

Historisierungs- und Aktualisierungskonzepte für Forschungslandkarte

Bachelorarbeit

zur Erlangung des Grades eines Bachelor of Science
im Studiengang Informationsmanagement

vorgelegt von

Juan Camilo Orrego

207110353

Betreuer: Prof. Dr. Maria A. Wimmer, Institut für Wirtschafts- und Verwaltungsinformatik,
Forschungsgruppe Verwaltungsinformatik / E-Government

Erstgutachter: Prof. Dr. Maria A. Wimmer
Zweitgutachter: Ansgar Mondorf

Koblenz, im Juni 2012

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien der Forschungsgruppe für Qualifikationsarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden Ja Nein

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. Ja Nein

Koblenz, den

Unterschrift

Zusammenfassung

Der Forschungsbereich E-Government gewinnt in Deutschland und Europa immer mehr an Bedeutung. Die Anzahl der Akteure wächst ständig, sodass es für Außenstehende und Interessierte schwierig ist, den Überblick zu behalten. Es ist diese enorme Informationsflut, die eine Forschungslandkarte für den E-Government-Bereich für sinnvoll erscheinen lässt. Auf dieser Internetplattform soll ein Überblick über alle relevanten Akteure, die im Bereich E-Government Forschung betreiben, gegeben werden. Um einen vollständigen Überblick geben zu können, müssen auch Informationen angezeigt werden, die in der Vergangenheit gültig waren. Voraussetzung hierfür ist die Historisierung relevanter Inhalte. Hierbei wird bei Änderung eines Objekts oder Attributs das Objekt oder Attribut nicht überschrieben, sondern eine neue Version in der Datenbank angelegt. Auf diese Weise sind die Benutzer in der Lage, die Forschungslandkarte zeitbezogen zu navigieren. Erfahrungen haben gezeigt, dass die Historisierung von Webinhalten bereits bei der konzeptionellen Phase und beim Design einer Online-Plattform eingeplant und modelliert werden soll, um so den temporalen Anforderungen zu genügen. Diese Bachelorarbeit beschäftigt sich mit der Herausforderung der Historisierung der Inhalte der Forschungslandkarte und legt ein Konzept vor, das die temporale Datenhaltung bei der Konzeptionierung der Plattform berücksichtigt.

Abstract

The E-Government research area has gained in importance in Europe and specially Germany in the last few years, causing the number of researchers, institutes and publications to increase rapidly. This makes it difficult for outsiders to get an overview of the relevant actors in the E-Government field. This issue can be addressed by implementing a research map for the E-Government field, where all relevant actors and objects and their information are shown on the map according to their location. In order to give a complete overview, information which was valid at a certain time in the past needs to be available on the research map. This can be only achieved if the contents of the research map are historicized. This means that a new version of an object needs to be created and saved in the database, if changes occur to the object. Older versions need to be retained on the database, so that the user is able to navigate the Webseite based on temporal information. Past experience has shown that the temporal aspects of historicization should be managed and planned during the conceptual phase of the Webseite rather than during implementation. This Bachelor thesis proposes a concept for the E-Government research map which includes the modeling of relevant temporal dimensions needed to historicize the contents of the research map.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Ausgangslage, Problemstellung und Zielsetzung	1
1.2	Struktur der Arbeit	3
2	Forschungslandkarte E-Government	6
2.1	Bedarf an Forschungslandkarte und Ziele des Projekts	6
2.2	Eigenschaften und Entitäten der Forschungslandkarte	7
3	Content Management Systeme	9
3.1	Content	9
3.2	Content Management	11
3.2.1	Content Life Cycle	12
3.2.2	Content Life Cycle im Kontext der Forschungslandkarte E-Government	14
3.3	Content Management Systeme	17
3.3.1	Web Content Management Systeme	18
3.3.2	Aufbau und Funktionen von Content Management Systemen	18
4	Temporale Datenhaltung	25
4.1	Notwendigkeit von Historisierung	25
4.2	Temporale Datenbanken	26
4.3	Temporale Webdokumente	28
4.4	Temporale Erweiterung der Datenmodellierung und Modell R _{ETTE}	29
4.5	Historisierung von Metadaten	32
4.6	Zusammenfassung	33
5	Metadaten	34
5.1	Begriffserklärungen	34
5.2	Arten von Metadaten	34
5.3	Funktionen von Metadaten	35
5.4	Dublin Core	36
5.4.1	Strukturierung von Metadaten	36
5.4.2	Dublin Core metadata	37
5.5	Tagging	38
5.5.1	Taggingbegriff und Folksonomien	38
5.5.2	Tagging innerhalb der Forschungslandkarte	41
6	Semantic Web Technologies	43
6.1	Einführung Semantisches Web	43
6.2	Ontologien	43
6.2.1	Wissensrepräsentation	43
6.2.2	Ontologien	45
6.2.3	Zusammenfassung	51
7	Content Management Systeme als Grundlage für Forschungslandkarte	52
7.1	Anforderungen an Content Management Systeme im Kontext der Forschungslandkarte E-Government	52
7.2	Eignung von Drupal für Forschungslandkarte	54
8	Herausforderungen im Kontext der E-Government-Forschungslandkarte: Historisierung und Aktualisierung	59
8.1	Historisierung der Inhalte	59
8.2	Aktualisierung der Inhalte	61
8.3	Abgrenzung und Gewichtung der Historisierung und Aktualisierung	63
9	Struktur und Lösungskonzept für die Historisierung der Forschungslandkarte	64
9.1	Datenmodellierung	64
9.1.1	Datenmodellierung der Forschungslandkarte	65
9.1.2	Temporale Datenmodellierung der Forschungslandkarte	66
9.2	Metadateninformationen in der Forschungslandkarte	70
10	Anforderungsspezifikation für die Historisierung der Forschungslandkarte	75

11	Handlungsempfehlungen für die Historisierung der Forschungslandkarte	80
12	Fazit und Ausblick	83
13	Literaturverzeichnis	86
	Anhang 1: Instanzerstellung bei Forschungslandkarte	89
	Anhang 2: DCMI Recommended Qualifiers	90

Abbildungsverzeichnis

Abbildung 1-1: Mock-up Forschungslandkarte E-Government (Projektpraktikum Forschungslandkarte E-Government 2010)	1
Abbildung 2-1: Klassendiagramm Forschungslandkarte (Entnommen aus Projekthandbuch des Projektpraktikums Forschungslandkarte E-Government 2010)	8
Abbildung 3-1: Begriffsabgrenzung Content (in Anlehnung an Baumann, 2003)	10
Abbildung 3-2: Die Bestandteile von Dokumenten (Gersdorf, 2003)	11
Abbildung 3-3: Definition und Festlegung der Attribute innerhalb des Content Management Lifecycle (Maass & Stahl, 2003)	12
Abbildung 3-4: Content Life Cycle in Forschungslandkarte.....	14
Abbildung 3-5: Hauptkomponenten eines CMS (Bodendorf, 2005).....	19
Abbildung 3-6: Datenorganisation im Content Repository (Bodendorf, 2005)	22
Abbildung 3-7: Publishing System (Bodendorf, 2005).....	23
Abbildung 4-1: Graphische Darstellung der unterschiedlichen Attributarten (Kaiser, 1999)	31
Abbildung 4-2: Graphische Darstellung der Beziehungstypen im Modell RETTE (Kaiser, 1999)	32
Abbildung 5-1: The Fifteen Elements of „Simple Dublin Core“ (Sugimoto, Baker, & Weibel, 2002)	37
Abbildung 5-2: TagCloud E-Government Forschungslandkarte (Entnommen aus Projekthandbuch des Projektpraktikums Forschungslandkarte E-Government 2010).....	40
Abbildung 6-1: Classification of the Methods of Knowledge Representation with regard to Expressiveness and Coverage of the Knowledge Domain. Source: Adapted from Peters & Weller (2008, 101, Fig. 1). (Peters, 2009).....	45
Abbildung 6-2: Musician ontology visualized as a semantic network (Gašević, Djurić, & Devedžić, 2009)	47
Abbildung 6-3: The Musician ontology represented in OWL [excerpt] (Gašević, Djurić, & Devedžić, 2009)	48
Abbildung 7-1: Überblick über den Drupal-Core (Van Dyk, 2009)	55
Abbildung 7-2: Suchformular (Graf, Drupal 6: Websites entwickeln und verwalten mit dem Open Source-CMS, 2008).....	58
Abbildung 9-1: ER-Modell der Forschungslandkarte E-Government.....	66
Abbildung 9-2: RETTE Modell der Forschungslandkarte E-Government	67

Tabellenverzeichnis

Tabelle 1: Anforderungen an das Web Content Management System im Kontext der Forschungslandkarte E-Government	54
Tabelle 2: Anforderungen an Historisierung der Forschungslandkarte.....	79

1 Einleitung

1.1 Ausgangslage, Problemstellung und Zielsetzung

2010 wurde das Projekt Forschungslandkarte E-Government vom Forschungsbereich Verwaltungsinformatik der Universität Koblenz-Landau ins Leben gerufen. Ziel des Projekts war es, einen Überblick über den Stand der Forschung im Bereich E-Government in Deutschland zu schaffen. Dies soll anhand einer Forschungslandkarte erreicht werden, auf der alle relevanten Akteure bzw. Objekte und Informationen über diese für den Forschungsbereich E-Government ortsbezogen angezeigt werden sollen. Die Akteure bzw. Objekte sind Forschungsinstitutionen, Personen, Konferenzen, Projekte und Publikationen. Abbildung 1-1 zeigt ein vorläufiges Mock-up der Forschungslandkarte.

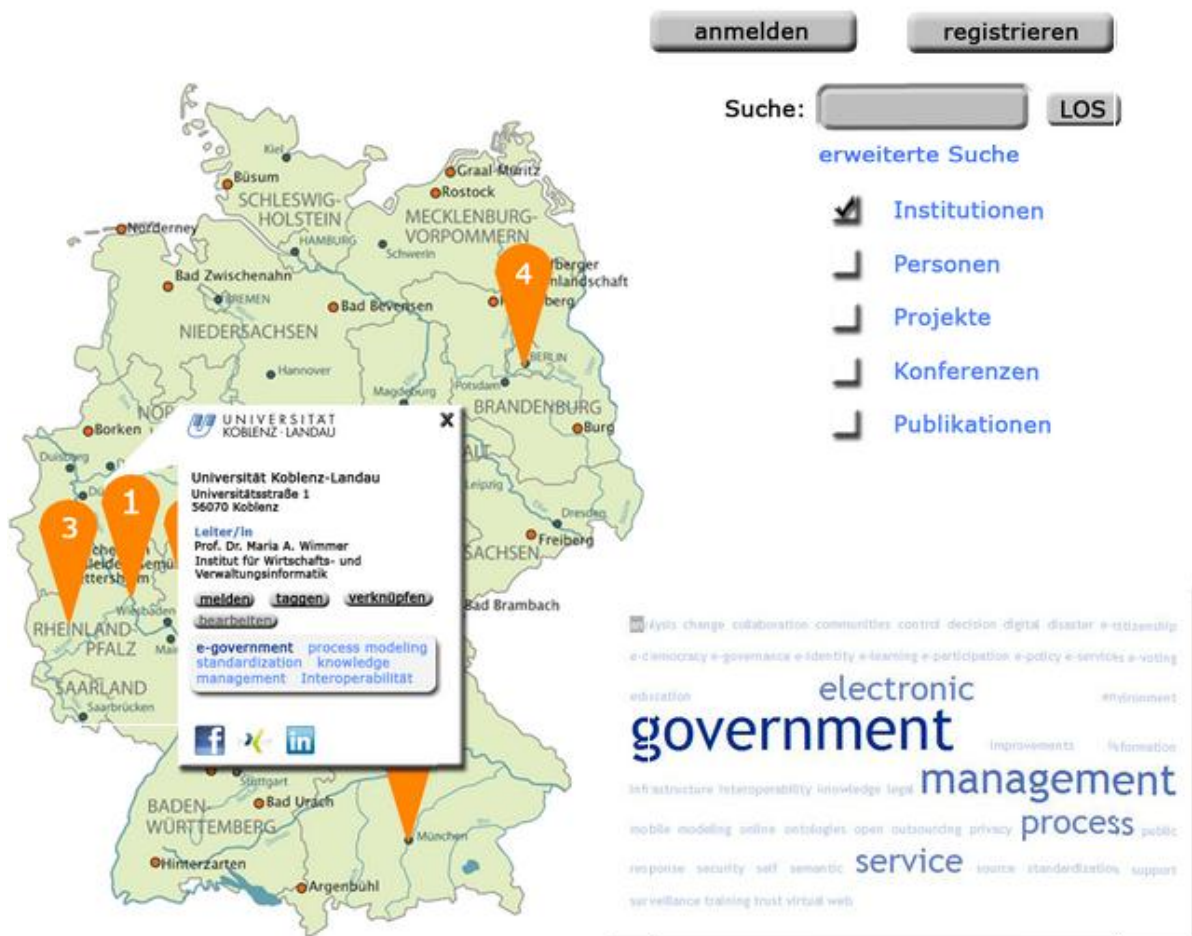


Abbildung 1-1: Mock-up Forschungslandkarte E-Government (Projektpraktikum Forschungslandkarte E-Government 2010)

Einleitung

Die Forschungslandkarte soll einen vollständigen Überblick der E-Government Forschung geben. Daher ist es notwendig, dass nicht nur aktuelle Informationen auf der Plattform bezogen werden können, sondern auch Informationen aus den letzten Jahren bereitgestellt werden. Genauer gesagt muss die Historie der einzelnen Instanzen auf der Forschungslandkarte nachvollziehbar sein. Auf diese Weise sind die Nutzer in der Lage zu erfahren, wie sich ein bestimmter Akteur oder Objekt mit der Zeit entwickelt hat.

Das Problem der Historisierung von Inhalten im digitalen Zeitalter ist kein neues. Bereits in den 70er Jahren begann man Methoden zu entwickeln, um digitale Objekte in Datenbanken zu historisieren (vgl. Etzion, Gal, & Segev, 1998). Mit Historisierung ist das Festhalten der Änderungen eines Objekts während dessen Lebensdauer gemeint. Objekte der realen Welt unterliegen einem ständigen Wandel, sodass sich ihre Eigenschaften ändern können. Ist das Objekt für die modellierte Welt innerhalb eines Systems relevant, müssen diese Änderungen erfasst und im System umgesetzt werden. Dies führt dazu, dass Objekte häufig aktualisiert oder sogar gelöscht werden müssen. Durch die Historisierung wird bei der Änderung eines Objekts dieses nicht bloß überschrieben, sondern eine neue Version des Objekts in der Datenbank angelegt. Über die Versionen wird die Historie eines Objekts abgebildet. Wird die Historie der Inhalte in der Datenbank dokumentiert, kann man sich anhand einer Filterfunktion den Zustand eines beliebigen Objekts zu einem beliebigen Zeitpunkt in der Vergangenheit anzeigen lassen. Um diese Anforderung zu erfüllen, müssen die Inhalte der Forschungslandkarte historisiert werden.

Mit der Herausforderung der Historisierung beschäftigt sich die vorliegende Bachelorarbeit. Es soll ein Konzept entwickelt werden, das die Historisierung der Inhalte der Forschungslandkarte ermöglicht. Dabei müssen relevante zeitliche Aspekte und Dimensionen bereits bei der konzeptionellen Phase der Forschungslandkarte mitberücksichtigt werden. Somit lautet die Problemstellung der vorliegenden Bachelorarbeit, die bisherigen Konzeptionierungsergebnisse der Forschungslandkarte so zu erweitern, dass die Historisierung der Inhalte möglichst problemlos bei der Implementierung der Online-Plattform umgesetzt werden kann.

