 2008. Lecture Notes in Computer Science (5173)*. Berlin, Heidelberg.
- Bizer, C., & Cyganiak, R. (2006). D2r server-publishing relational databases on the semantic web. *Poster at the 5th international semantic web conference (175)*, S. 1-2. URL: https://www.researchgate.net/profile/Christian_Bizer/publication/228837475_D2R_Server_-_Publishing_Relational_Databases_on_the_Semantic_Web/links/02e7e51d283213e97200000/D2R-Server-Publishing-Relational-Databases-on-the-Semantic-Web.pdf.
- Borgo, S., Massimiliano, C., Garbacz, P., & Vermaas, P. E. (2010). Formalizations of Functions within the DOLCE Ontology. *Proceedings of the 8th International Symposium on Tools and Methods of Competitive Engineering, TMCE 2010 (1)*, S. 113-126.
- Bruschke, J., & Wacker, M. (2014). Application of a Graph Database and Graphical User Interface for the CIDOC CRM. *Access and Understanding – Networking in the Digital Era. Session J1. The 2014 annual conference of CIDOC, the International Committee for Documentation of ICOM*. Dresden. URL: http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2014/J-1_Bruschke_Wacker_paper.pdf [abgerufen: 31.10.17].
- Bumanns, G. (2010). Mapping between Relational Databases and OWL. *Scientific Papers, University of Latvia, 2010 (756). Computer Science and Information Technologies*, S. 99-117. URL: http://dspace.lu.lv/dspace/bitstream/handle/7/2248/LUR-756_Datorika.pdf?sequence=1&isAllowed=y#page=99 [abgerufen: 31.10.17].
- Carver, G. (2012). ArcheoInf, the CIDOC-CRM and STELLAR: Workflow, Bottlenecks, and Where do we Go from Here? *CAA2012. Proceedings of the 40th Conference in Computer Applications and Quantitative Methods in Archaeology. Southampton, United Kingdom, 26-30 March 2012*, S. 498-508. URL: [http://www.ariadne-infrastructure.eu/ita/content/download/3763/21702/file/Geoff%20Carver-STELLAR\[2\].pdf](http://www.ariadne-infrastructure.eu/ita/content/download/3763/21702/file/Geoff%20Carver-STELLAR[2].pdf) [abgerufen: 31.10.17].
- Cioara, T., Moldovan, D., Antal, M., Valea, D., Pop, C., Anghel, I., & Salomie, I. (2015). Tools for Mapping Ontologies to Relational Databases: A Comparative Evaluation. *Intelligent Computer Communication and Processing (ICCP), 2015 IEEE International Conference*, S. 77-83.

- Crofts, N. (1999). Implementing the CIDOC CRM with a relational database. *MCN Spectra*, 24 (1), S. 1-6. URL: <https://www.researchgate.net/publication/260363113> [abgerufen: 31.10.17].
- Crofts, N., Doerr, M., & Gill, T. (2003). The CIDOC Conceptual Reference Model. A Standard for Communicating Cultural Contents. *Cultivate Interactive* (9), S. 1-14. URL: https://www.researchgate.net/publication/260363253_The_CIDOC_Conceptual_Reference_Model_A_Standard_for_Communicating_Cultural_Contents [abgerufen: 31.10.17].
- Crofts, N., Doerr, M., Gill, T., Stead, S., & Stiff, M. (2011). *Definition of the CIDOC Conceptual Reference Model. Version 5.0.4. ICOM/CIDOC CRM Special Interest Group*. URL: http://old.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf [abgerufen: 31.10.17].
- Decker, S., Prasenjit, M., & Melnik, S. (2000). Framework for the Semantic Web: An RDF Tutorial. *IEEE Internet Computing*, 4(6), S. 68-73. URL: <http://www.andrew.cmu.edu/user/mm6/95-733/PowerPoint/rdf tutorial.pdf> [abgerufen: 31.10.17].
- Diehr, F. (2013). *Ontologisch basiertes Datenmodell für die Beschreibung wissenschaftlicher Sammlungen. Ein Beitrag zur Identifikation semantischer Zielressourcen und zum Mapping existierender Quellattribute*. Berlin. URL: <http://edoc.hu-berlin.de/18452/14861> [Stand 31.10.17].
- Doerr, M. (2000). *Mapping of the Dublin Core metadata element set to the CIDOC CRM. ICS/FORTH, Technical Report 274*. URL: http://www.cidoc-crm.org/sites/default/files/dc_to_crm_mapping.pdf [abgerufen: 31.10.17].
- Doerr, M. (2003). The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24 (3), S. 75.
- Doerr, M. (2009). Ontologies for Cultural Heritage. In S. Staab, & R. Studer (Hrsg.), *Handbook on Ontologies*. Heidelberg, S. 463-486.
- Doerr, M., Felicetti, A., Hermon, S., Hiebel, G., Kritsotaki, A., Masur, A., . . . Tsiadaki, D. (2016). *Definition of the CRMarchaeo. An Extension of CIDOC CRM to support the archaeological excavation process*. URL: http://www.cidoc-crm.org/crmarchaeo/sites/default/files2/CRMarchaeo_v1.4.1.pdf [abgerufen: 31.10.17].
- Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G., & Theodoridou, M. (2014). *CRMsci: the Scientific Observation Model. An Extension of CIDOC-CRM to support scientific observation*. URL: <http://www.cidoc-crm.org/crmsci/sites/default/files2/CRMsci1.2.pdf> [abgerufen: 31.10.17].
- Doerr, M., Theodoridou, M., Aspöck, E., & Masur, A. (2015). Mapping Archaeological Databases to CIDOC CRM. *Proc. 43rd Computer Applications and Quantitative Methods in Archaeology Conference (CAA 2015 SIENA)*, S. 443-452.
- Dröge, E., Iwanowa, J., Trkulja, V., Hennicke, S., & Gradmann, S. (2013). Wege zur Integration von Ontologien am Beispiel einer Spezifizierung des Europeana Data Model. In H.-C. Hobohm (Hrsg.), *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Proceedings des 13. Internationalen Symposiums für Informationswissenschaften (ISI 2013)*. Potsdam, 19. bis 22. März 2013. Glückstadt, S. 273-284.
- Ermert, A., Gottschewski, J., Hagedorn-Saupe, M., Hansen, H.-J., Heuchert, R., Saro, C., . . . Stein, R. (2004). *Das CIDOC Conceptual Reference Model: Eine Hilfe für den Datenaustausch?* URL: http://old.cidoc-crm.org/docs/cidoc_paper_german.pdf [abgerufen: 31.10.17].
- FORTH ICS. (2007). *Datamodels and Standards*, URL: https://www.ics.forth.gr/isl/index_main.php?l=e&c=229 [abgerufen: 31.10.17].