Eine zweite Herausforderung der Forschungslandkarte ist die automatische Aktualisierung der Inhalte. Hierbei handelt es sich um die automatisierte Datenpflege. Ziel ist es, den Verwaltungsaufwand der Projektverantwortlichen zu minimieren, indem

Einleitung

die Inhalte der Forschungslandkarte dezentral und möglichst automatisiert aktuell gehalten werden. Dabei gibt es unterschiedliche Ebenen der Aktualisierung. Die erste Ebene umfasst die Änderungen der Instanzen in der realen Welt, die erfasst und im System umgesetzt werden müssen. Die zweite Aktualisierungsebene sind weitere Aktualisierungen, die aus den Änderungen aus der realen Welt (1. Ebene) resultieren. Ein Beispiel hierfür ist die Änderung der Forschungsfelder eines Instituts, falls die verantwortliche Person das Institut wechselt und somit keine Forschung mehr in diesen Bereichen betrieben wird. Darüber hinaus muss überlegt werden, ob die automatisierte Integration neuer Inhalte, anhand von Crawlern beispielsweise, sinnvoll und möglich ist. Die Herausforderung der Aktualisierung wird durch die Historisierung der Inhalte begleitet, da relevante Aktualisierungen historisiert und dokumentiert werden müssen. Daher ist es nicht möglich, die Aktualisierung der Inhalte völlig unbeachtet zu lassen. Allerdings kann die Datenpflege vorerst durch die Nutzer durchgeführt werden. Dies wird auch dem Ziel der Aufwandsminimierung der Projektverantwortlichen gerecht. Da die vollständige Betrachtung der Herausforderung der automatisierten Datenpflege durch das System den Rahmen dieser Bachelorarbeit sprengen würde, soll diese nur kurz angesprochen werden.

1.2 Struktur der Arbeit

Im Folgenden wird der Aufbau der Bachelorarbeit beschrieben. Die Arbeit ist in drei Teile aufgeteilt.

Im ersten Teil der Arbeit wird die Forschungslandkarte E-Government vorgestellt. Zuerst wird der Bedarf einer Forschungslandkarte für den Forschungsbereich E-Government dargestellt und die Ziele der Plattform beschrieben. Darüber hinaus wird der Aufbau der Forschungslandkarte näher betrachtet. Dabei werden die Entitäten, ihre Attribute, die Beziehungen zwischen Entitäten sowie weitere Eigenschaften der Forschungslandkarte beschrieben.

Im darauffolgenden Teil 2 der Arbeit wird anhand der bereits vorhandenen Literatur die theoretische Grundlage der Arbeit gelegt.

Die Forschungslandkarte soll mit Hilfe eines Content Management Systems realisiert werden. So werden in Kapitel 3 zuerst die wichtigsten Konzepte, worauf Content Management Systeme basieren, vorgestellt. Hierzu gehören Begriffe wie Content,

Einleitung

Content Management und Content Life Cycle. In einem weiteren Schritt wird der Aufbau von Content Management Systemen und Web Content Management Systemen beschrieben und deren Funktionen aufgezeigt.

Kapitel 4 legt die Grundlage für die Historisierung. Hier wird das Prinzip der temporalen Datenhaltung vorgestellt. In diesem Kapitel wird gezeigt, warum Inhalte historisiert werden und welche Methoden und technische Hilfsmittel dazu genutzt werden. Es wird sowohl auf die temporale Datenhaltung in Datenbanken als auch auf die temporale Datenhaltung im Rahmen von Webseiten Bezug genommen. Des Weiteren wird das Modell R_{ETTE} vorgestellt, mit dem relevante temporale Aspekte bereits bei der Konzeptionierungsphase von Plattformen berücksichtigt werden können.

Kapitel 5 behandelt das Thema Metadaten. Zuerst wird der Begriff Metadaten erklärt und gezeigt, welche Arten von Metadaten existieren und wozu diese genutzt werden. Anhand dessen wird die Taggingfunktion der Forschungslandkarte vorgestellt, mit der Nutzer Themenfelder an Instanzen der Forschungslandkarte anhängen können.

Kapitel 6 behandelt die Grundlagen des Semantic Web. Zuerst werden die Prinzipien des Semantic Web erklärt und die Grundlagen sowie Bausteine beschrieben. Darüber hinaus werden neuartige Webtechnologien vorgestellt, die bei der Implementierung der Forschungslandkarte E-Government genutzt werden können. Insbesondere werden geeignete Technologien wie zum Beispiel Wissensrepräsentationsmethoden und Ontologien untersucht und deren Bedeutung für die Historisierung der Forschungslandkarte aufgezeigt.

Im dritten und letzten Teil der Arbeit wird ein Konzept entwickelt, das die Historisierung der Inhalte der Forschungslandkarte ermöglichen soll.

In Kapitel 7 werden Content Management Systeme im Kontext einer Forschungslandkarte betrachtet. In einem ersten Schritt werden die Anforderungen einer Forschungslandkarte an ein Content Management System erörtert. Basierend darauf wird das CMS Drupal vorgestellt und gezeigt, warum Drupal für dieses Projekt besonders geeignet ist.

Einleitung

Kapitel 8 widmet sich den Herausforderungen der Forschungslandkarte für die vorliegende Ausarbeitung. Hier werden beide Herausforderungen (Historisierung und Aktualisierung) beschrieben und voneinander abgegrenzt.

In Kapitel 9 wird die Grundlage für die Historisierung der Forschungslandkarte gelegt. Hierbei wird ein Konzept entwickelt, das die Historisierung der Inhalte von Beginn an berücksichtigt. Zuerst wird die Datenmodellierung der Forschungslandkarte durchgeführt. Dabei wird das Entity-Relationship-Modell der Forschungslandkarte um temporale Aspekte erweitert. Durch diese Erweiterung wird festgelegt, welche Objekte der Forschungslandkarte historisiert werden müssen und welche nicht. Außerdem wird festgelegt, welche Metainformationen zusammen mit den Inhalten veröffentlicht werden sollen, sodass Nutzer wichtige Informationen über temporale Aspekte der Inhalte erhalten.

Kapitel 10 widmet sich der Anforderungsspezifikation für die Historisierung. In Tabellenform werden die Anforderungen erfasst, die das System erfüllen muss, um die Inhalte der Forschungslandkarte zu historisieren.

Basierend darauf werden in Kapitel 11 konkrete Handlungsempfehlungen für die Historisierung der Inhalte der Forschungslandkarte gegeben. Es werden Vorschläge ausgesprochen über das zu nutzende Content Management System, die Historisierung auf Datenbankebene und Metadatenstandards.

Schließlich werden im Fazit die Ergebnisse der vorliegenden Bachelorarbeit zusammengefasst und die Ziele auf Erreichung bzw. nicht Erreichung überprüft. Darüber hinaus werden weitere notwendige Schritte und der weitere Forschungsbedarf in Bezug auf das Projekt Forschungslandkarte E-Government aufgezeigt.

Teil 1: Forschungslandkarte E-Government

2 Forschungslandkarte E-Government

Das Projekt „Forschungslandkarte E-Government“ ist ein Projekt des Forschungsbereichs Verwaltungsinformatik im Institut für Wirtschafts- und Verwaltungsinformatik der Universität Koblenz-Landau. Unter der Organisation von Prof. Dr. Maria A. Wimmer wurde das Projekt im Rahmen eines Projektpraktikums im Sommersemester 2010 gestartet.

2.1 Bedarf an Forschungslandkarte und Ziele des Projekts

Aufgrund der enormen Informationsflut im Bereich des E-Governments ist es heutzutage beinahe unmöglich, sich einen Überblick über Institute, Personen etc., die in diesem Bereich tätig sind, zu verschaffen. Durch die Vielzahl an Akteuren gibt es unzählige Institutionen, Konferenzen, Projekte oder Publikationen, die für den Forschungsbereich relevant sind. Dadurch entsteht ein Informationsbedarf, welcher momentan aufgrund der fehlenden Informationsquellen nicht befriedigt werden kann.

Dieser Informationsbedarf soll zukünftig mit Hilfe einer Internetplattform in Form einer Forschungslandkarte gedeckt werden. Auf dieser webbasierten Forschungslandkarte soll ein Überblick über alle Institute, Forscher, Projekte, Konferenzen und Publikationen mit Forschungsschwerpunkt E-Government innerhalb Deutschlands gegeben werden. Dabei werden die einzelnen Instanzen ortsbezogen auf der Landkarte angezeigt.

Das System soll den Nutzern erlauben, sich zu registrieren. Sind diese angemeldet, so können sie ihre Informationen pflegen oder sich mit anderen Instanzen verknüpfen.

Als Zielvorgabe gilt die dynamische Aufnahme und Aktualisierung von Inhalten, um veraltete Informationen möglichst automatisch zu vermeiden und den Verwaltungsaufwand der Seite so gering wie möglich zu halten.

Dabei müssen relevante Informationen historisiert werden, um die zeitliche Entwicklung der Instanzen zu dokumentieren und um so einen vollständigen Überblick der E-Government Forschung in den letzten Jahren zu schaffen.

Ziel der vorliegenden Bachelorarbeit ist es, ein Konzept der Historisierung zu entwickeln, das den Nutzern anhand einer Filterfunktion ermöglicht, Informationen über Instanzen zu finden, die in der Vergangenheit gültig waren oder stattgefunden haben. Würde man beispielsweise die Forschungslandkarte nach Konferenzen und einem

Zeitintervall von 2010 bis 2011 filtern, sollen alle Konferenzen angezeigt werden, die zwischen 2010 und 2011 stattgefunden haben. Darüber hinaus ist es wichtig, relevante Attributs- oder Verknüpfungsänderungen zwischen Instanzen zu historisieren, um so beispielsweise nachvollziehen zu können, wo eine Person früher gearbeitet hat oder an welchen Projekte diese mitgearbeitet hat.

Am Ende der ersten Projektphase wurde eine erste Konzeptionierung bestehend aus Prozessmodellen, Szenarien, Use-Cases, Mock-Ups usw. erarbeitet, die als Anleitung für eine zukünftige Implementierung der Online-Plattform Forschungslandkarte E-Government dienen soll.

Die Herausforderung der Historisierung der Inhalte der Forschungslandkarte wurde während dieser ersten Projektphase nicht berücksichtigt.

2.2 Eigenschaften und Entitäten der Forschungslandkarte

In der ersten Konzeptionsphase des Projekts wurde das Content-Management-System Drupal als Tool für die Implementierung der Forschungslandkarte und der Veröffentlichung dessen Inhalte vorgeschlagen. So erwiesen sich nach einer ersten Untersuchung die Funktionen und Eigenschaften von Content-Management-Systemen und insbesondere von Drupal bei der Implementierung und Verwaltung einer solchen Online-Plattform für besonders nützlich und geeignet. Die Tauglichkeit von Drupal für das Projekt soll im Laufe dieser Arbeit erneut kurz angesprochen und belegt werden.

Eine wichtige Rolle bei der Aktualisierung und Historisierung und beim Projekt insgesamt spielen die Akteure. Bei der Konzeptionierung wurden die Entitäten festgelegt, die auf der Forschungslandkarte angezeigt werden sollen. Dazu gehören alle **Forschungsinstitute**, die im Bereich E-Government tätig sind. Diese werden mit Hilfe von Geokoordinaten ortsbezogen auf der Landkarte angezeigt. Wie bei allen anderen Entitäten auch werden weitere Informationen wie Kontakt, Link oder Forschungsfelder auf Wunsch angezeigt. Des Weiteren werden auch **Konferenzen** und **Projekte** auf der Karte ortsbezogen angezeigt. **Forscher oder Personen** werden am Ort des Instituts angezeigt, wo diese arbeiten, wohingegen **Publikationen** am Ort des Verlegers angezeigt werden.

Eine wichtige Eigenschaft, die bei der Implementierung des Systems berücksichtigt werden muss, ist die Möglichkeit, die verschiedenen Entitäten miteinander zu verknüpfen. Nur so können die Beziehungen zwischen den einzelnen Akteuren modelliert

werden. So können beispielsweise Personen in Instituten arbeiten, Konferenzen besuchen oder in Projekten mitarbeiten. Darüber hinaus existiert eine Verknüpfung zwischen Person und Publikation. Hat eine Person eine Publikation geschrieben oder an dieser mitgearbeitet, muss dieses modelliert werden können. Die Beziehungen zwischen den Entitäten ist ein Aspekt, den es bei der Aktualisierung und Historisierung der Entitäten besonders zu berücksichtigen gilt. So muss zum Beispiel die Frage geklärt werden, wie sich beispielsweise der Wechsel eines Professors zu einer anderen Universität auf die verknüpften Entitäten auswirkt. Auf der einen Seite ändert sich natürlich die Organisationszugehörigkeit der Person, auf der anderen Seite bleiben Publikationen der bisherigen Institution zugehörig.

Abbildung 2-1 gibt einen Überblick über die Entitäten des Systems, ihre Eigenschaften und die Beziehungen zueinander.

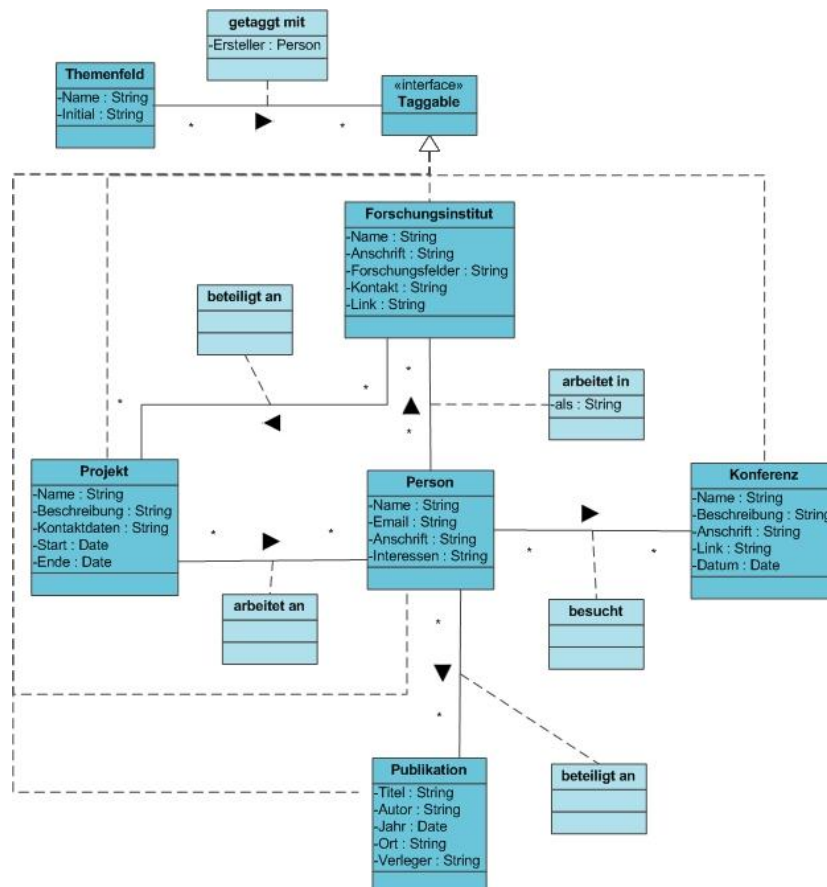


Abbildung 2-1: Klassendiagramm Forschungslandkarte (Entnommen aus Projekthandbuch des Projektpraktikums Forschungslandkarte E-Government 2010)

Nachdem die Forschungslandkarte vorgestellt wurde, soll im zweiten Teil der Arbeit die theoretische Grundlage anhand der Literatur gelegt werden, die die Historisierung der Inhalte ermöglicht.

Teil 2: Theoretische Grundlagen

3 Content Management Systeme

Wie bereits erwähnt, baut die E-Government Forschungslandkarte auf ein Content Management System. Anhand des CMS können dann die Inhalte und Informationen gespeichert, verwaltet und präsentiert werden.

Um eine verständliche Basis für die folgenden Kapitel dieser Arbeit zu schaffen, werden in diesem Kapitel Content Management Systeme im Allgemeinen vorgestellt und die wichtigsten Begriffe erklärt.