- Ghawi, R., & Cullot, N. (2007). Database-to-ontology mapping generation for semantic interoperability. *Third International Workshop on Database Interoperability (InterDB 2007) (91)*.
- Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino (Hrsg.), *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*. Amsterdam, S. 3-15.
- Hanke, P., Mandl, T., & Womser-Hacker, C. (2002). Ein „Virtuelles Bibliotheksregal“ für die Informationswissenschaft als Anwendungsfall semantischer Heterogenität. In R. Hammwöhner, C. Wolf, & C. Womser-Hacker (Hrsg.), *Information und Mobilität, Optimierung und Vermeidung von Mobilität durch Information. Proceedings des 8. Internationalen Symposiums für Informationswissenschaft (ISI 2002), Regensburg, 8. - 11. Oktober 2002*. Konstanz, S. 289-302.
- Hiebel, G., Doerr, M., Eide, Ø., & Theodoridou, M. (2015). *CRMgeo: a Spatiotemporal Model. An Extension of CIDOC to link the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement*. URL: http://www.cidoc-crm.org/crmgeo/sites/default/files2/CRMgeo1_2.pdf [Stand 31.10.17].
- Hiebel, G., Doerr, M., Hanke, K., & Masur, A. (2014). How to put archaeological geometric data into context? Representing mining history research with CIDOC CRM and extensions. *International Journal of Heritage in the Digital Era*, 3(3), S. 557-577.
- Hilbert, P. P. (1968). *Archäologische Untersuchungen am mittleren Amazonas. Beiträge zur Vorgeschichte des südamerikanischen Tieflandes. Marburger Studien zur Völkerkunde* (Bd. 1). Berlin.
- Hohmann, G. (2008). Die Anwendung des CIDOC CRM für die semantische Wissensrepräsentation in den Kulturwissenschaften. In P. Ohly, & J. Sielgerschmidt (Hrsg.), *Wissensspeicher in digitalen Räumen. Nachhaltigkeit, Verfügbarkeit, semantische Interoperabilität. Proceedings der 11. Tagung der Deutschen Sektion der Internationalen Gesellschaft für Wissensorganisation Konstanz 20.-22. Februar 2008*. Würzburg, S. 210-222.
- Hohmann, G. (2011). Die Anwendung von Ontologien zur Wissensrepräsentation und -kommunikation im Bereich des kulturellen Erbes. In S. Schomburg, C. Leggewie, H. Lobin, & C. Puschmann (Hrsg.), *Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland. 20./21. September 2010, Köln, 2. ergänzte Fassung*. Köln, S. 33-41.
- Hohmann, G., & Fichtner, M. (2015). Chancen und Herausforderungen in der praktischen Anwendung von Ontologien für das Kulturerbe. In C. Y. Robertson-Von Trotha, & R. H. Schneider (Hrsg.), *Digitales Kulturerbe : Bewahrung und Zugänglichkeit in der wissenschaftlichen Praxis; Kulturelle Überlieferung - digital ; Band 2*. Karlsruhe/Baden, S. 115-128.
- Kokkelink, S., & Schwänzl, R. (2002). *Expressing Qualified Dublin Core in RDF / XML*. URL: <http://dublincore.org/documents/dcq-rdf-xml/> [abgerufen: 31.10.17].
- Kondylakis, H., Doerr, M., & Plexousakis, D. (2006). Mapping language for information integration. *ICS/FORTH Technical Report 385*, S. URL: ftp://139.91.151.170/tech-reports/2006/2006.TR385_Mapping_Language_Information_Integration.pdf [abgerufen: 31.10.17].
- Lagoze, C., & Hunter, J. (2006). The ABC ontology and model. *Journal of Digital Information*, Article No. 77, 2001-11-06, 2(2).
- Le Boeuf, P. (2012). A strange model named FRBRoo. *Cataloging & Classification Quarterly*, 50 (5-7), S. 422-438.