3.1 Content

Die wichtigste Funktion einer Forschungslandkarte ist die Bereitstellung von Informationen und Inhalten, in unserem Fall in Bezug auf den Forschungsbereich E-Government. Diese Inhalte werden in der Literatur als Content bezeichnet. Allerdings existieren viele verschiedene Definitionen für diesen Begriff. Eine für den Autor dieser Arbeit geeignete Definition bietet Kronz an:

„Content sind alle (potentiellen), elektronisch verfügbaren Inhalte einer Webseite, die für den Kunden, Partner oder Mitarbeiter einer Organisation von Interesse sind.“ (Kronz, 2003, S. 5)

Obwohl das oben genannte Zitat aus dem Bereich des E-Commerce stammt, kann man es trotzdem auch im Kontext der Forschungslandkarte anwenden. So kann man alle Inhalte als Content bezeichnen, die den Nutzern der Forschungslandkarte bereitgestellt werden sollen. Das sind konkret die Informationen über Institutionen, Personen, Konferenzen, Projekte und Publikationen.

Laut Kronz besitzt Content der Definition nach bestimmte prägende Eigenschaften. Content muss zu jeder Zeit elektronisch abrufbar sein. Inhalte, die nicht elektronisch verfügbar sind, werden als Wissen bezeichnet und deshalb als Content nicht berücksichtigt (Kronz, 2003, S. 5). Die Verfügbarkeit muss also durch die Webtechnologie gewährleistet sein. So müssen die Daten und Informationen oft aufbereitet werden, damit die Nutzer die Inhalte über das Web abrufen und auf diese zugreifen können. Darüber hinaus muss ein Interesse seitens der Nutzer an den Informationen bestehen, denn erst dann werden diese laut Kronz (2003, S. 5) als Content angesehen.

Im Rahmen dieser Bachelorarbeit soll der Begriff Content etwas tiefgründiger betrachtet werden als in der oberen Definition. So kann laut Nix Content nicht nur auf die redaktionellen Inhalte reduziert werden, sondern muss alle Arten von digitalen Informationen und Inhalten umfassen, wie zum Beispiel auch Metadaten, die die redaktionellen Objekte beschreiben und deren Struktur festlegen (2005). Abbildung 3-1 zeigt den Aufbau von Content.

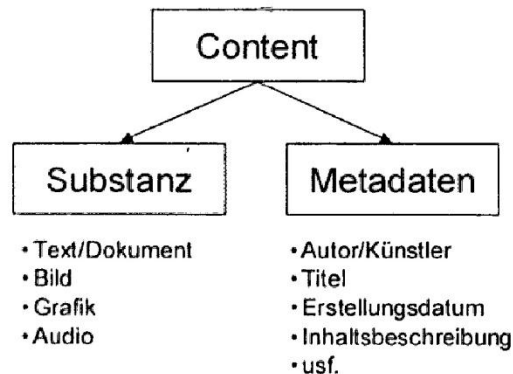


Abbildung 3-1: Begriffsabgrenzung Content (in Anlehnung an Baumann, 2003)

Alle diese einzelnen Informationselemente werden schließlich gebündelt und bilden auf diese Weise logische Einheiten, die sich zu beliebigen Medienprodukten kombinieren lassen (Bodendorf, 2005, S. 95). Um die Wiederverwendung von Content zu erleichtern und eine einfachere Handhabung zu gewährleisten, soll nach Bodendorf eine Trennung von Struktur, Inhalt und Layout erfolgen (Abbildung 3-2). Die Struktur beschreibt den Aufbau eines Dokuments, während der Inhalt die genaue Ausprägung des Dokuments bzw. des Informationselements festlegt. Schließlich beschreiben die Layoutinformationen, in welcher Weise der Content im Zielmedium dargestellt wird (Bodendorf, 2005, S. 95).

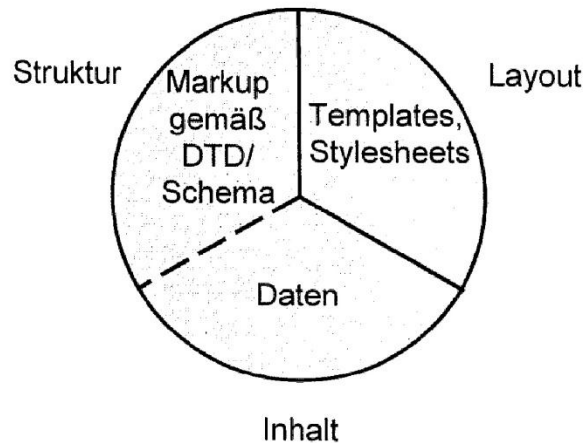


Abbildung 3-2: Die Bestandteile von Dokumenten (Gersdorf, 2003)

Zusammenfassend besteht der Content also aus einzelnen Elementen und agiert als eine Art Behälter für diese Elemente oder andere Behälter. Laut Spörrer speichert der Behälter die Inhalte nicht selbst, sondern referenziert nur auf diese. So können Inhalte mehrmals von verschiedenen Behältern genutzt und wiederverwendet werden (Spörrer, 2009, S. 6).

Vereinfachend soll in dieser Arbeit der deutschsprachige Begriff Inhalt als Synonym für Content genutzt werden.

3.2 Content Management

Wie bereits erwähnt stellen die Inhalte der Forschungslandkarte die wichtigste Ressource des Projekts dar. Deshalb und aufgrund der immer wachsenden Informationsmengen entsteht die Notwendigkeit, diese Inhalte zu managen und zu verwalten. Dies ist im Groben die Aufgabe des Content Managements. Dabei umfasst Content Management alle Aktivitäten und Prozesse, die bei der Strukturierung und Verwaltung von Content während des gesamten Lebenszyklus notwendig sind (Gersdorf, 2003). Daraus resultiert das Ziel des Content Managements, nämlich den Nutzern und Interessenten aktuellen Content zur Verfügung zu stellen. Einen ähnlichen Gedankengang offenbart Kronz. Sie definiert das Ziel erfolgreichen Content Managements als die Fähigkeit, die richtigen Inhalte den richtigen Nutzern zur richtigen Zeit bereitzustellen. Dabei handelt es sich bei den „richtigen“ Inhalten um Inhalte, die korrekt sind und den Qualitätsanforderungen des Herausgebers genügen. Diese Inhalte sollen dann den „richtigen“ Nutzern bereitgestellt werden. Diese sind diejenigen, die das Recht haben, auf die zur Verfügung gestellten Inhalte zuzugreifen (Kronz, 2003).

Bezogen auf die Forschungslandkarte sind das alle Personen, die auf die Plattform zugreifen, um Kontaktinformationen, Informationen über Publikationen und alle anderen Inhalte zu beziehen. Mit der „richtigen“ Zeit ist die Aktualität und zeitliche Relevanz des Contents gemeint. Das oben genannte Ziel, den Nutzern die richtigen Inhalte bereitzustellen, unterliegt einer Reihe besonderer Aktivitäten, ohne die es nicht möglich wäre, Inhalte strukturiert zur Verfügung zu stellen. So wird in der Literatur Content Management oft als eine Kette von Vorgängen und Handlungen definiert, die notwendig für die Verwaltung des Contents sind. So versteht man unter Content Management „die systematische und strukturierte Beschaffung, Erzeugung, Aufbereitung, Verwaltung, Präsentation, Verarbeitung, Publikation und Wiederverwendung von Inhalten.“ (Rothfuss & Ried, 2001, S. 52).

3.2.1 Content Life Cycle

Ein in der Literatur oft eingesetztes Referenzmodell für das Content Management ist der Content Life Cycle (Abbildung 3-3). Das Modell zeigt die Phasen, die die Inhalte während des gesamten Lebenszyklus durchlaufen und verdeutlicht dabei die Aufgaben des Content Managements.

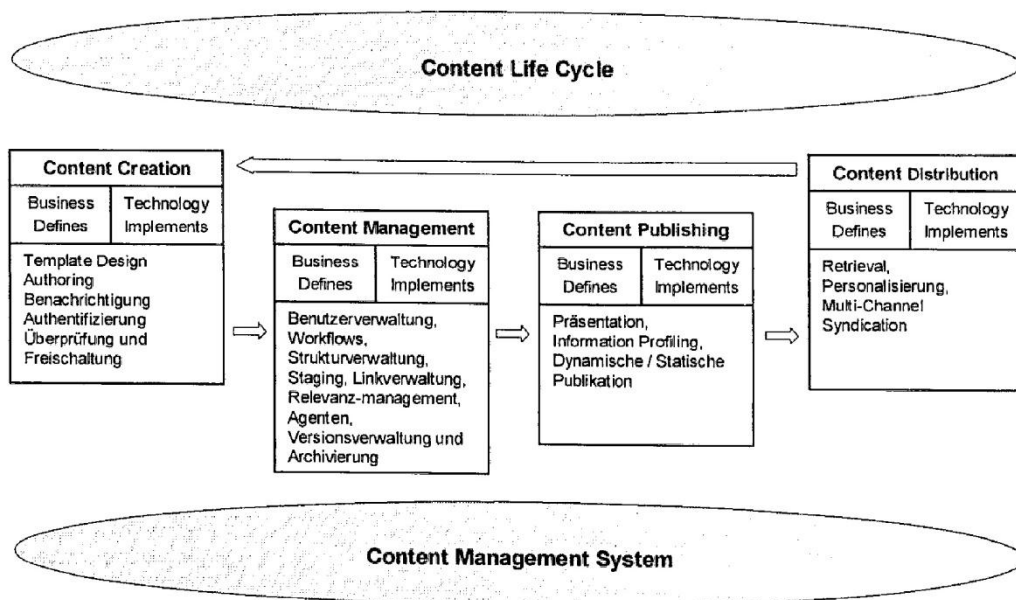


Abbildung 3-3: Definition und Festlegung der Attribute innerhalb des Content Management Lifecycle (Maass & Stahl, 2003)

Laut Maass & Stahl besteht der Lebenszyklus von Inhalten aus vier Phasen, welche in viele Unteraufgaben aufgeteilt werden können. Die vier Phasen sind **Content Creation**,

Content Management, Content Publishing und Content Distribution (Maass & Stahl, 2003). Im folgenden Abschnitt sollen nun die einzelnen Phasen näher erläutert und später deren Bedeutung im konkreten Fall der Forschungslandkarte E-Government gezeigt werden.

Content Creation:

In der ersten Phase gilt es, den Content zu generieren. Dabei müssen die Inhalte wie zum Beispiel Bilder, Texte oder Videos von den Autoren unabhängig vom Layout erstellt werden. Alternativ zu der Generierung vom eigenen Content gibt es die Möglichkeit, fremden Content zu beziehen und unverändert weiterzugeben (Bodendorf, 2005, S. 98). Neben der Contenterstellung muss außerdem noch das Layout erstellt werden. Dies ist die Aufgabe des sogenannten Content Designers.

Dieser muss im Rahmen der Content Creation das Layout der Inhalte in Form von Template-Schablonen oder Stylesheets bereitstellen. Weitere Aufgaben in der ersten Phase sind laut Maass & Stahl die Benachrichtigung und Authentifizierung der erstellten Inhalte (2003).

Content Management:

Während der Phase des Content Managements erfolgt die Organisation der Inhalte. Hierbei wird eine Struktur geschaffen, die die spätere Nutzung der Inhalte erleichtern soll. So wird in dieser Phase der Content durch Metainformationen angereichert. Dies ist sehr wichtig, um den Zugriff auf den Content zu gewährleisten und die effizientere Nutzung der Inhalte zu ermöglichen. Auf die Bedeutung der beschreibenden Metadaten wird zu einem späteren Zeitpunkt in dieser Arbeit eingegangen. Darüber hinaus erfolgt in dieser Phase die Versionierung des Contents, die Benutzerverwaltung und im Rahmen der Qualitätssicherung wird der Content freigegeben oder unter Umständen abgelehnt. Eine weitere Aufgabe des Content Managements ist die Archivierung der Inhalte.

Content Publishing:

In der dritten Phase wird die Publikation der Inhalte verwaltet und durchgeführt. Hierzu gehören „die dynamische und statische Präsentation und Publikation des Content“, also die Zusammenführung von Inhalt und Layout zu den notwendigen Aktivitäten (Maass & Stahl, 2003).

Content Distribution:

Die Aufgaben der vierten und letzten Phase sollen die Distribution der Inhalte unterstützen und ermöglichen. Hierzu zählen zum Beispiel die Gewährleistung und Bereitstellung von Retrievalmöglichkeiten und die Publikation über verschiedene Medien bzw. Kanäle. Auch die Personalisierung der Inhalte soll während dieser Phase angestrebt werden.

3.2.2 Content Life Cycle im Kontext der Forschungslandkarte E-Government

Das Modell des Content Life Cycle ist ein stark abstrahiertes Referenzmodell, das versucht, den Lebenszyklus von Content im Allgemeinen darzustellen. Um diese theoretische Abstraktion zu erreichen, werden hierbei laut Bodendorf vereinfachende Thesen angenommen, die in der Praxis aber oft in einer anderen Form vorkommen. So wird hier angenommen, dass Content immer den gleichen Lebenszyklus durchläuft (Bodendorf, 2005, S. 98). Dies ist allerdings nicht immer der Fall, denn jede Art von Inhalt hat einen eigenen Lebenszyklus, welcher aus verschiedenen Phasen besteht, die unterschiedlich umfangreich sind. Im folgenden werden die unterschiedlichen Phasen des Lebenszyklus von Inhalten und Informationen, die im Kontext der Forschungslandkarte E-Government gespeichert und veröffentlicht werden sollen, grob erläutert (Abbildung 3-4).

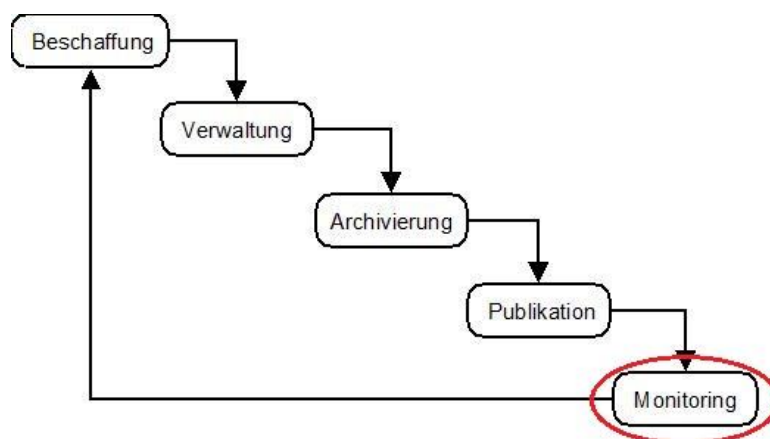


Abbildung 3-4: Content Life Cycle in Forschungslandkarte

In der ersten Phase erfolgt die Beschaffung des Contents. Im Falle der Forschungslandkarte wird kein eigener Content generiert, sondern die Inhalte werden ausschließlich aus fremden Quellen bezogen. Im Idealfall sollen die Informationen dynamisch oder von den Akteuren selbst in das System eingespeist werden. Dies hat das Ziel, den enormen und unzumutbaren Aufwand im Falle einer manuellen Suche

nach den Informationen und Inhalten zu reduzieren, und die Vollständigkeit und Aktualität der Inhalte zu gewährleisten. Eine komplett automatisierte Informationsbeschaffung ist vorerst allerdings nicht möglich, sodass die Informationen über Konferenzen, Projekte etc. durch die Nutzer selbst eingespeist werden müssen. Diese erste Phase des Lebenszyklus von Inhalten, die auf der Forschungslandkarte angezeigt werden sollen, ist Teil der **Content Creation** Phase aus dem vorgestellten Referenzmodell von Maass & Stahl (2003).