- Marketakis, Y., Minadakis, N., Kondylakis, H., Konsolaki, K., Samaritakis, G., Theodoridou, M., . . . Doerr, M. (2017). X3ML mapping framework for information integration in cultural heritage and beyond. *International Journal on Digital Libraries*, 18 (4), S. 301-319.
- Meghini, C., Scopingo, R., Richards, J., Wright, H., Geser, G., Cuy, S., . . . Vlachidis, A. (2017). ARIADNE: A research Infrastructure for Archaeology. *Journal on Computing and Cultural Heritage*, 10 (3). doi:10.1145/3064527
- Menzel, C. (2003). Reference Ontologies — Application Ontologies. "Either/Or or Both And?". *KI Workshop on Reference Ontologies and Application Ontologies (16)*, S. 1-10. URL: http://ceur-ws.org/Vol-94/ki03rao_menzel.pdf [30.10.2017].
- Minadakis, N., Marketakis, Y., Kondylakis, H., Flouris, G., Theodoridou, M., de Jong, G., & Doerr, M. (2015). X3ML Framework: An Effective Suite for Supporting Data Mappings. *Proceedings of the Workshop: Extending, Mapping and Focusing the CRM, TPDL 2015. Poznan, POL*.
- Neches, R., Fikes Richard, F. T., Gruber, T., Patil, R., Senator, T., & Swartout, W. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine*, 12 (3), S. 36-56.
- Neves, E. (2008). Ecology, Ceramic Chronology and Distribution, Long-term History, and Political Change. In H. Silvermann (Hrsg.), *Handbook of South American archaeology*. New York, S. 359-379.
- Nilsson, M., Powell, A., Johnston, P., & Naeve, A. (2008). *Expressing Dublin Core Metadata Using the Resource Description Framework (RDF)*. Dublin Core Metadata Initiative. DCMII Recommendation, 14. URL: <http://dublincore.org/documents/dc-rdf> [abgerufen: 31.10.2017].
- Nussbaumer, P., & Haslhofer, B. (2007). *Putting the CIDOC CRM into Practice. Experiences and Challenges. TR-200*. University of Vienna. URL: <http://eprints.cs.univie.ac.at/404/1/covered.pdf> [abgerufen: 30.10.2017].
- Oberle, D., Guarino, N., & Staab, S. (2009). What is an ontology? In S. Staab, & R. Studer (Hrsg.), *Handbook on Ontologies. 2nd edition*. Berlin/Heidelberg, S. 1-20.
- Oldman, D., & CRM Labs. (2014). *The CIDOC Conceptual Reference Model (CIDOC-CRM): Primer*. http://83.212.168.219/CIDOC-CRM/sites/default/files/CRMPrimer_v1.1.pdf [abgerufen: 01.01.2018].
- Oldman, D., Theodoridou, M., & Samaritakis, G. (2016). *Using Mapping Memory Manager (3M) with CIDOC CRM. Draft Version 5b*. URL: <http://139.91.183.3/3M/Manuals/en/manual.pdf> [abgerufen: 31.10.17].
- Olensky, M. (2010). Semantic interoperability in Europeana. An examination of CIDOC CRM in digital cultural heritage documentation. *Bulletin of IEEE Technical Committee on Digital Libraries. Volume 6 (2)*. URL: <http://www.ieee-tcdl.org/Bulletin/v6n2/Olensky/olensky.html> [abgerufen: 31.10.17].
- OpenInfRA. (2015). *OpenInfRA - Open Information System for Research in Archaeology. Grobkonzept für ein webbasiertes Informationssystem zur Dokumentation archäologischer Forschungsprojekte. V2.4. Arbeitsstand vom 16.04.2015*. URL: https://www-docs.tu-cottbus.de/openinfra/public/Grobkonzept_OpenInfRA.pdf [abgerufen: 28.10.2017].
- Ore, C. E., Doerr, M., Le Bœuf, P., & Stead, S. (2015). *Definition of the CIDOC Conceptual Reference Model. The version 6.0 of the reference document*. URL: http://www.cidoc-crm.org/sites/default/files2/cidoc_crm_version_6.0.pdf [abgerufen: 31.10.17].
- Pohl, A., & Danowski, P. (2013). Linked Open Data in der Bibliothekswelt. In P. Danowski (Hrsg.), *(Open) Linked Data in Bibliotheken*. Berlin [u.a.], S. 1-45.

- PostGIS Development Group. (2005). PostGIS 2.4.4dev Manual.
- PostgreSQL Global Development Group. (2017). PostgreSQL 10.1 Documentation.
- Radiokohlenstoff Datierung von Keramik. ((o.D.)). URL: <https://www.radiocarbon.com/deutsch/ams-datierung-keramik.htm> [abgerufen: 31.10.2017].
- Sahoo, S., Halb, W., Hellmann, Sebastian, Kingsley, I., Thibodeau, T., . . . Ezzat, A. (2009). A Survey of Current Approaches for Mapping of Relational Databases to RDF. *W3C RDB2RDF Incubator Group Report (1)*, S. 113-130.
- Scheuermann, L. (2006). Ontologien in den historischen Wissenschaften. *Historische Sozialforschung*, 31 (3) , S. 308-316.
- Schwer, J. (2016). Universität Augsburg. *unpubl.*
- Severiens, T., Demidova, E., Kucheria, S., Plümer, J., Polzer, H., Schwänzl, R., & Wenneker, S. (2006). *Einführung in RDF. RDF in wissenschaftlichen Bibliotheken*. URL: <https://www.iwi-iuk.uni-osnabrueck.de/seminarNotes/2/rdf-reader-v2.pdf> [abgerufen: 31.10.17].
- Spanos, D. E., Periklis, S., & Mitrou, N. (2012). Bringing relational databases into the semantic web: A survey. *Semantic Web* 3(2), S. 169-209.
- Spiridonidou, T. (2003). Dublin Core Metadata. *Grundlegende Aspekte des Semantic Web, Institut für Informatik, Humboldt Universität zu Berlin*. Berlin. URL: <https://www.researchgate.net/publication/260356866> [abgerufen: 31.10.17].
- Thiemeyer, H. (2008). Böden als Archive. Methoden aus der Bodenkunde bereichern die Archäologie. *Frankfurter Forschung* (3), S. 82-85.
- Trachsel, M. (2008). *Ur- und Frühgeschichte. Quellen, Methoden, Ziele*. Zürich.
- van Ruymbeke, M., Hallot, P., & Billen, R. (2017). Enhancing CIDOC-CRM and compatible models with the concept of multiple interpretation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (4), S. 287-294. URL: https://orbi.uliege.be/bitstream/2268/214320/1/Annals_CIPA_2017_Enhancing_VanRuymbek e.pdf [abgerufen: 31.10.17].
- von Groote-Bidlingmaier, C. (2016). *Developing a Data Driven Approach for the Analysis*. Augsburg: unpubl. Diss.
- von Groote-Bidlingmaier, C., Hilbert, K., Schwer, J., & Timpf, S. (2015). Interactive WebGIS for Archaeological Settlement Pattern Analysis—A Requirement Analysis. *27th International Cartographic Conference (ICC)*. Aug. 2015. Rio de Janeiro.
- von Groote-Bidlingmaier, C., Timpf, S., & Hilbert, K. (2014). Erstellung von potenziellen Einflussgebieten auf der Basis von archäologischen Ausgrabungsorten— am Beispiel prähistorischer Kulturen im Amazonasegebiet. In J. Strobl, J. Blaschke, G. Griesebner, & B. Zagel, *Angewandte Geoinformatik 2014 - Beiträge zum 26. AGIT-Symposium Salzburg*. S. 77-86.

Verzeichnis Internetlinks

Alle zuletzt abgerufen: 31.10.17

WEB1

<http://terrapreta.geo.uni-augsburg.de/>

WEB2

<http://portal.iphan.gov.br/sgpa/?consulta=cnsa>

WEB3

<https://www.postgresql.org/>

WEB4

<http://postgis.net/>

WEB5

<http://cidoc-crm.gnm.de/wiki/Hauptseite>

WEB6

<http://139.91.183.3/3M/Login>

WEB7

<http://www.geonames.org/>

WEB8

<https://www.w3.org/RDF/Validator/>

WEB9

<http://portal.iphan.gov.br/pagina/detalhes/872>

WEB10

<http://portal.iphan.gov.br/pagina/detalhes/236>

Anhang

A. Testdatensatz einer normalisierten XML-Datei aus der TP-Datenbank

```
<root>
  <findings>
    <row>
      <finding_id>1023</finding_id>
      <site_id>1</site_id>
      <finding_depth_cm>90</finding_depth_cm>
      <tradition_id>1</tradition_id>
      <culture_id>1</culture_id>
      <archaeologist_id>1</archaeologist_id>
      <bib_id>1</bib_id>
      <c14_year>380 +/- 60 BP</c14_year>
      <c14_info>(Si 4777)</c14_info>
    </row>
    <row>
      <finding_id>1024</finding_id>
      <site_id>2</site_id>
      <finding_depth_cm>90</finding_depth_cm>
      <tradition_id>2</tradition_id>
      <culture_id>2</culture_id>
      <archaeologist_id>2</archaeologist_id>
      <bib_id>2</bib_id>
      <c14_year>1655 +/- 65 BP</c14_year>
      <c14_info>(Si 4778)</c14_info>
    </row>
  </findings>
  <sites>
    <row>
      <site_id>1</site_id>
      <site_code>AM-AL-01</site_code>
      <site_name>Santa Luzia</site_name>
      <site_length_m>400</site_length_m>
      <site_width_m>200</site_width_m>
      <site_lat>-2.754039</site_lat>
      <site_lon>-67.583092</site_lon>
      <geom>0101000020E6100000CAA31B6151E550C0B344679
        9450806C0</geom>
    </row>
    <row>
      <site_id>2</site_id>
      <site_code>AM-AL-02</site_code>
      <site_name>São Joaquim</site_name>
      <site_length_m>250</site_length_m>
      <site_width_m>250</site_width_m>
      <site_lat>-2.751925</site_lat>
      <site_lon>-67.551203</site_lon>
      <geom>0101000020E6100000419DF2E846E350C0C05B20
        41F10306C0</geom>
    </row>
  </sites>
</root>
```

```

<traditions>
  <row>
    <tradition_id>1</tradition_id>
    <tradition_name>EingefassteSchraffur</tradition_name>
    <tradition_from>2500</tradition_from>
    <tradition_to>1500</tradition_to>
  </row>
  <row>
    <tradition_id>2</tradition_id>
    <tradition_name>Früh-Keramisch</tradition_name>
    <tradition_from>7000</tradition_from>
    <tradition_to>3000</tradition_to>
  </row>
</traditions>
<cultures>
  <row>
    <culture_id>1</culture_id>
    <culture_name>Konduri</culture_name>
    <culture_from>800</culture_from>
    <culture_to>400</culture_to>
  </row>
  <row>
    <culture_id>2</culture_id>
    <culture_name>Guarita</culture_name>
    <culture_from>1000</culture_from>
    <culture_to>300</culture_to>
  </row>
</cultures>
<bib>
  <row>
    <bib_id>1</bib_id>
    <bib_name>Hilbert. 1968</bib_name>
  </row>
  <row>
    <bib_id>2</bib_id>
    <bib_name>Simões et al. 1983; 1987</bib_name>
  </row>
</bib>
<archaeologists>
  <row>
    <archaeologist_id>1</archaeologist_id>
    <archaeologist_name>Steere</archaeologist_name>
    <archaeologist_from>1871</archaeologist_from>
    <archaeologist_to>1871</archaeologist_to>
  </row>
  <row>
    <archaeologist_id>2</archaeologist_id>
    <archaeologist_name>Vanzolini</archaeologist_name>
    <archaeologist_from>1974</archaeologist_from>
    <archaeologist_to>1974</archaeologist_to>
  </row>
</archaeologists>
</root>