In der zweiten Phase werden die Inhalte und Informationen verwaltet. Hierzu gehören einige der Aufgaben aus dem Bereich des **Content Managements** aus dem Modell von Maass & Stahl (2003). Eine Aufgabe in dieser Phase ist es, die notwendige Struktur des Contents zu schaffen, um die Archivierung und die Veröffentlichung dieser zu ermöglichen. Bei der Verwaltung werden die Inhalte automatisch mit Metadaten angereichert. Diese kann man sich als Informationen über die Inhalte vorstellen, welche die Inhalte der Forschungslandkarte selbst beschreiben. So legen diese beispielsweise die technischen Eigenschaften (z.B. Dateiformat oder Dateigröße) oder strukturellen Eigenschaften (z.B. Autor, Erstellungsdatum...) fest und ermöglichen dabei beispielsweise die automatische Publikation von Dokumenten (Schumann, 2007). Die Anreicherung der Inhalte mit Metadaten erfolgt im Falle der Forschungslandkarte automatisch durch das Content Management System anhand der Daten, die die Systemnutzer bei der Entitätenerstellung eingegeben haben. Darüber hinaus werden Metadaten über die Tagging Funktion durch die Nutzer des Systems erzeugt. Die Frage, ob zusätzliche Metadaten, die für die Historisierung der Seite notwendig sind, manuell ergänzt werden müssen, soll zu einem späteren Zeitpunkt in dieser Bachelorarbeit untersucht werden.

Eine weitere Aufgabe während der Verwaltungsphase ist es, die Informationen mit dem nötigen Layout für die Veröffentlichung zu versehen. Dies erfolgt, wie auch die Erzeugung der Metadaten, automatisch durch das CMS. Nachdem die Nutzer die Informationen in der Maske eingegeben haben, werden die Entitäten samt Layout durch das System erzeugt.

Eine weitere Aktivität, die für die spätere Veröffentlichung der Inhalte relevant ist, ist die Verknüpfung der Publikationen mit den Personen, die diese geschrieben oder an ihr mitgearbeitet haben. Im Idealfall sollen sich alle Personen selbst mit den Publikationen verknüpfen, an denen sie beteiligt sind. Dies wird in der Realität aber nicht hundertprozentig möglich sein, denn anzunehmen, dass alle Personen, die an

Publikationen beteiligt waren auch im System registriert sind, ist utopisch. Selbst wenn die Autoren registriert sind, werden sich wahrscheinlich nur die wenigsten die Mühe machen, sich mit allen Publikationen zu verknüpfen. Aus Gründen der Vollständigkeit sollen aber alle Autoren einer Publikation in der Forschungslandkarte angezeigt oder zumindest genannt werden. So ist es die Aufgabe bei der Verwaltung, die nötigen Verknüpfungen zwischen Publikationen und den registrierten Nutzern des Systems zentral zu erstellen. Die fehlenden Angaben zu Autoren, welche nicht im System registriert sind, müssen dann anhand von Metadaten ergänzt werden.

Eine Aufgabe, die in der Literatur oft zum Content Management dazugehört, ist die Qualitätssicherung und Freigabe des Contents. Im Falle der Forschungslandkarte erfolgt die Prüfung der Inhalte nur im Falle der Erstellung eines neuen Instituts. Alle anderen Instanzen wie Projekte, Konferenzen oder Publikationen, werden ohne Freigabe direkt in der Datenbank gespeichert. Dies wurde während der Konzeptionierung der Forschungslandkarte festgelegt (Anhang 1: Instanzerstellung bei Forschungslandkarte). Somit fällt diese Aufgabe während der Verwaltung des Contents aus. Ausnahme ist hierbei die Institutserstellung. Wird eine neue Instanz der Entität Institut durch einen Nutzer erstellt, werden die Eingaben durch einen Prüfer auf Korrektheit geprüft. Ist alles korrekt, wird die neue Instanz in der Datenbank gespeichert. Dies hat das Ziel falsche bzw. redundante Eingaben zu vermeiden.

Im folgenden Schritt des Lebenszyklus erfolgt die Archivierung der Inhalte. Diese werden nun zusammen mit den Metainformationen in einer Datenbank abgespeichert. Alternativ kann man die Metadaten getrennt von den Inhalten abspeichern und diese mittels Referenz verknüpfen.

In der vierten Phase erfolgt die Publikation (**Content Publishing/Distribution**) der Inhalte. Die Inhalte sind dann jedem Nutzer, der auf die Forschungslandkarte zugreift, auf verschiedene Möglichkeiten frei zugänglich. So kann man beispielsweise auf der Forschungslandkarte auf die ortsbezogenen Entitäten zugreifen oder sich diese über die Suchfunktion auflisten lassen.

Nach der Publikation findet das Monitoring der Inhalte statt. Diese Phase ist von hoher Bedeutung, um die Aktualität der Inhalte innerhalb der Forschungslandkarte zu gewährleisten. Da es sich bei den Instanzen der Forschungslandkarte um echte

Personen, Institute etc. handelt, sind die Attribute der Instanzen einem ständigen Wandel unterzogen. So genügt es beispielsweise, dass ein Professor das Institut wechselt, um wasserfallartige Änderungen der Attribute der betroffenen Entitäten auszulösen. So muss nicht nur die Organisationszugehörigkeit des Professors geändert werden, sondern zum Beispiel auch die Forschungsfelder des Instituts, falls diese mit dem Professor abwandern. Um die Inhalte aktuell zu halten, müssen alle Änderungen der Attribute, die sich in der Realität abspielen, erfasst und die entsprechenden Änderungen in der Forschungslandkarte durchgeführt werden. Dies ist eine der Aufgaben beim Monitoring. Da es sich bei der Forschungslandkarte um eine Informationsplattform handelt, ist es sehr wichtig, dass die Informationen so weit wie möglich aktuell sind, denn veraltete Inhalte senken den Wert der Forschungslandkarte ungemein. So ist es unbedingt notwendig, dass das Monitoring der Seite in regelmäßigen Zeitabständen durchgeführt wird. Sind die Änderungen erstmal erfasst, müssen nun die entsprechenden Attributsänderungen durchgeführt werden. Dies soll anhand von semantischen Verknüpfungen anhand von Metadaten, Ontologien usw. automatisch erfolgen.

3.3 Content Management Systeme

Wie aus den letzten Kapitelabschnitten ersichtlich, ist die Aufgabe des Content Managements, die betroffenen unstrukturierten Informationen innerhalb der verschiedenen Phasen des Lebenszyklus zu verwalten (Gersdorf, 2003). So sind laut Gersdorf „Content Management Systeme Anwendungssysteme, welche die dafür notwendige informationstechnische Umsetzung für alle oder zumindest die wichtigsten Teilaufgaben realisieren“ (2003). Dabei sollen laut Hengartner die Inhalte vom System so gehandhabt werden, dass eine einfache und strukturierte Verwendung und Verwaltung ermöglicht wird (2003). Es handelt sich also bei Content Management Systemen laut Koop, Jäckel und Anja um IT-Systeme oder Tools zur Organisation, Verwaltung und Durchführung von Content Management (2001).

Der Unterschied zwischen Content Management und Content Management Systemen, wie in den obigen Definitionen bereits angedeutet, bringen Rothfuss & Ried deutlich zum Ausdruck: Es handelt sich beim Content Management um einen Prozess bzw. eine Kette von Aufgaben, wohingegen es sich bei Content Management Systemen um ein technisches Hilfsmittel handelt, das für die Umsetzung und Durchführung der Content Management Aufgaben genutzt wird (Rothfuss & Ried, 2001, S. 57). Der Einsatz von

Content Management Systemen hat laut Bodendorf letztendlich das Ziel, die Kosten der Content Erstellung und Verwaltung zu senken und gleichzeitig eine Steigerung der Qualität der Inhalte zu erzielen (2005, S. 100).

3.3.1 Web Content Management Systeme

Web Content Management Systeme (WCMS) sind eine spezielle Ausprägung von Content Management Systemen, die die Erstellung, Verwaltung und Publikation von Informationen in Webseiten ermöglicht (vgl. Zschau, 2003). Dabei steht laut Rothfuss & Ried die Optimierung der Verwaltung und Verwendung von Inhalten, welche über das Internet publiziert werden sollen, im Vordergrund (2001). Hierbei handelt es sich nicht nur um selbst erstellten Content, sondern auch um Content, der aus Drittsystemen und Informationsanbietern durch das Web Content Management System eingebunden wird (Stahl & Wolfgang, 2003, S. 269).

Aufgrund der Fülle an Informationen ist die Nutzung eines WCMS für umfangreiche Webseiten unentbehrlich. Die Anbindung des WCMS an eine Datenbank ermöglicht die Speicherung und Verwaltung von verwendeten digitalen Objekten in der Datenbank. Auf diese Weise können Bilder, Videos etc. leicht wiederverwendet werden. Dies ist heutzutage sehr wichtig, um den redaktionellen Aufwand für eine Webseite zu reduzieren. Daher werden laut Zschau immer mehr Web Content Management Systeme für die Erstellung und Verwaltung komplexer Webseiten verwendet (2003, S. 52).

Darüber hinaus gibt es weitere Vorteile, die die Nutzung von WCMS für komplexe Webseiten sinnvoll macht. So sind diese eine Hilfe, wenn die Inhalte einer Webseite von Mitarbeitern gepflegt werden sollen, die über geringe technische Fähigkeiten verfügen. Denn aufgrund der einfachen und intuitiven Nutzeroberfläche von WCMS, benötigt man keine besondere Programmierkenntnisse für die Verwaltung von Content. Ein weiterer Vorteil ist die Verwaltung der Inhalte einer Webseite von verschiedenen Standorten aus (Zschau, 2003, S. 53).

3.3.2 Aufbau und Funktionen von Content Management Systemen

Aufgrund der verschiedenen Bedürfnisse von Webseiten, müssen Content Management Systeme unternehmensspezifisch zugeschnitten und personalisiert werden. Deshalb existieren heutzutage unzählige CMS-Lösungen, die verschiedene Anforderungen und Funktionen abdecken sollen. Trotzdem ist es möglich, Content Management Systeme in

ihrer einfachsten Form und anhand der Grundmerkmale zu charakterisieren. Abbildung 3-5 zeigt die technische Architektur von Content Management Systemen nach (Bodendorf, 2005). Diese besteht aus den drei Komponenten Editorial System, Content Repository und Publishing System.

Das Content Repository bildet die Basis des Systems, wo die Inhalte gespeichert werden, nachdem diese im Editorial System erzeugt bzw. bearbeitet wurden. Über das Publishing System werden die Inhalte dann an die Rezipienten weitergeleitet. Die Komponenten Publishing System und Editorial System entsprechen dem Front- und Backend Prinzip von Graf (2006). Das Frontend ist die Webseite an sich, wie sie die Endnutzer zu sehen bekommen, während im Backend die Inhalte von den Redakteuren bzw. Administratoren generiert und verwaltet werden. Nun werden die einzelnen Komponenten eines Content Management Systems und deren Funktionen näher erleutert.

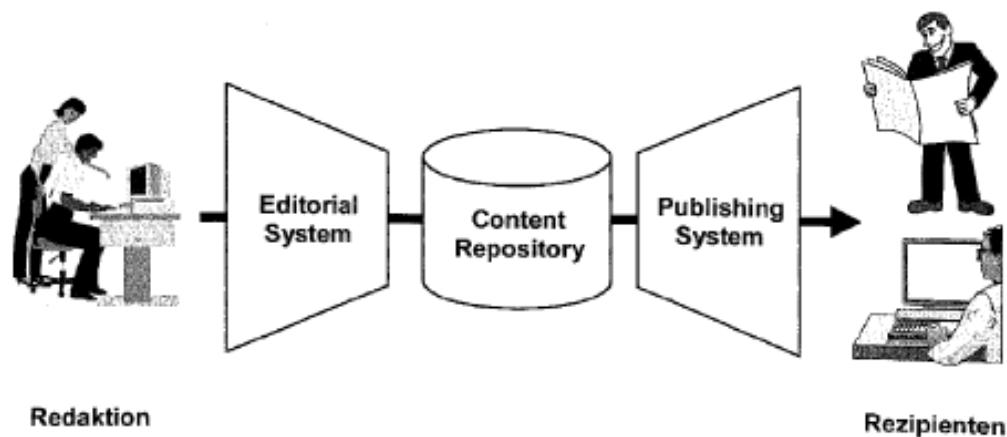


Abbildung 3-5: Hauptkomponenten eines CMS (Bodendorf, 2005)

Editorial System:

Im Editorial System werden Inhalte aller Art erzeugt und bearbeitet. Hierzu stehen laut Bodendorf Editoren zur Verfügung, die über eine Schnittstelle direkt in das Content Management System eingebunden sind, was zu einer Reduzierung des Aufwands bei der Einpflege von Content in das System führt (2005, S. 103). Bei der Bearbeitung von Content soll laut Rothfuss & Ried das Content Management System fähig sein, alle Änderungen des Inhalts zu protokollieren, um Fehler im Nachhinein entdecken zu können. Außerdem soll das CMS über eine Rollbackfunktion verfügen, die es möglich

macht, eine alte Version des geänderten Inhalts wiederherzustellen. Dies ist allerdings nur möglich, wenn das CMS durch eine ausgeprägte Datensicherungsfunktion die Inhalte regelmäßig speichert (Rothfuss & Ried, 2001, S. 75).

Darüber hinaus muss eine Verbindung zwischen dem Editorial System und dem Content Repository bestehen. Nur so kann durch Referenzen zwischen verschiedenen Inhalten, Content zu einer Einheit gebündelt und verwaltet werden (vgl. Bodendorf, 2005). Hierzu ist auch eine Sprache notwendig, die es ermöglicht, Beziehungen zwischen Inhaltsobjekten zu modellieren (Rothfuss & Ried, 2001).

Im Editorial System werden außerdem noch Inhalte mit dem gewünschten Layout versehen. Das Layout wird bei Content Management Systemen mittels Vorlagen (Templates) realisiert. Dies hat laut Zschau den Vorteil, dass Informationen nicht mehr in einem festen HTML-Code verankert sind, sondern aufgrund der Trennung ihrer Struktur medienneutral wiederverwendet werden können (2003).

Abgesehen von der Content Erstellung werden über das Editorial System, wie bereits erwähnt, auch die Inhalte verwaltet. Hierzu gehört laut Gersdorf unter anderem die Verwaltung von Links, die Bereitstellung eines Versionierungsmanagements und die Verwaltung von Zugriffsberechtigungen (2003). Letzteres ist ein wichtiger Bestandteil von Content Management Systemen. Laut Graf kann man anhand von Benutzernamen und Rollen, verschiedene Zugriffsrechte für verschiedene Nutzer einstellen und so beispielsweise zwischen Nutzern ohne Bearbeitungsrechte und Administratoren unterscheiden (2006). Dies spielt auch bei der Forschungslandkarte eine große Rolle. Man kann durch Zugriffsrechte zwischen den verschiedenen Nutzern (registriert vs. Gast) unterscheiden und dabei die Inhalte der Forschungslandkarte anzeigen, auf welche diese Zugriffsrechte haben. Über die Zugriffsverwaltung wird laut Bodendorf darüber hinaus das gleichzeitige Arbeiten an Content ermöglicht, um dabei Inkonsistenzen zu vermeiden (2005, S. 100).

Einen besonderen Stellenwert für die Forschungslandkarte E-Government haben, wie bereits oben erwähnt, Metadaten. So ist laut Rothfuss & Ried eine wichtige Funktion von Content Management Systemen, die Anreicherung der Inhalte mit Metainformationen während der Verwaltungsphase. Hierbei wird zwischen vordefinierten Metadaten, die sich bereits im CMS befinden und denen, die durch die Benutzer selbst angelegt werden, unterschieden (Rothfuss & Ried, 2001, S. 77). Die Leistungsfähigkeit des CMS hängt also auch davon ab, wie dieses mit den benutzerdefinierten Metainformationen umgehen und diese verstehen kann. Dies kann

beispielsweise dadurch erreicht werden, indem man für die Festlegung der Metainformationen eine standardisierte Datenbeschreibungssprache nutzt (Rothfuss & Ried, 2001, S. 78).