```

B. Auszug der im 3M generierten RDF/XML Datei

```
<rdf:RDF
  xmlns:crmsci="http://www.ics.forth.gr/isl/CRMext/CRMsci/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:tp="http://www.tp.de/database/"
  xmlns:crm="http://www.cidoc-crm.org/cidoc-crm/"
  xmlns:crmgeo="http://www.ics.forth.gr/isl/CRMext/CRMgeo/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:crmarchaeo="http://www.ics.forth.gr/isl/CRMarchaeo/"
  <crm:E22_Man-Made_Object rdf:about="http://www.tp.de/database/Finding/1023">
    <crm:P53_has_former_or_current_location>
      <crm:E53_Place rdf:about="uid:AAF">
        <crm:P39i_was_measured_by>
          <crm:E16_Measurement rdf:about="uid:AAAG">
            <crmsci:O12_has_dimension>
              <crm:E54_Dimension rdf:about="uid:AAAH">
                <crm:P90_has_value>90</crm:P90_has_value>
                <crm:P91_has_unit>
                  <crm:E58_Measurement_Unit rdf:about="http://www.tp.de/database/Concept/Unit/cm"/>
                </crm:P91_has_unit>
                <crm:P2_has_type>
                  <crm:E55_Type rdf:about="http://www.tp.de/database/Concept/Dimension/depth"/>
                </crm:P2_has_type>
              </crm:E54_Dimension>
            </crmsci:O12_has_dimension>
          </crm:E16_Measurement>
        </crm:P39i_was_measured_by>
      </crm:E53_Place>
    </crm:P53_has_former_or_current_location>
    <crm:P3_has_note>(Si 4777)</crm:P3_has_note>
    <crm:P70i_is_documented_in>
      <crm:E31_Document rdf:about="http://www.tp.de/database/Bib/1">
        <crm:P48_has_preferred_identifier>
          <crm:E42_Identifier rdf:about="uid:AAAQ">
            <rdfs:label>Hilbert. 1968</rdfs:label>
          </crm:E42_Identifier>
        </crm:P48_has_preferred_identifier>
      </crm:E31_Document>
    </crm:P70i_is_documented_in>
    <crm:P108i_was_produced_by>
      <crm:E12_Production rdf:about="uid:AAAD">
        <crm:P4_has_time-span>
          <crmgeo:SP10_Declarative_Time-Span rdf:about="uid:AAAE">
            <crm:P82_at_some_time_within>380 +/- 60 BP</crm:P82_at_some_time_within>
          </crmgeo:SP10_Declarative_Time-Span>
        </crm:P4_has_time-span>
        <crm:P10_falls_within>
          <crm:E4_Period rdf:about="http://www.tp.de/database/Culture/1">
            <crm:P4_has_time-span>
              <crm:E52_Time-Span rdf:about="uid:AABV">
                <crm:P82b_end_of_the_end>400</crm:P82b_end_of_the_end>
                <crm:P82a_begin_of_the_begin>800</crm:P82a_begin_of_the_begin>
              </crm:E52_Time-Span>
            </crm:P4_has_time-span>
            <crm:P1_is_identified_by>
              <crm:E41_Appellation rdf:about="uid:AABU">
                <rdfs:label>Konduri</rdfs:label>
              </crm:E41_Appellation>
            </crm:P1_is_identified_by>
          </crm:E4_Period>
        </crm:P10_falls_within>
      </crm:E12_Production>
    </crm:P108i_was_produced_by>
    <crm:P10_falls_within>
      <crm:E4_Period rdf:about="http://www.tp.de/database/Tradition/1">
        <crm:P4_has_time-span>
          <crm:E52_Time-Span rdf:about="uid:AABR">
            <crm:P82b_end_of_the_end>1500</crm:P82b_end_of_the_end>
            <crm:P82a_begin_of_the_begin>2500</crm:P82a_begin_of_the_begin>
          </crm:E52_Time-Span>
        </crm:P4_has_time-span>
        <crm:P1_is_identified_by>
          <crm:E41_Appellation rdf:about="uid:AABQ">
            <rdfs:label>EingefassteSchraffur</rdfs:label>
          </crm:E41_Appellation>
        </crm:P1_is_identified_by>
      </crm:E4_Period>
    </crm:P10_falls_within>
  </crm:E12_Production>
</crm:P108i_was_produced_by>
  <crmsci:O19i_was_object_found_by>
    <crmsci:S19_Encounter_Event rdf:about="uid:AAAB">
      <crm:P9i_forms_part_of>
        <crm:E7_Activity rdf:about="http://www.tp.de/database/Activity/1/Steere">
          <crm:P11_had_participant>
            <crm:E39_Actor rdf:about="uid:AAAS">
              <crm:P11i_participated_in>
                <crmarchaeo:A9_Archaeological_Excavation rdf:about="uid:AAAU">
```

```

    <crm:P4_has_time-span>
      <crm:E52_Time-Span rdf:about="uuid:AAAW">
        <crm:P82b_end_of_the_end>1871</crm:P82b_end_of_the_end>
      </crm:E52_Time-Span>
    </crm:P4_has_time-span>
    <crm:P4_has_time-span>
      <crm:E52_Time-Span rdf:about="uuid:AAAV">
        <crm:P82a_begin_of_the_begin>1871</crm:P82a_begin_of_the_begin>
      </crm:E52_Time-Span>
    </crm:P4_has_time-span>
  </crmarchaeo:A9_Archaeological_Excavation>
</crm:P11i_participated_in>
<crm:P1_is_identified_by>
  <crm:E41_Appellation rdf:about="uuid:AAAT">
    <rdfs:label>Steere</rdfs:label>
  </crm:E41_Appellation>
</crm:P1_is_identified_by>
</crm:E39_Actor>
</crm:P11_had_participant>
<crm:P14_carried_out_by rdf:resource="uuid:AAAS"/>
</crm:E7_Activity>
</crm:P9i_forms_part_of>
<crm:P117_occurs_during>
  <crmarchaeo:A9_Archaeological_Excavation rdf:about="uuid:AAAC">
    <crm:P8_took_place_on_or_within>
      <crm:E27_Site rdf:about="http://www.tp.de/database/Site/1">
        <crmgeo:Q11i_is_approximated_by>
          <crmgeo:SP6_Declarative_Place rdf:about="uuid:AABI">
            <crm:P2_has_type>
              <crm:E55_Type rdf:about="http://www.tp.de/database/Concept/hexEWKT"/>
            </crm:P2_has_type>
            <rdfs:label>0101000020E610000CAA31B6151E550C0B3446799450806C0</rdfs:label>
          </crmgeo:SP6_Declarative_Place>
        </crmgeo:Q11i_is_approximated_by>
      </crm:E27_Site>
    </crm:P8_took_place_on_or_within>
  </crmarchaeo:A9_Archaeological_Excavation>
</crm:P117_occurs_during>
  <crm:E47_Spatial_Coordinates rdf:about="uuid:AABH">
    <crm:P2_has_type>
      <crm:E55_Type rdf:about="http://www.tp.de/database/Concept/WGS84_Longitude"/>
    </crm:P2_has_type>
    <rdfs:label>-67.583092</rdfs:label>
  </crm:E47_Spatial_Coordinates>
</crm:P1_is_identified_by>
<crm:P1_is_identified_by>
  <crm:E47_Spatial_Coordinates rdf:about="uuid:AABG">
    <crm:P2_has_type>
      <crm:E55_Type rdf:about="http://www.tp.de/database/Concept/WGS84_Latitude"/>
    </crm:P2_has_type>
    <rdfs:label>-2.754039</rdfs:label>
  </crm:E47_Spatial_Coordinates>
</crm:P1_is_identified_by>
<crm:P43_has_dimension>
  <crm:E54_Dimension rdf:about="uuid:AABF">
    <crm:P90_has_value>200</crm:P90_has_value>
    <crm:P2_has_type>
      <crm:E55_Type rdf:about="http://www.tp.de/database/Concept/Dimension/width"/>
    </crm:P2_has_type>
    <crm:P91_has_unit>
      <crm:E58_Measurement_Unit rdf:about="http://www.tp.de/database/Concept/Unit/meter"/>
    </crm:P91_has_unit>
  </crm:E54_Dimension>
</crm:P43_has_dimension>
<crm:P43_has_dimension>
  <crm:E54_Dimension rdf:about="uuid:AABE">
    <crm:P90_has_value>400</crm:P90_has_value>
    <crm:P2_has_type>
      <crm:E55_Type rdf:about="http://www.tp.de/database/Concept/Dimension/length"/>
    </crm:P2_has_type>
    <crm:P91_has_unit rdf:resource="http://www.tp.de/database/Concept/Unit/meter"/>
  </crm:E54_Dimension>
</crm:P43_has_dimension>
<crm:P1_is_identified_by>
  <crm:E42_Identifier rdf:about="uuid:AABD">
    <rdfs:label>AM-AL-01</rdfs:label>
  </crm:E42_Identifier>
</crm:P1_is_identified_by>
<crm:P1_is_identified_by>
  <crm:E41_Appellation rdf:about="uuid:AABC">
    <rdfs:label>Santa Luzia</rdfs:label>
  </crm:E41_Appellation>
</crm:P1_is_identified_by>
</crm:E27_Site>
</crm:P8_took_place_on_or_within>
</crmarchaeo:A9_Archaeological_Excavation>
</crm:P117_occurs_during>
</crm:sci:S19_Encounter_Event>
</crm:sci:O19i_was_object_found_by>
<crm:P1_is_identified_by>
  <crm:E42_Identifier rdf:about="uuid:AAAA"/>
</crm:P1_is_identified_by>
</crm:E22_Man-Made_Object>