Content Repository:

Die Content Repository oder Content Base ist die Komponente eines Content Management Systems, in der die Inhalte gespeichert werden, nachdem diese generiert oder bearbeitet wurden. Somit bildet sie die Basis des Systems. Heutige Content Management Systeme basieren auf sehr komplexen Datenspeichersystemen, weil nicht nur heterogene Inhalte gespeichert werden müssen. Vielmehr müssen laut Rothfuss & Ried verschiedenartige Inhalte wie Bilder, Videos, Text, Audiodateien etc. gespeichert werden, die darüber hinaus in unterschiedlichen Formaten vorliegen (2001, S. 71). Das Speichern von komplexem Content wird seit geraumer Zeit durch die Entwicklung von sogenannten Datenbank Management Systemen ermöglicht, kurz DBMS. Moderne DBMS sind in der Lage, Inhalte verschiedener Art wie Bilder, Video oder Sounddateien in der Datenbank zu speichern (Rothfuss & Ried, 2001, S. 68). Der Zugang zu der Datenbank und letztendlich auch zu den Inhalten, wird über eine Schnittstelle des Content Management Systems ermöglicht.

Laut Bodendorf gibt es drei Aspekte, die berücksichtigt werden müssen, um die effiziente Speicherung und Publikation der Inhalte zu gewährleisten (2005):

- Trennung von Layout und Inhalt
- Abbildung von Makro- und Mikrostrukturen
- Einbindung dynamischer Informationsarten.

Die **Trennung von Inhalt und Layout** wird oft in der Literatur als eine der wichtigsten Prinzipien von Content Management Systemen erwähnt. Hierbei geht es laut Bodendorf um die getrennte Speicherung der Inhalte vom Layout. Diese getrennte Speicherung bringt wichtige Vorteile mit sich, ohne die die Publikation großer Mengen an Informationen nicht möglich wäre, oder aber nur verbunden mit einem unzumutbaren Aufwand (Bodendorf, 2005, S. 104). So muss laut Zschau nicht jedesmal ein neues Dokument erstellt werden, da die Trennung von Inhalt und Layout die einfache Wiederverwendung von Content ermöglicht. Der Inhalt wird lediglich aus der Datenbank ausgelesen und anhand vordefinierter Templates (Vorlagen) ausgegeben und dargestellt (Zschau, 2003). Weitere Vorteile des Trennungsprinzips sind nach

Bodendorf die einfache und schnelle Publikation von Content in verschiedenen Zielmedien (Cross Media Publishing) und die einfache Überarbeitung des Designs, beispielsweise einer Webseite, da nur die Layoutvorlagen angepasst werden müssen (2005, S. 104). Abbildung 3-6 zeigt die Datenorganisation im Content Repository.

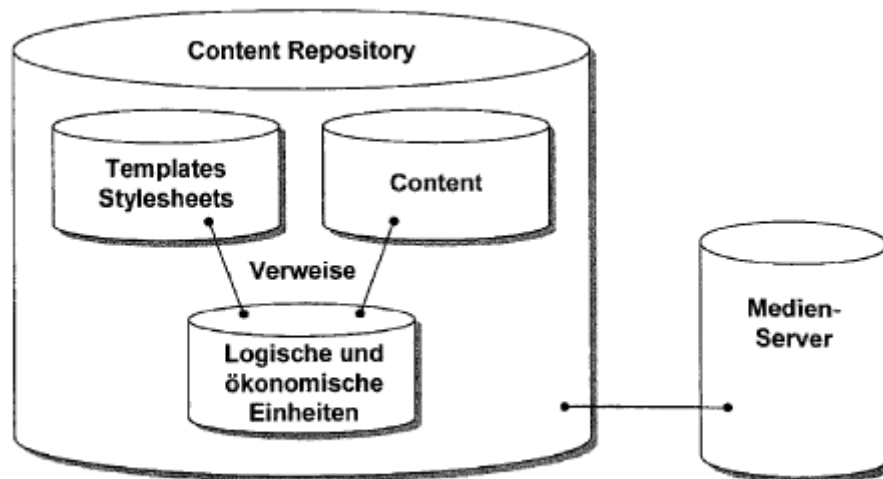


Abbildung 3-6: Datenorganisation im Content Repository (Bodendorf, 2005)

Laut Bodendorf bestehen Informationsobjekte aus einer **Makro- und einer Mikrostruktur** (2005, S. 104). Bei der Makrostruktur handelt es sich um zusammengesetzte Komponenten bestehend aus mehreren Informationsbausteinen. Um eine neue Einheit, bestehend aus bereits verwendeten Objekten zusammenzuführen, soll man nicht die Inhalte duplizieren, sondern lediglich auf diese mittels Referenzen verweisen, um unter anderem den benötigten Speicherplatz zu reduzieren. Die Mikrostruktur eines Informationsobjektes soll laut Bodendorf semantisch strukturiert sein, um das Prinzip der Trennung von Inhalt und Layout nicht zu verletzen, da eine semantische Strukturierung im Gegensatz zu der layoutorientierten Strukturierung keine Aussagen über das Aussehen der Inhalte trifft (2005).

Darüber hinaus soll wie bereits erwähnt das Content Repository die Speicherung und **Einbindung dynamischer Informationsarten** unterstützen, also von komplexen Informationsarten wie Videos oder Tönen, da diese in der heutigen Zeit und besonders im Online-Bereich immer wichtiger werden (Bodendorf, 2005, S. 105).

Publishing System:

Die dritte Komponente eines Content Management Systems ist das Publishing System. Über das Publishing System werden die Inhalte veröffentlicht und den jeweiligen Rezipienten angeboten. Das CMS kann dabei mittels Retrieval oder Navigation auf die benötigte Informationen zugreifen (Hengartner, 2003).

Der Prozess, der bei der Veröffentlichung der Inhalte durchgeführt wird, ist je nach Zielmedium immer unterschiedlich. Für diese Ausarbeitung ist nur der Publikationsprozess von Online-Angeboten von Bedeutung. Zu den Funktionen dieses Veröffentlichungsprozesses, der durch das Publishing System durchgeführt wird, zählen laut Gersdorf „...die Kontextspezifische Selektion relevanter Contents, deren geeignete Kombination, die medienspezifische optische Aufbereitung sowie die eigentliche Ausgabe“ (2003). Soll ein Nutzer also Content aufrufen, so muss laut Bodendorf das Publishing System in der Lage sein, anhand der in Form von Metadaten gespeicherten Verweise, die notwendigen Inhalte sowie die für das Layout relevanten Templates aus dem Content Repository zu erheben und diese dann beispielsweise als HTML-Seiten ausgeben (siehe Abbildung 3-7) (2005, S. 106).

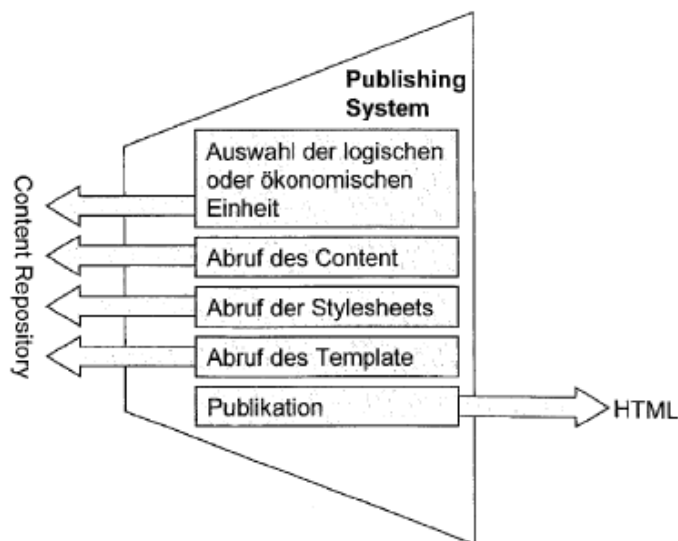


Abbildung 3-7: Publishing System (Bodendorf, 2005)

Laut Bodendorf kann Content in zwei verschiedenen Formen generiert werden. Bei der **statischen Generierung** werden die vollständigen HTML-Seiten, die aus verlinkten HTML-Dokumenten bestehen, generiert und an den Rezipienten weitergegeben. Diese stehen dann auf dem Server als starre HTML-Seiten zur Verfügung. Dieses hat laut Bodendorf zwar Vorteile in Bezug auf die benötigte Leistung der Hardware, führt aber

dazu, dass eine halbwegs automatisierte und personalisierte Contentgenerierung nicht möglich ist. Für personalisierte Medienprodukte, welche an die Benutzerbedürfnisse angepasst sind, benötigt man eine **dynamische Contentgenerierung** (Bodendorf, 2005, S. 107). Diese ist außerdem notwendig, wenn laut Bodendorf „Inhalte häufig erneuert oder geändert werden und eine zeitnahe Veröffentlichung nötig ist...“ (2005, S. 107). Dann wird die Webseite mit dem benötigten Content beim Aufruf durch den Nutzer dynamisch aus dem Content Repository erstellt. Dies führt laut Gersdorf allerdings zu einem höheren Aufwand und letztendlich zu höheren Kosten im Bezug auf die Planung und Generierung von Content (2003).

Eine Funktion von Content Management Systemen, welche komponentenübergreifend, also von der Contentbeschaffung bis zur Publikation, zur Unterstützung des Publikationsprozesses von Content gebraucht wird, ist das Workflow-Management. Hierbei geht es laut Gersdorf, um die „...Steuerung und Überwachung des Publikationsprozesses ...“ (2003). Über ein Workflowmodell wird dann der Ablauf der zu bearbeitenden Aufgaben und die dafür zuständigen Mitarbeiter festgelegt und die benötigten Daten und Dokumente bereitgestellt, was zu einer Vereinfachung und klaren Darstellung des Arbeitsflusses führt (Bodendorf, 2005, S. 101).

Die Inhalte der Forschungslandkarte sollen anhand eines Content Management Systems verwaltet werden. Zwar können durch das CMS einige temporale Informationen, wie Versionsmanagement oder Meldungen über neuen Content verwaltet werden, eine vollständige Historisierung der Inhalte wird allerdings nicht unterstützt. Dies kann durch das Konzept der temporalen Datenhaltung erreicht werden, welches im nächsten Kapitel vorgestellt wird.

4 Temporale Datenhaltung

Dieses Kapitel behandelt das Prinzip der temporalen Datenhaltung in der Informationstechnik. Das Ziel der temporalen Datenhaltung ist es, die Entwicklung von Informationen oder Daten zu dokumentieren und auf diese Weise eine Historie aufzubauen, an der man die zeitliche Veränderung von Dokumenten nachvollziehen kann. Somit bildet dieses Verfahren die Grundlage für die Historisierung der Inhalte.

4.1 Notwendigkeit von Historisierung

Digitale Objekte unterliegen einem ständigen Wandel, der daraus resultiert, dass diese sich während des Lebenszyklus oft verändern, aktualisiert oder gelöscht werden. Ohne das Konzept der temporalen Datenhaltung wird der Wert eines Objekts einfach überschrieben. Dies führt dazu, dass nur der aktuelle Wert oder Zustand des Objekts gespeichert wird. Ist es allerdings notwendig, die zeitliche Entwicklung von Objekten festzuhalten, benötigt man die temporale Datenhaltung (Myrach, 2005). Bei der temporalen Datenhaltung wird bei Änderung eines Objekts eine neue Version des Objekts erstellt und in der Datenbank gespeichert. Da die alte Version weiterhin auf der Datenbank gespeichert bleibt, ist es möglich nachzuvollziehen, welche Änderungen an diesem Objekt durchgeführt wurden, und welche Zustände es in der Vergangenheit hatte. Auf diese Weise entsteht eine Historie. Dieser Prozess wird in der Informationstechnik als Historisierung bezeichnet. Laut Frick & Najar existieren zwei Arten der Historisierung (2009, S. 29):

Bei der **vollständigen Historisierung** wird der gesamte Datenbestand in regelmäßigen Zeitabständen festgehalten. Dies führt dazu, dass verschiedene Versionen des gesamten Datensatzes in der Datenbank gespeichert werden.

Bei der **inkrementellen Historisierung** werden die Änderungen an einzelnen Objekten oder deren Bestandteilen dokumentiert.

In der Forschungslandkarte sollen die Änderungen an den einzelnen Instanzen festgehalten werden, sodass sich diese Ausarbeitung nur mit dem Thema der inkrementellen Historisierung beschäftigt.

4.2 Temporale Datenbanken

Durch die Notwendigkeit der Historisierung von digitalen Objekten, begann man bereits in den 70er Jahren Methoden zu entwickeln, um zeitliche Bezüge in Datenbanken abzubilden (Etzion, Gal, & Segev, 1998). Als Ergebnis dieser Forschungsaktivitäten wurden temporale Datenbanken entwickelt. Den obengenannten Autoren zufolge ermöglicht eine temporale Datenbank die Speicherung verschiedener Werte des gleichen Objekts, die zu unterschiedlichen Zeitpunkten bestanden haben. Die Abbildung temporaler Zuordnungen der realen Welt bildet laut Goralwalla, Özsu und Szafron die technische Voraussetzung verschiedener Applikationen, wie beispielsweise ökonomische Zeitreihen, Online-Banking, Echtzeitsysteme, Online-Buchungen oder Lagerhaltungssoftware (1998).

Im Gegensatz zu konventionellen Datenbanken, ist mit temporalen Datenbanken die Abbildung und Verwaltung von notwendigen Zeitgrößen möglich. Dabei werden laut Atzeni, Del Nostro & Paolozzi drei verschiedene zeitliche Dimensionen unterschieden (2012). Die **benutzerdefinierte Zeit** sind zeitliche Werte, die vom Benutzer festgelegt sind und wie nicht-temporale Attribute in der Datenbank behandelt werden. Ein Beispiel hierfür ist laut Ebner das Geburtsdatum einer Person. Dieses Datum ändert sich, außer im Falle einer falschen Eingabe, nie. Wurde das Datum falsch eingegeben, bedarf es keiner Historisierung, da das Datum einfach überschrieben werden kann (Ebner, 2002, S. 25). Daher müssen solche Eingaben im Kontext der temporalen Datenhaltung nicht berücksichtigt werden. Für die Historisierung relevante Zeitgrößen sind die **Gültigkeitszeit** und die **Transaktionszeit**. Bei der Gültigkeitszeit handelt es sich um die Zeit, in der ein Objekt oder ein Datenelement in der realen Welt gültig ist. Ein Element kann sowohl in der Vergangenheit als auch in der Zukunft gültig sein (Atzeni, Del Nostro, & Paolozzi, 2012).

Die Transaktionszeit bezeichnet den Zeitpunkt, zu welchem ein Objekt oder ein Datenelement in der Datenbank gespeichert, entfernt oder aktualisiert wurde. Im Gegensatz zu der Gültigkeitszeit, die vom Benutzer festgelegt wird, wird die Transaktionszeit vom System bestimmt. Da diese den Zeitpunkt der Speicherung des Elements in der Datenbank repräsentiert, kann die Transaktionszeit nicht in der Zukunft liegen (vgl. Ebner, 2002, S. 25). Temporale Datenbanken, die sowohl die Gültigkeits-

als auch die Transaktionszeit verarbeiten können, werden als bitemporale Datenbanken bezeichnet (Kumar, Tsotras, & Faloutsos, 1998).

Werden Daten in einer Datenbank mit einem zeitlichen Bezug versehen, spricht man laut Zapf & Myrach von Zeitstempelung (2000). Durch Zeitstempel lassen sich unterschiedliche zeitliche Dimensionen abbilden. Wird nur ein Zeitstempel an ein Datenelement angehängt, repräsentiert dies einen Zeitpunkt. Ist es nötig ein Zeitintervall zu repräsentieren, setzt man einen Zeitstempel als Startzeitpunkt und einen weiteren als Endzeitpunkt. Darüber hinaus lässt sich anhand eines Zeitstempels und einer Dauer eine Zeitperiode abbilden (Ebner, 2002, S. 26). Auf diese Weise ist es möglich, anhand von Zeitstempelung, die für die Historisierung notwendigen Zeitdimensionen (Transaktionszeit und Gültigkeitszeit) an Datenelemente anzuhängen. Somit ist die Zeitstempelung ein zentraler Begriff der temporalen Datenhaltung.

Eine Entscheidung, die im Kontext von temporaler Datenhaltung getroffen werden muss, ist die der Granularität der Zeitstempelung. Laut Balzer unterscheidet man in der Theorie temporaler Datenbanken zwei Arten der Zeitstempelung (2001). Bei der **Tupelzeitstempelung** wird an einen Tupel ein Zeitstempel angehängt. Dies führt dazu, dass bei jeder Änderung der gesamte Tupel historisiert wird. Dies geschieht, indem eine neue Version des Tupels in der Datenbank gespeichert wird. Laut Ebner bezeichnet ein Tupel in der Datenbank Theorie, einen Datensatz (2002, S. 26). Für den Kontext der Forschungslandkarte soll der Begriff Tupel als Synonym für ein Objekt bzw. eine Instanz verstanden werden. Im Gegensatz dazu werden bei der **Attributzeitstempelung** einzelne Attribute mit Zeitstempeln versehen. So wird im Falle einer Änderung eines Attributs, nur das Attribut historisiert und nicht das ganze Objekt. Für welche Art der Zeitstempelung man sich schließlich entscheidet, hängt laut Balzer vom Kontext ab, den man abbilden möchte. Resultieren die Änderungen des Datensatzes überwiegend aus Änderungen einzelner Attribute, soll man sich für die Attributzeitstempelung entscheiden. Dies hat den Vorteil, dass redundante Datenhaltung verhindert wird, da bei jeder Attributsänderung nur das jeweilige Attribut historisiert wird. Allerdings ist dieser Prozess mit einem höheren Implementierungsaufwand verbunden, da deutlich mehr Zeitstempel gesetzt werden müssen (Balzer, 2001, S. 19).

4.3 Temporale Webdokumente

Im Kapitelabschnitt 4.2 wurde das Konzept temporaler Datenbanken vorgestellt. Wie in der Datenbanktheorie, können auch Nutzer im Web von der temporalen Verwaltung von Internetdokumenten profitieren. So besteht ein mögliches Interesse seitens der Nutzer an temporalen Informationen von Webdokumenten, wie zum Beispiel dem letzten Aktualisierungszeitpunkt oder eine frühere Version der Inhalte einer Webseite. Die große Herausforderung bei der Historisierung von Webseiten ist es, die Konzepte der temporalen Datenhaltung in Datenbanken auf das Web zu übertragen. Dabei müssen Webseiten auf eine Weise strukturiert werden, die die Abbildung der Transaktions- und Gültigkeitszeit erlaubt (vgl. Ebner, 2002, S. 32).

Laut Ebner kann man durch Zeitstempelung die **Transaktionszeit von Webseiten oder Dateien** festlegen. Hierbei legt man anhand von Transaktionszeitstempeln ein Zeitintervall fest von dem Moment an, an dem die Webseite erstellt oder geändert wurde bis zum Zeitpunkt des Löschens oder erneuten Ändern der Seite (Ebner, 2002). Dies sagt aus, dass die Webseite bzw. die Datei in diesem Intervall online unter der entsprechenden URL verfügbar war oder ist. Auf diese Weise ist es möglich, den aktuellen sowie den vergangenen Zustand bzw. Version eines Objekts anzusehen.

Auch das Abbilden der **Gültigkeitszeit von Internetinhalten** ist theoretisch möglich. Hierzu muss man laut Ebner die Gültigkeit der einzelnen Bestandteile eines Objekts anhand von Zeitstempeln markieren (2002, S. 33). So ist man in der Lage, den Zustand eines Objekts zu einem bestimmten Zeitpunkt in der Vergangenheit aber auch in der Zukunft anzusehen. Dies ist allerdings nur möglich, wenn der Browser diese temporalen Bezüge in Form von Gültigkeitstags unterstützt bzw. erkennt, was aber nicht immer der Fall ist.

Obwohl die zeitlichen Anforderungen an Webdokumente, die die Historisierung der Inhalte ermöglichen, ähnlich mit denen der temporalen Datenbanktheorie sind, gibt es laut Atzeni, Del Nostro, & Paolozzi einige Unterschiede, die betrachtet werden müssen. Während bei temporalen Datenbanken das Hauptaugenmerk auf dem Speichern der historischen Entwicklung von Daten und deren Abfrage anhand einer Datenbanksprache liegt, muss die Historie einer Webseite so zur Verfügung gestellt werden, dass Nutzer diese über den Browser nachvollziehen können (Atzeni, Del Nostro, & Paolozzi, 2012). Wie bereits erwähnt werden die meisten Webseiten mit

einem Content Management System verwaltet. Daher ist es für die Historisierung von Webseiten laut Atzeni, Del Nostro, & Paolozzi unbedingt notwendig, dass die zeitlichen Dimensionen der Inhalte innerhalb des CMS repräsentiert werden können (2012).

Um Webseiten historisieren zu können, muss zusammenfassend ein Konzept entwickelt werden, das die Übertragung der zeitlichen Aspekte der temporalen Datenbanktheorie auf Content Management Umgebungen ermöglicht. Dabei müssen die temporalen Informationen der Daten bereits bei der Konzeptionierung und Datenmodellierung der Webseite mitberücksichtigt werden.

4.4 Temporale Erweiterung der Datenmodellierung und Modell R_{ETTE}

Aufgrund fehlender Datenbanklösungen, die die zeitliche Dimension von Daten auf eine Weise berücksichtigen, mit der die Historisierung von Inhalten problemlos umgesetzt werden kann, erkannten viele Autoren die Bedeutung, temporale Aspekte bei der Konzeptionierung von Systemen zu modellieren (vgl. Kaiser, 1999). Unterschiedliche Ansätze temporaler Datenmodellierung basierend auf dem Entity-Relationship-Modell nach Chen (1976) wurden in den letzten Jahren vorgestellt. Ein Überblick temporal erweiterter Datenmodelle geben Gregersen & Jensen (1999). Anhand der vorgestellten Modelle ist der Designer eines Systems in der Lage, das Entity-Relationship-Modell mit temporalen Aspekten zu erweitern, um festzulegen, welche Objekte historisiert werden sollen und welche nicht. Allerdings führt die Integration der Zeitdimension in das ER-Modell zu einer erhöhten Komplexität bei der Modellierung der Datenschicht, sodass eine einfache, aber zugleich aussagekräftige Notation benötigt wird (vgl. Atzeni, Del Nostro, & Paolozzi, 2012). Nun wird das Modell R_{ETTE} vorgestellt, mit dem die temporalen Aspekte der Inhalte der Forschungslandkarte modelliert werden sollen.

Das Modell R_{ETTE} nach Kaiser (1999) schlägt eine Notation vor, mit der man die Zeitdimension von Daten auf der konzeptionellen Ebene modellieren kann. Genauer gesagt handelt es sich hierbei um eine temporale Erweiterung des Entity-Relationship-Modells nach Chen (1976). Um die Komplexität zu minimieren, beschränkt sich das R_{ETTE} Modell auf das Abbilden der Gültigkeitszeit. Diese wird als Intervall anhand der Zeitstempel T_B (Anfangszeitpunkt) und T_E (Endzeitpunkt) festgelegt. Wie beim gewöhnlichen Chen Modell, werden beim R_{ETTE} Modell Entitäten, Attribute und Beziehungen modelliert. Der Designer kann entscheiden, welche Objekte, ob Entitäten,

Attribute oder Beziehungen, zeitbezogen sind. Ist ein Objekt zeitbezogen, ist man daran interessiert, die Historie des Objekts auch nach dessen Zeitablauf festzuhalten.

Entitäten:

Entitäten werden im RETTE Modell nicht gesondert, sondern wie gewohnt durch ein Rechteck gekennzeichnet. Ob eine Entität historisiert werden soll erkennt man daran, dass anhand der Zeitstempelattribute T_B und T_E die Gültigkeitszeit der Entität festgelegt wird. Die Zeitstempel werden als zusätzliche Attribute an zeitbezogene Entitäten angehängt. Auf diese Weise werden verschiedene Versionen von Entitäten abgebildet, sodass die Historie einer Instanz anhand der Versionen festgehalten wird.

Die Historisierung einer Entität ist nur möglich, falls diese anhand eines zeitunabhängigen Schlüssels eindeutig identifizierbar ist. Der zeitunabhängige Schlüssel wird anhand eines oder mehrerer zeitunabhängigen Attribute festgelegt. Die Historisierung ist nur möglich, wenn sich der Schlüssel auch bei der Änderung einer Entität nicht ändert. Somit wird im Fall der Änderung einer Entität ein neues Entitätsexemplar mit dem gleichen Schlüssel in der Datenbank angelegt. Die Gesamtzahl aller Entitätsexemplare mit gleichem Schlüssel repräsentiert ein Objekt. Die einzelnen Entitätsexemplare mit gleichem Schlüssel sind Versionen eines Objekts (vgl. Kaiser, 1999). Verfügt eine Entität nicht über einen eindeutigen Schlüssel oder wird dieser bei der Historisierung geändert, kommt es laut Sondermann zu sogenannten Phantomänderungen des Datenbestandes. Eine Entität ohne Schlüssel ist nämlich nicht identifizierbar, sodass nicht festgestellt werden kann, ob sich die Entität geändert hat oder eine neue hinzugefügt wurde (Sondermann, 2006).

Attribute:

Im RETTE Modell werden drei Arten von Attributen unterschieden: Zeitstempel-, Zeitunabhängige- und Zeitabhängigeattribute.

Zeitstempelattribute legen wie bereits erwähnt die Gültigkeitszeit einer Entität fest.

Zeitunabhängige Attribute können sich während des Lebenszyklus einer Entität nicht ändern. Das gängige Beispiel hierfür ist das Geburtsdatum einer Person.

Zeitabhängige Attribute können sich ändern und sind somit für die Historisierung relevant. Das RETTE Modell unterscheidet drei Arten von zeitabhängigen Attributen, die durch zusätzliche Symbole im Modell dargestellt werden (Abbildung 4-1):

- Zeitabhängige Attribute im weiteren Sinn können sich ändern, das Festhalten älterer Werte dieser Attribute ist allerdings nicht relevant. Es wird also nur der aktuelle Wert in der aktuellen Version des Objekts überschrieben. Ein Beispiel hierfür ist die Adresse einer Person.
- Zeitabhängige Attribute im engeren Sinn können sich ebenfalls im Zeitablauf eines Objekts ändern. Diese Änderungen müssen allerdings dokumentiert werden, sodass die Historie des Attributs nachvollziehbar ist. Das bedeutet, dass eine neue Version des entsprechenden Objekts im Falle einer Änderung in der Datenbank gespeichert werden und alte Versionen erhalten bleiben müssen. Im Falle der Forschungslandkarte wären die Forschungsfelder eines Instituts ein zeitabhängiges Attribut im engeren Sinn.
- Eine weitere Art zeitabhängiger Attribute sind zyklische Attribute. Diese müssen ebenfalls historisiert werden. Ändert sich ein zyklisches Attribut nicht, obwohl dies möglich gewesen wäre, muss diese Tatsache abgebildet werden. Zyklische Attribute sind für die Forschungslandkarte nicht relevant und sollen nicht weiter berücksichtigt werden.

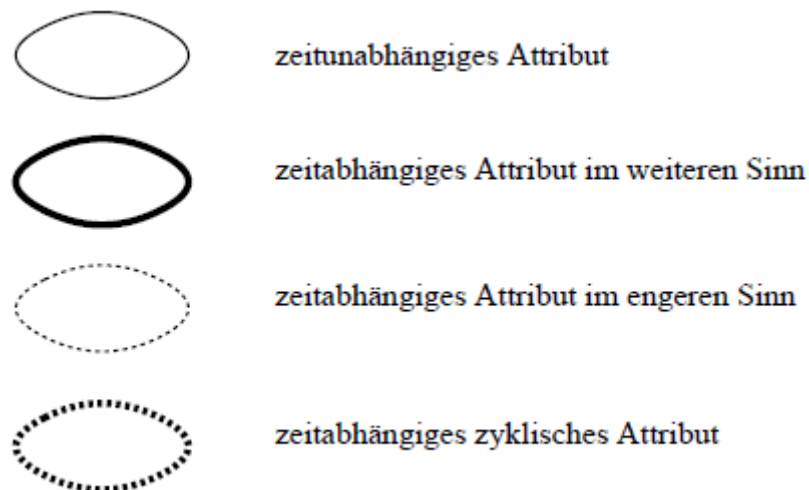


Abbildung 4-1: Graphische Darstellung der unterschiedlichen Attributarten (Kaiser, 1999)

Beziehungen:

Auch Beziehungen können im Modell RETTE zeitbezogen sein. Soll eine Beziehung bzw. ein Beziehungstyp historisiert werden, werden die Attribute T_B und T_E an diese angehängt (Abbildung 4-2). Darüber hinaus können Beziehungen weitere

zeitunabhängige und temporale Attribute besitzen. Laut Kaiser können zeitbezogene Beziehungstypen sowohl zwischen konventionellen als auch zwischen zeitbezogenen Entitäten bestehen (1999).

Ein Beispiel für einen zeitbezogenen Beziehungstyp ist die Verknüpfung „arbeitet in“ zwischen Person und Forschungsinstitut in der Forschungslandkarte. Ist diese Beziehung zeitbezogen, kann die Historie dieses Beziehungstyps festgehalten werden. Auf diese Weise ist es möglich nachzuvollziehen, welche Personen aktuell in einem bestimmten Institut beschäftigt sind und in der Vergangenheit beschäftigt waren.

Das Modell RETTE erlaubt Beziehungen mit folgenden Kardinalitäten, wobei das letztere die einzige relevante Art bei der Forschungslandkarte ist:

- 1 : 1 Beziehungen
- 1 : n Beziehungen
- n : m Beziehungen

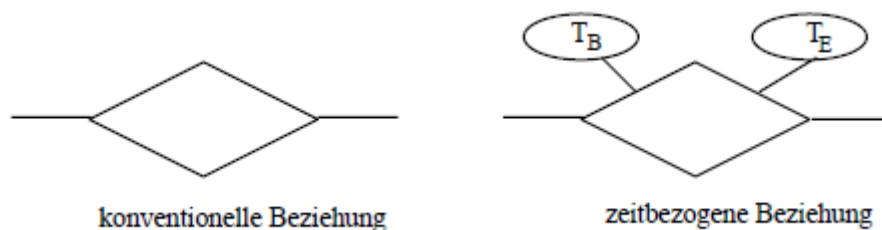


Abbildung 4-2: Graphische Darstellung der Beziehungstypen im Modell RETTE (Kaiser, 1999)

4.5 Historisierung von Metadaten

Webdokumente bestehen nicht nur aus den eigentlichen Inhalten, sondern wie bereits erwähnt auch aus Metadaten, die die Inhalte beschreiben. Laut Balzer kann die Änderung der Metadaten unter Umständen zu einer Änderung der Inhalte führen, die sie beschreiben (2001, S. 39). Daher kann eine korrekte und vollständige Historisierung nur erfolgen, wenn auch die Metadaten historisiert werden. Dabei besteht eine zeitliche Interdependenz zwischen Daten- und Metadatenebene, die unbedingt berücksichtigt werden muss. Beispielsweise darf laut Ramsch & Sosna zu jedem Zeitpunkt nur eine gültige Version von Metadaten bestehen (1997).

Des Weiteren muss bei der Konzeptionierung der Forschungslandkarte die Thematik der Historisierung der Metadaten berücksichtigt werden. Dies soll aber zu einem

späteren Zeitpunkt geschehen, da sonst der Rahmen dieser Ausarbeitung gesprengt würde.