```

```

<crm:E22_Man-Made_Object rdf:about="http://www.tp.de/database/Finding/1024">
  <crm:P53_has_former_or_current_location>
    <crm:E53_Place rdf:about="uuid:AAAN">
      <crm:P39i_was_measured_by>
        <crm:E16_Measurement rdf:about="uuid:AAAO">
          <crmsci:O12_has_dimension>
            <crm:E54_Dimension rdf:about="uuid:AAAP">
              <crm:P90_has_value>90</crm:P90_has_value>
              <crm:P91_has_unit rdf:resource="http://www.tp.de/database/Concept/Unit/cm"/>
              <crm:P2_has_type rdf:resource="http://www.tp.de/database/Concept/Dimension/depth"/>
            </crm:E54_Dimension>
          </crmsci:O12_has_dimension>
        </crm:E16_Measurement>
      </crm:P39i_was_measured_by>
    </crm:E53_Place>
  </crm:P53_has_former_or_current_location>
  <crm:P3_has_note>(Si 4778)</crm:P3_has_note>
  <crm:P70i_is_documented_in>
    <crm:E31_Document rdf:about="http://www.tp.de/database/Bib/2">
      <crm:P48_has_preferred_identifier>
        <crm:E42_Identifier rdf:about="uuid:AAAR">
          <rdfs:label>Simões et al. 1983; 1987</rdfs:label>
        </crm:E42_Identifier>
      </crm:P48_has_preferred_identifier>
    </crm:E31_Document>
  </crm:P70i_is_documented_in>
  <crm:P108i_was_produced_by>
    <crm:E12_Production rdf:about="uuid:AAAL">
      <crm:P4_has_time-span>
        <crmgeo:SP10_Declarative_Time-Span rdf:about="uuid:AAAM">
          <crm:P82_at_some_time_within>1655 +/- 65 BP</crm:P82_at_some_time_within>
        </crmgeo:SP10_Declarative_Time-Span>
      </crm:P4_has_time-span>
      <crm:P10_falls_within>
        <crm:E4_Period rdf:about="http://www.tp.de/database/Culture/2">
          <crm:P4_has_time-span>
            <crm:E52_Time-Span rdf:about="uuid:AABX">
              <crm:P82b_end_of_the_end>300</crm:P82b_end_of_the_end>
              <crm:P82a_begin_of_the_begin>1000</crm:P82a_begin_of_the_begin>
            </crm:E52_Time-Span>
          </crm:P4_has_time-span>
          <crm:P1_is_identified_by>
            <crm:E41_Appellation rdf:about="uuid:AABW">
              <rdfs:label>Guarita</rdfs:label>
            </crm:E41_Appellation>
          </crm:P1_is_identified_by>
        </crm:E4_Period>
      </crm:P10_falls_within>
    </crm:E12_Production>
  </crm:P108i_was_produced_by>
  <crm:P10_falls_within>
    <crm:E4_Period rdf:about="http://www.tp.de/database/Tradition/2">
      <crm:P4_has_time-span>
        <crm:E52_Time-Span rdf:about="uuid:AABT">
          <crm:P82b_end_of_the_end>3000</crm:P82b_end_of_the_end>
          <crm:P82a_begin_of_the_begin>7000</crm:P82a_begin_of_the_begin>
        </crm:E52_Time-Span>
      </crm:P4_has_time-span>
      <crm:P1_is_identified_by>
        <crm:E41_Appellation rdf:about="uuid:AABS">
          <rdfs:label>Früh-Keramisch</rdfs:label>
        </crm:E41_Appellation>
      </crm:P1_is_identified_by>
    </crm:E4_Period>
  </crm:P10_falls_within>
  </crm:E12_Production>
  <crm:P108i_was_produced_by>
  <crmsci:O19i_was_object_found_by>
    <crmsci:S19_Encounter_Event rdf:about="uuid:AAAJ">
      <crm:P9i_forms_part_of>
        <crm:E7_Activity rdf:about="http://www.tp.de/database/Activity/2/Vanzolini">
          <crm:P11_had_participant>
            <crm:E39_Actor rdf:about="uuid:AAAX">
              <crm:P11i_participated_in>
                <crmarchaeo:A9_Archaeological_Excavation rdf:about="uuid:AAAZ">
                  <crm:P4_has_time-span>
                    <crm:E52_Time-Span rdf:about="uuid:AABB">
                      <crm:P82b_end_of_the_end>1974</crm:P82b_end_of_the_end>
                    </crm:E52_Time-Span>
                  </crm:P4_has_time-span>
                  <crm:P4_has_time-span>
                    <crm:E52_Time-Span rdf:about="uuid:AABA">
                      <crm:P82a_begin_of_the_begin>1974</crm:P82a_begin_of_the_begin>
                    </crm:E52_Time-Span>
                  </crm:P4_has_time-span>
                </crmarchaeo:A9_Archaeological_Excavation>
              </crm:P11i_participated_in>
            </crm:E39_Actor>
          </crm:P11_had_participant>
        </crm:E7_Activity>
      </crm:P9i_forms_part_of>
    </crmsci:S19_Encounter_Event>
  </crm:P108i_was_produced_by>
  <crm:P1_is_identified_by>
    <crm:E41_Appellation rdf:about="uuid:AAAY">
      <rdfs:label>Vanzolini</rdfs:label>
    </crm:E41_Appellation>
  </crm:P1_is_identified_by>