4.6 Zusammenfassung

Zusammenfassend lässt sich festhalten, dass die Historisierung der Inhalte der Forschungslandkarte nicht ohne das Konzept der temporalen Datenhaltung umgesetzt werden kann. Allerdings wurde dieses Konzept für die Anwendung in Datenbanken entwickelt. Die Forschungslandkarte E-Government ist eine Online-Plattform, welche durch ein Content Management System verwaltet werden soll, wodurch besondere Anforderungen für die temporale Verwaltung der Inhalte entstehen. Die Herausforderung besteht darin, die zeitlichen Konzepte der temporalen Datenhaltung auf Content Management Systeme und das Web zu übertragen. Dabei ist es wichtig, dass temporale Aspekte bereits während des Designs und der Konzeptionierung der Forschungslandkarte berücksichtigt werden, da eine vollständige und korrekte Historisierung der Inhalte in Form bitemporaler Datenhaltung nicht durch standardisierte Funktionen von Content Management Systemen unterstützt wird.

Die Inhalte der Forschungslandkarte bestehen nicht nur aus dem Content, sondern sind eine Einheit aus Content und Metainformationen, die die Inhalte beschreiben. Der Begriff Metadaten soll im nächsten Kapitel näher betrachtet werden und die Arten von Metadaten und deren Funktionen aufgezeigt werden.

5 Metadaten

Metadaten ist einer der bedeutsamsten und wichtigsten Begriffe im Kontext von Informations- und internetbasierten Technologien. Aufgrund der Reichhaltigkeit des Begriffes haben Metadaten in der Literatur unzählige Verwendungen und oft verschiedene Bedeutungen. Das folgende Kapitel zielt darauf ab, den Begriff Metadaten näher zu erklären und deren Funktionen und Bedeutung aufzuzeigen.

5.1 Begriffserklärungen

Der Begriff Metadaten hat seine Wurzeln im Griechischen sowie im Lateinischen. Meta stammt aus dem griechischen und bedeutet „über“, während Datum im lateinischen als „das Gegebene“ übersetzt werden kann (Schumann, 2007). Vereinfachend werden in der Literatur Metadaten als „Daten über Daten“¹ definiert, also Informationen, die elektronische Daten beschreiben. Elektronische Daten werden dabei nicht einmalig mit Metainformationen angereichert, vielmehr können Metadaten während des gesamten Content Life Cycles entstehen, sei es bei der Beschaffung, der Verwaltung oder der Archivierung der Inhalte. Dabei bilden die Metadaten zusammen mit den inhaltlichen Daten laut Dobratz, et al. die logische Einheit „digitales Objekt“ (2009).

5.2 Arten von Metadaten

Es gibt viele verschiedene Arten von Metadaten in den unterschiedlichen Anwendungsbereichen der Informationstechnik. Aufgrund dieser Reichhaltigkeit wurde von bestimmten Institutionen der Versuch gestartet, Standards für Metadaten durchzusetzen. Eine Institution, die sich unter anderem dieses Ziel gesetzt hat, ist die National Information Standards Organization. NISO ist eine Non-Profit Organisation, die das Ziel verfolgt, technische Standards zu entwickeln, um Informationen effektiv verwalten zu können.

Nach NISO² gibt es folgende Arten von Metadaten:

Beschreibende Metadaten:

Beschreibende Metadaten beschreiben digitale Objekte, um die Identifikation und Retrieval der Objekte zu ermöglichen. Mögliche beschreibende Metadaten sind beispielsweise Autor, Titel, Schlagwörter oder eine Zusammenfassung. Der

¹ Institut für Museumsforschung SMB-PK / nestor Arbeitsgruppe (2009): Langzeiterhaltung digitaler Daten in Museen, <http://www.langzeitarchivierung.de/eng/publikationen/handbuch/informationsblaetter.htm>

² NISO Press, National Information Standards Organization (2004): Understanding Metadata, <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

bekannteste Standard für beschreibende Metadaten ist Dublin Core, was in Kapitelabschnitt 5.4 näher beschrieben wird.

Strukturelle Metadaten:

Strukturelle Metadaten beschreiben, wie Objekte zusammengesetzt sind und in welchem Zusammenhang sie zu anderen Objekten oder Daten stehen. Laut Wiemer legen strukturelle Metadaten die Struktur von Dokumenten fest, sodass man beispielsweise zwischen einzelnen Kapiteln oder Punkten innerhalb eines Dokuments springen kann (2008, S. 12). Strukturelle Metadaten werden z. B. durch den „Metadata Encoding and Transmission Standard“ (METS) strukturiert.

Administrative Metadaten:

Administrative Metadaten werden dazu genutzt, um Ressourcen besser verwalten zu können. Beispiele sind Erstelldatum, Dateiformat, Speicherort und andere technische Metadaten sowie Informationen zu Urheber- und Verwertungsrechten und Informationen bezüglich der Archivierung und Langzeiterhaltung von Objekten.

5.3 Funktionen von Metadaten

Die Definition, „Daten über Daten“³ ist eine sehr allgemeine und vereinfachende Aussage über Metadaten. So versteht man intuitiv Metadaten als Informationen, die ein Objekt zusammenfassend beschreiben. Dies ist auch eine der Funktionen von Metadaten, denn so legen diese bestimmte Eigenschaften fest, wie zum Beispiel Autor eines Dokuments oder zusammenfassende Schlagwörter. Aufgrund der vielen verschiedenen Datentypen, die heutzutage genutzt werden, ist Metadaten allerdings ein sehr umfassender Begriff, bestehend aus verschiedenen Arten von Metadaten, mit verschiedenen Funktionen.

Metadaten werden laut Kashyap, Bussler & Moran beispielsweise dazu genutzt, um Eigenschaften von Daten oder Dokumenten zu speichern, welche für den **Informationsretrieval** nötig sind (2008, S. 25). So kann anhand von Metainformationen der Zugriff auf Daten und Dokumente ermöglicht werden, welche in Datenbanksystemen gespeichert sind. Laut Dobratz, et al. kann dabei der Zugriff auf ein Objekt nur gesichert werden, indem das Objekt anhand von Metadaten inhaltlich, formal und strukturell beschrieben wird (2009).

³ Institut für Museumsforschung SMB-PK / nestor Arbeitsgruppe (2009): Langzeiterhaltung digitaler Daten

Metadaten

Wie bereits erwähnt haben Metadaten auch eine zusammenfassende Funktion. Hierbei werden Dokumente mit **inhaltlichen Eigenschaften** wie Autor oder Schlagwörtern versehen, außerdem kann das Dokument mit einer kleinen Zusammenfassung des Inhalts ergänzt werden.

Metadaten erfüllen darüber hinaus eine weitere Funktion, welche für die vorliegende Arbeit von großer Bedeutung ist. So werden Metainformationen laut Kashyap, Bussler, & Moran dafür verwendet, um die **Attribute von Objekten festzulegen und Beziehungen zwischen heterogenen Objekten zu beschreiben** (2008, S. 25). Dies ist sehr wichtig im Kontext der Forschungslandkarte E-Government, denn anhand der Metadaten werden bei der Konzeptionierung und später bei der Implementierung die Instanzen und deren Relation zueinander erfasst und modelliert. Auch die **Interoperabilität** von Ressourcen wird durch Metadaten ermöglicht, die Daten in einer Weise beschreiben, die diese maschinenlesbar machen. So können anhand von Metadatenschemata und Übertragungsprotokollen Daten zwischen heterogenen Systemen ausgetauscht werden (Guenther & Radebaugh, 2004).

Metadaten werden außerdem dazu genutzt, Veränderungen an Daten und Dokumenten zu dokumentieren. Dies spielt bei der **Versionierung** eine übergeordnete Rolle und hilft die **Authentizität** der Inhalte zu gewährleisten (Dobratz, et al., 2009).

Anhand von technischen Metadaten werden laut Dobratz, et al. weitere Kriterien, wie die **Integrität** oder die **Interpretierbarkeit** der Daten gesichert und weitere Vorhaben, wie z. B. die **Langzeiterhaltung und Archivierung** der Daten ermöglicht (2009).

Zusammenfassend erhöhen Metadaten die Verwendbarkeit der beschriebenen Daten, denn „je umfassender und differenzierter die Metadaten sind, desto flexibler sind die Recherche, die automatische Verknüpfung mit anderen digitalen Objekten und desto höher ist letztendlich die wissenschaftliche Nutzbarkeit des Objektes“ (Guenther & Radebaugh, 2004).

5.4 Dublin Core

5.4.1 Strukturierung von Metadaten

Mit der zunehmenden Bedeutung des Internets begann man Versuche zu unternehmen, um Metadaten zu strukturieren. Als Ergebnis wurden unterschiedliche Metadatenschemata entwickelt, die einheitliche und oft standardisierte Vokabulare zur Beschreibung von Ressourcen festlegen. Die Standardorganisation NISO definiert

Metadaten schemata als Sammlungen von Metadatenelementen, um bestimmte Ressourcen zu beschreiben. Hierbei werden laut Guenther & Radebaugh die Namen von Metadatenelementen, deren Syntax und in manchen Fällen auch Regeln über die Darstellung oder die Formulierung von Content spezifiziert (2004). Eines der wichtigsten und am meisten anerkannten Metadaten schemata ist das Dublin Core metadata, welches im nächsten Kapitelabschnitt vorgestellt werden soll.

5.4.2 Dublin Core metadata

Dublin Core ist eine Sammlung von Metadatenelementen, die dazu genutzt werden, um Ressourcen im Web zu beschreiben⁴. Dublin Core wurde von der „Dublin Core Metadata Initiative“ (DCMI) entwickelt, um Objekte im Internet mit semantischen Metadaten zu bereichern und auf diese Weise das domänenübergreifende Auffinden von Ressourcen zu ermöglichen (Weibel & Koch, 2000). Laut Sugimoto, Baker, & Weibel wurde während eines Workshops, der in Dublin, Ohio in 1995 gehalten wurde, das klassische Dublin Core verfasst, ein Metadaten set mit 13 Elementen, um Ressourcen zu klassifizieren. Bis 1998 wurde es um zwei weitere Elemente ergänzt (Sugimoto, Baker, & Weibel, 2002). Dies ist das sogenannte „Simple Dublin Core“. Die Abbildung 5-1 zeigt die 15 Elemente des „Simple Dublin Core“.

Identifier	Definition
Title	A name given to the resource.
Creator	An entity primarily responsible for making the content of the resource.
Subject	The topic of the content of the resource.
Description	An account of the content of the resource.
Publisher	An entity responsible for making the resource available.
Contributor	An entity responsible for making contributions to the content of the resource.
Date	A date associated with an event in the life cycle of the resource.
Type	The nature or genre of the content of the resource.
Format	The physical or digital manifestation of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Source	A reference to a resource from which the present resource is derived.
Language	A language of the intellectual content of the resource.
Relation	A reference to a related resource.
Coverage	The extent or scope of the content of the resource.
Rights	Information about rights held in and over the resource.

Abbildung 5-1: The Fifteen Elements of „Simple Dublin Core“ (Sugimoto, Baker, & Weibel, 2002)

⁴ <http://dublincore.org/>

Die 15 Elemente des „Simple Dublin Core“ entsprechen in etwa den möglichen Eigenschaften einer Internetressource. Allerdings ist die Aussagekraft dieser Elemente relativ beschränkt, sodass mit der Zeit Erweiterungen entwickelt wurden, um Ressourcen genauer beschreiben zu können. Anhand sogenannter Qualifier werden die Elemente des „Simple Dublin Core“ semantisch ergänzt und deren Bedeutung spezifiziert (Baker, 2000). Werden Qualifier zusammen mit den Elementen eingesetzt, spricht man vom „Qualified Dublin Core“.

Die DCMI erkennt zwei unterschiedliche Arten von Qualifier an⁵:

Element Refinements spezifizieren die Bedeutung eines Elements. Die Bedeutung des Elements an sich bleibt die gleiche, sodass der Qualifier, falls ein Client diesen nicht versteht, einfach ignoriert werden kann. Ein Beispiel hierzu ist der Qualifier „Created“ vom Element „Date“, welches das Erstelldatum einer Ressource festlegt.

Encoding scheme identifiziert ein Element als ein Teil eines standardisierten Vokabulars oder einer formalen Notation. Beispielsweise sagt laut Sugimoto, Baker, & Weibel der Qualifier „LCSH“ aus, dass der Wert eines Elements vom Typ „Subject“ dem Standard der Library of Congress Subject Headings folgt (2002).

Anhang 2: DCMI Recommended Qualifiers zeigt eine Liste aller Qualifier, die durch die DCMI den Status „Recommended“ erhalten haben.

5.5 Tagging

5.5.1 Taggingbegriff und Folksonomien

Im Kapitelabschnitt 5.3 wurde die Bedeutung von Metadaten und deren Funktionen für elektronische Objekte erläutert. Wie bereits ausgeführt gibt es Metadaten, die vom System initialisiert festgelegt werden, andere werden durch die Benutzer des Systems eingegeben. Eine besondere Erscheinungsform von benutzergenerierten Metadaten sind sogenannte Tags. Ein Tag ist laut Alby ein Begriff oder ein Schlagwort, welcher ein bestimmtes Objekt beschreibt (2008, S. 127). Tags werden durch die Benutzer an Objekte angehängt, wodurch Tags beliebige Ressourcen wie zum Beispiel Webseiten, Fotos, Videos usw. durch die Nutzer selbst klassifiziert und kategorisiert werden

⁵ <http://dublincore.org/documents/usageguide/qualifiers.shtml>

können. Diesen Vorgang nennt man Tagging. Eine wichtige Eigenschaft des Tagging-Prozesses ist die Freiheit, die dem Benutzer beim Erstellen eines Tags eingeräumt wird. Im Gegensatz zu anderen Wissensrepräsentationstechniken haben die Benutzer keine Vorgaben oder Richtlinien, die sie beim Tagging einzuhalten haben (Peters, 2009, S. 153). Tags sind also freiwählbare Schlagwörter, welche eine Ressource beschreiben. Da eine Ressource zu mehreren Kategorien gehören kann, können unterschiedliche Tags einem Objekt angehängt werden. So kann beispielsweise ein Video einer E-Government Tagung mit Tags wie „Konferenz“, „E-Government“, „Wirtschaftsinformatik“ usw. versehen werden. Dies sagt aus, dass das Video unter anderem zu diesen Kategorien gehört bzw. diese Themen behandelt.

Die Gesamtzahl bzw. die Sammlung der Tags eines Systems oder einer Plattform bezeichnet man als eine Folksonomie (Alby, 2008, S. 127).

Wie Euzenat & Shvaiko richtig andeuten, handelt es sich bei Folksonomien um eine sehr simple Wissensrepräsentation durch die Benennung von Objekten (2007, S. 30). Bei Tags von unterschiedlichen Benutzern ist es nicht möglich, Beziehungen der einzelnen Tags untereinander darzustellen oder zu modellieren, denn laut Brooks & Montanez dürfen Benutzer keine Angaben zu Beziehungen von Tags machen, wie zum Beispiel, dass zwei Tags zusammengehören oder in einer Teil-Ganzes-Beziehung zueinander stehen. Dies führt dazu, dass Tagging alleine nicht ausreichend ist, um Dokumente zu beschreiben und lesbar für Maschinen zu machen (Brooks & Montanez, 2006).

Darüber hinaus besitzen Folksonomien keine Hierarchie, da alle Tags an sich gleich wichtig sind. Eine Art Ranking entsteht nur dadurch, dass populäre Tags, also die, die oft vorkommen, hervorgehoben werden. So hängt laut Peters die Bedeutung eines Tags und dadurch auch der Ressource davon ab, wie oft ein Tag durch verschiedene Benutzer an eine Ressource angehängt wird (2009, S. 160). Dies kann dann beispielsweise durch eine Tag Cloud visualisiert werden, in der die wichtigen Tags durch eine größere Schrift gekennzeichnet sind (Abbildung 5-2).