```

```

    </crm:E39_Actor>
  </crm:P11_had_participant>
  <crm:P14_carried_out_by rdf:resource="uuid:AAAX"/>
  </crm:E7_Activity>
</crm:P9i_forms_part_of>
<crm:P117_occurs_during>
  <crmarchaeo:A9_Archaeological_Excavation rdf:about="uuid:AAAK">
    <crm:P8_took_place_on_or_within>
      <crm:E27_Site rdf:about="http://www.tp.de/database/Site/2">
        <crmgeo:Q11i_is_approximated_by>
          <crmgeo:SP6_Declarative_Place rdf:about="uuid:AABP">
            <crm:P2_has_type rdf:resource="http://www.tp.de/database/Concept/hexEWKT"/>
            <rdfs:label>0101000020E6100000419DF2E846E350C0C05B2041F10306C0</rdfs:label>
          </crmgeo:SP6_Declarative_Place>
        </crmgeo:Q11i_is_approximated_by>
        <crm:P1_is_identified_by>
          <crm:E47_Spatial_Coordinates rdf:about="uuid:AABO">
            <crm:P2_has_type rdf:resource="http://www.tp.de/database/Concept/WGS84_Longitude"/>
            <rdfs:label>-67.551203</rdfs:label>
          </crm:E47_Spatial_Coordinates>
        </crm:P1_is_identified_by>
        <crm:P1_is_identified_by>
          <crm:E47_Spatial_Coordinates rdf:about="uuid:AABN">
            <crm:P2_has_type rdf:resource="http://www.tp.de/database/Concept/WGS84_Latitude"/>
            <rdfs:label>-2.751925</rdfs:label>
          </crm:E47_Spatial_Coordinates>
        </crm:P1_is_identified_by>
        <crm:P43_has_dimension>
          <crm:E54_Dimension rdf:about="uuid:AABM">
            <crm:P90_has_value>250</crm:P90_has_value>
            <crm:P2_has_type rdf:resource="http://www.tp.de/database/Concept/Dimension/width"/>
            <crm:P91_has_unit rdf:resource="http://www.tp.de/database/Concept/Unit/meter"/>
          </crm:E54_Dimension>
        </crm:P43_has_dimension>
        <crm:P43_has_dimension>
          <crm:E54_Dimension rdf:about="uuid:AABL">
            <crm:P90_has_value>250</crm:P90_has_value>
            <crm:P2_has_type rdf:resource="http://www.tp.de/database/Concept/Dimension/length"/>
            <crm:P91_has_unit rdf:resource="http://www.tp.de/database/Concept/Unit/meter"/>
          </crm:E54_Dimension>
        </crm:P43_has_dimension>
        <crm:P1_is_identified_by>
          <crm:E42_Identifier rdf:about="uuid:AABK">
            <rdfs:label>AM-AL-02</rdfs:label>
          </crm:E42_Identifier>
        </crm:P1_is_identified_by>
        <crm:P1_is_identified_by>
          <crm:E41_Appellation rdf:about="uuid:AABJ">
            <rdfs:label>São Joaquim</rdfs:label>
          </crm:E41_Appellation>
        </crm:P1_is_identified_by>
      </crm:E27_Site>
    </crm:P8_took_place_on_or_within>
  </crmarchaeo:A9_Archaeological_Excavation>
</crm:P117_occurs_during>
</crm:sci:S19_Encounter_Event>
</crm:sci:O19i_was_object_found_by>
<crm:P1_is_identified_by>
  <crm:E42_Identifier rdf:about="uuid:AAAI"/>
</crm:P1_is_identified_by>
</crm:E22_Man-Made_Object>
</rdf:RDF>

```

C. Auszug der URI Generation Policy File

```
<generator_policy>
  <generator name="FindingURI" prefix="tp">
    <pattern>Finding/{term}</pattern>
  </generator>
  <generator name="TraditionURI" prefix="tp">
    <pattern>Tradition/{term}</pattern>
  </generator>
  <generator name="CultureURI" prefix="tp">
    <pattern>Culture/{term}</pattern>
  </generator>
  <generator name="ArchaeologistURI" prefix="tp">
    <pattern>Archaeologist/{term}</pattern>
  </generator>
  <generator name="SiteURI" prefix="tp">
    <pattern>Site/{term}</pattern>
  </generator>
  <generator name="BibURI" prefix="tp">
    <pattern>Bib/{term}</pattern>
  </generator>
  <generator name="DimensionURI" prefix="tp">
    <pattern>Dimension/{term}</pattern>
  </generator>
  <generator name="ConceptURI" prefix="tp">
    <pattern>Concept/{term}</pattern>
  </generator>
  <generator name="ConceptURI_2step" prefix="tp">
    <pattern>Concept/{topterm}/{term}</pattern>
  </generator>
  <generator name="OneLevelCustomURI" prefix="tp">
    <pattern>{level1}/{term}</pattern>
  </generator>
  <generator name="TwoLevelCustomURI" prefix="tp">
    <pattern>{level1}/{level2}/{term}</pattern>
  </generator>
  <generator name="OneLevelCustomURIwCount" prefix="tp">
    <pattern>{level1}/{count}/{term}</pattern>
  </generator>
  <generator name="TwoLevelCustomURIwCount" prefix="tp">
    <pattern>{level1}/{level2}/{count}/{term}</pattern>
  </generator>
  <generator name="SimpleLabel">
    <pattern>{label}</pattern>
  </generator>
</generator_policy>
```


E. Gesamter RDF-Graph des Artefaktes mit der ID 1023