Abbildung 5-2: TagCloud E-Government Forschungslandkarte (Entnommen aus Projekthandbuch des Projektpraktikums Forschungslandkarte E-Government 2010)

Ein Problem, das beim Tagging oft eintritt, ist die inkohärente Schreibweise von gleich bedeutenden Tags. So können oft inhaltlich gleiche Tags unterschiedliche Schreibweisen haben. Beispielsweise kann ein Nutzer das Wort „E-Government“ als Tag eingeben, während ein anderer die Schreibweise „eGovernment“ verwendet. Dies würde aber dazu führen, dass zwei unterschiedliche Tags gespeichert werden, was aber nicht sein soll, da eigentlich das Gleiche gemeint ist. Um diesem Problem entgegenzuwirken, soll man für das Tagging sogenannte „Tag Recommender Systems“ verwenden. Dabei bekommt der Benutzer während des Taggingprozesses Empfehlungen vom System über die Schreibweise des Wortes, das er eingeben möchte. Auf diese Weise werden doppelte Einträge von Tags vermieden, ohne dass irgendein inhaltlicher Einfluss auf die Benutzer ausgeübt wird (Peters, 2009, S. 204).

Tagging und Folksonmies mögen zwar eine relativ beschränkte Art der Wissensrepräsentation darstellen, trotzdem bringen diese einige wichtige Vorteile für die Indexierung von elektronischen Ressourcen mit sich.

Tagging kann aufgrund der Einfachheit des Prozesses intuitiv durch die Nutzer durchgeführt werden. Dies führt dazu, dass Ressourcen aus Sicht vieler Benutzer beschrieben werden, was letztendlich einen informationellen Mehrwert für die beschriebene Ressource mit sich bringt. Diesen hohen Informationsgehalt ermöglicht die Zugänglichkeit und das Auffinden der Ressource anhand der angehängten Tags, sowie eine Brücke zwischen zwei Ressourcen, wenn diese mit den gleichen Tags versehen wurden und kann somit als eine solide Grundlage für die spezifische und erweiterte Suche von Ressourcen innerhalb eines Systems eingesetzt werden. Darüber

hinaus ist Tagging im Gegensatz zu der Indexierung von Dokumenten durch Experten sehr kostengünstig, und somit laut Peters geeignet für die Indexierung von Masseninformatoren im Internet (2009, S. 218).

5.5.2 Tagging innerhalb der Forschungslandkarte

Bei der Konzeptionierung der Forschungslandkarte E-Government wurde dem Taggen eine hohe Bedeutung eingeräumt. Benutzer werden in der Lage sein, alle Instanzen mit Tags zu versehen, die den Inhalt der Instanzen beschreiben. Scrollt man dann beispielsweise über eine Instanz auf der Forschungslandkarte, werden nicht nur die Informationen über diese Instanz angezeigt, sondern auch die Tags in Form einer Tag Cloud.

Mittels der Tagging-Funktion sollen wichtige Funktionalitäten der Forschungslandkarte realisiert werden. Zum einen ermöglicht die Tagging Funktion das genaue Auffinden von Instanzen wie beispielsweise Instituten, Projekte oder Publikationen anhand der angehängten Tags. Die Instanzen können auf zwei verschiedene Arten gefunden werden: Über die Suchfunktion oder über die Tag Cloud. In der einfachen Suchfunktion kann ein Nutzer ein Stichwort eingeben, z. B. „eGovernment“. Nun werden unter anderem alle Instanzen auf der Karte angezeigt, die mit diesem Tag versehen wurden. Eine weitere Möglichkeit der Suche wird über die Tag Cloud, die neben der Karte platziert ist, realisiert. In dieser sind die vom Benutzer eingegebenen Tags visualisiert und die populären Tags über eine größere Schrift hervorgehoben. Ein Benutzer kann dann durch die verschiedenen Tags navigieren und somit sich die für ihn interessanten Informationen selektieren.

Auch die Funktion der erweiterten Suche wird erst durch das Taggen von Ressourcen möglich gemacht. Bei der erweiterten Suche können Nutzer, abgesehen von freiwählbaren Stichwörtern, auch bestimmte Attribute eingeben. So kann der Nutzer beispielsweise gezielt nach der Kontaktperson eines Instituts suchen oder aber nach Instanzen mit bestimmten Forschungsfeldern. Aufgrund der Bedeutung der Forschungsfelder für das Konzept der Forschungslandkarte sind diese gesondert zu betrachten.

E-Government ist ein sehr umfassender Begriff, der in viele unterschiedliche Teilbereiche bzw. Forschungsfelder aufgeteilt werden kann. Um die Reichhaltigkeit des Begriffes möglichst vollständig abzudecken, hat man sich entschieden, Forschungsfelder anhand von Tags in das System einzubeziehen. So werden manche

Metadaten

Themenfelder vom System vorgegeben, alle anderen sollen dann von den Nutzern selbst eingefügt werden. Hierbei haben die Nutzer die Möglichkeit, alle Instanzen mit Forschungsfeldern zu taggen. Diese können entweder aus einer Liste in Form eines Dropdown-Menüs ausgewählt oder durch eine Freitexteingabe eingegeben werden. So werden die Instanzen mit Forschungsfeldern in Form von Tags verknüpft. Bei der erweiterten Suche kann man dann Inhalte anhand der Forschungsfelder suchen und sich diese anschließend auf der Karte anzeigen lassen.

Metadaten haben das Ziel, Ressourcen im Internet semantisch zu ergänzen, um so die Auffindbarkeit und Interoperabilität der Inhalte zu gewährleisten. Content mit Semantik zu versehen ist von großer Bedeutung, um Inhalte interpretierbar für Maschinen zu gestalten. Voraussetzung hierfür ist, das System auf eine strukturierte Datenebene aufzubauen. Hierzu stehen unterschiedliche Technologien aus dem Bereich des Semantic Webs zur Verfügung. Das nächste Kapitel gibt eine kurze Einführung in die Prinzipien des Semantic Webs und Ontologien.

6 Semantic Web Technologies

6.1 Einführung Semantisches Web

Semantisches Web ist eine Erweiterung des World Wide Webs. Im Gegensatz zum World Wide Web, das auf eine Weise entwickelt wurde, die Menschen den Zugang und die Verarbeitung von Informationen ermöglichen soll, ist das Ziel des Semantischen Web, Informationen mit einer Semantik zu versehen, die diese maschinenlesbar und –interpretierbar machen (Domingue, Fensel, & Hendler, 2011). Auf diese Weise können Systeme automatisierte Aufgaben übernehmen wie die Wiederverwendung von Informationen oder das automatisierte Beziehen von Informationen aus fremden Quellen, was die Mensch-Computer Interaktion verbessert und unterstützt. Das Semantische Web soll letztendlich Informationen mit einer maschinenlesbaren Bedeutung versehen und diese Informationen verlinken, sodass ein universales Medium für den Austausch von Daten und Informationen entsteht (Kashyap, Bussler, & Moran, 2008, S. 3).

Die Inhalte, die den Nutzern über das Semantische Web zur Verfügung gestellt werden, bestehen laut Kashyap, Bussler, & Moran aus strukturierten oder unstrukturierten Daten, welche mit domainspezifischen Metadaten versehen werden, die sie bezogen auf einen Wissensbereich beschreiben (2008, S. 6). Diese Metadatenannotationen werden aus einem Vokabular zusammengesetzt, das durch eine Ontologie für einen bestimmten Wissensbereich festgelegt ist. Dabei beschreibt die Ontologie die Entitäten innerhalb des Wissensbereichs und ihre Beziehungen zueinander. Im Folgenden sollen Ontologien näher betrachtet werden, da sie eine tragende Rolle für das Verlinken und Verteilen von Informationen über das Semantische Web spielen.

6.2 Ontologien

6.2.1 Wissensrepräsentation

Wissensrepräsentation ist ein Bereich der künstlichen Intelligenz. Künstliche Intelligenz zielt darauf ab, die menschliche Intelligenz so auf einem Computer nachzubilden, dass dieser in der Lage ist, selbständig komplexe Funktionen auszuführen und Probleme zu lösen. Dabei wird Wissen über eine bestimmte Problemstellung durch das System gesammelt, ausgewertet und verarbeitet, um ähnliche Ergebnisse wie durch menschliche logische Schlussfolgerung zu erzielen (vgl. Gašević, Djurić, & Devedžić,

2009, S. 3). Im Gegensatz zu den Menschen sind Computersysteme laut Gašević, Djurić, & Devedžić nicht in der Lage, sich das benötigte Wissen durch selbständiges Lernen zu beschaffen (2009, S. 4). Dies hat zur Folge, dass das Wissen durch Menschen in das System hinzugefügt werden muss. Anhand von Wissensrepräsentationstechniken, die das Wissen in einer Weise repräsentieren, die maschinenlesbar ist, sodass es durch das System verarbeitet werden kann, wird dies möglich gemacht.

Durch die Wissensrepräsentation ist das System in der Lage, Fakten und Konzepte über einen bestimmten Wissensbereich und deren Beziehungen zueinander zu erkennen und diese so einzusetzen und zu manipulieren, dass bestimmte Probleme gelöst werden können (Gašević, Djurić, & Devedžić, 2009, S. 3). So werden durch Wissensrepräsentation Entitäten der realen Welt innerhalb des Systems abgebildet und die Interaktion dieser zueinander anhand von Regeln, Fakten etc. modelliert.

Es existieren viele verschiedene Methoden der Wissensrepräsentation wie z. B. Ontologien, Thesauri oder Folksonomien. Diese unterscheiden sich in ihren Eigenschaften wie beispielsweise der Ausdruckstärke und Komplexität.

Abbildung 6-1 zeigt die Abhängigkeit der Ausdruckstärke dieser Methoden zur Größe des Wissensbereiches, die sie abdecken können. Einfache Wissensrepräsentationsmethoden wie Folksonomien (Kapitelabschnitt 5.4.1), können aufgrund ihrer Einfachheit auf sehr große Wissensbereiche angewendet werden. Im Gegensatz dazu eignen sich komplexe und ausdrucksstarke Methoden wie Ontologien nur für kleine Wissensbereiche, da die Anzahl der semantischen Beziehungen, die sie beschreiben, mit der Größe des Wissensbereichs wächst.

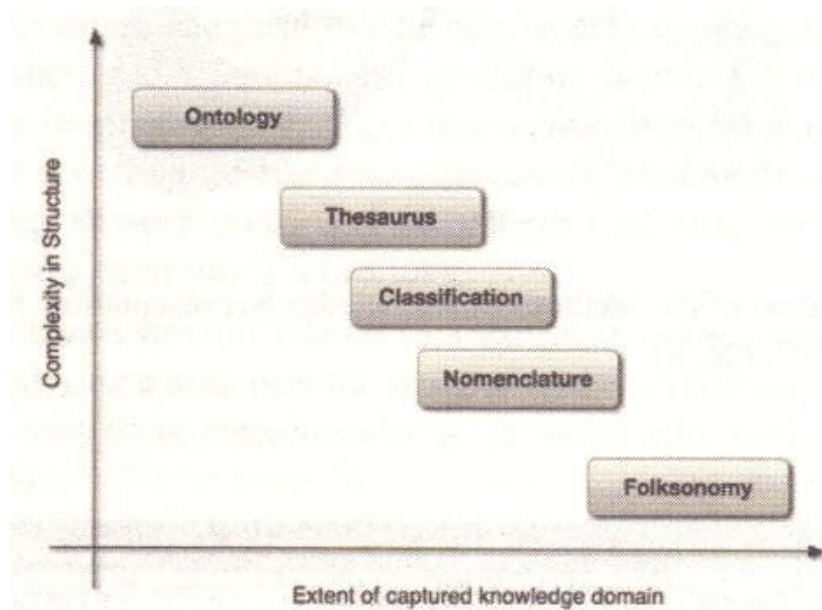


Abbildung 6-1: Classification of the Methods of Knowledge Representation with regard to Expressiveness and Coverage of the Knowledge Domain. Source: Adapted from Peters & Weller (2008, 101, Fig. 1). (Peters, 2009)

6.2.2 Ontologien

Wie in Kapitelabschnitt 6.2.1 ausgeführt, gibt es eine hohe Anzahl an unterschiedlichen Wissensrepräsentationsmethoden. Nun werden Ontologien behandelt. Diese Methode der Wissensrepräsentation stellt ein Modell zur Verfügung, mit dem es möglich ist, semantisch reichhaltiges Wissen zu repräsentieren. Darüber hinaus kann eine Ontologie anhand von formalen Wissensrepräsentationssprachen interoperabel gestaltet werden, sodass diese mit anderen Ontologien verknüpft werden kann (Breslin, Passant, & Vrandecic, 2011). Diese Idee von „shared“-Wissen ist eine der Prinzipien des Semantic Web.

6.2.2.1 Begriffserklärung und Darstellung von Ontologien

Das Wort Ontologie in seiner ursprünglichen Form kommt aus der griechischen Sprache und ist ein Begriff bzw. eine Disziplin der Philosophie, die sich mit Forschungsfragen der Existenz bzw. des Daseins befasst (Gašević, Djurić, & Devedžić, 2009, S. 45). Dabei geht es laut Grimm et al. um die Fragen „welche Sachen gibt es?“ und „zu welchen Kategorien gehören die Sachen, die existieren?“ (2011). Eine Ontologie beschränkt sich dabei immer auf eine bestimmte Domain, also einen bestimmten

Bereich, der beschrieben wird. Eine Ontologie beschreibt also die Sachen, die innerhalb dieser Domain existieren.

In der Informationstechnik werden Ontologien dazu genutzt, um Informationen im Netz zu beschreiben und zu modellieren. Laut Grimm et al. hat die Idee der ontologischen Kategorisierung einer Domain Vorteile für Informationssysteme (2011). In der Wissensrepräsentation werden Ontologien für die maschinenlesbare Repräsentation einer Domain anhand ihrer Begrifflichkeiten (z. B. Klassen, etc.) und deren Beziehungen zueinander (Peters, 2009, S. 125) verwendet. Auf diese Weise definiert eine Ontologie eine Menge an Begrifflichkeiten, die dazu genutzt werden, um einen bestimmten Wissensbereich zu beschreiben. Eine genauere Definition liefert Tom Gruber:

„Ontologies are specifications of the conceptualizations at a semantic level“⁶

Diese Definition beinhaltet die verschiedenen Eigenschaften von Ontologien und soll im folgenden Absatz näher betrachtet und in ihre Einzelteile zerlegt werden.

Eine Ontologie ist laut dem ersten Teil der Definition von Gruber eine **Spezifikation**. Dies bedeutet, dass Ontologien durch eine formale Sprache beschrieben werden müssen, da dies der einzige Weg ist, um das Wissen über die beschriebene Domain maschinenlesbar zu gestalten. Nur so sind Informationssysteme in der Lage, auf das repräsentierte Wissen zuzugreifen und dieses zu interpretieren (Peters, 2009, S. 125).

Eine weitere Eigenschaft von Ontologien ist die der **Konzeptualisierung**. Laut Gašević, Djurić, & Devedžić sollen Ontologien eine abstrakte, vereinfachte und allgemein gehaltene Sicht auf die Welt ermöglichen (2009, S. 46). Das Ziel ist also nicht eine spezielle Situation in der Welt darzustellen, sondern möglichst alle Situationen und Zustände abzudecken, die vorkommen könnten.

Darüber hinaus gibt es laut Grimm et al. weitere Eigenschaften von Ontologien:

⁶ Gruber, T. (2005). *Ontology of Folksonomy: A Mash-Up of Apples and Oranges*, <http://tomgruber.org/writing/ontology-of-folksonomy.htm>

Ontologien sollen nur **explizites Wissen** darstellen, da nicht-explizites Wissen nicht maschinenlesbar ist und daher nicht von Informationssystemen interpretiert werden kann.

Ontologien repräsentieren nicht das Wissen eines Individuums, sondern objektives Wissen, welches als das Wissen über eine Domain, von einer Gruppe von Menschen oder einer Gesellschaft akzeptiert wird. Ein solcher **Kompromiss über eine Konzeptualisierung (Konsens)** des Wissens einer Domain ist eine der Grundlagen für die Interoperabilität und Wiederverwendung von Wissen (Grimm et al., 2011).

Gašević, Djurić, & Devedžić verdeutlichen den Ontologiebegriff anhand eines einfachen Beispiels. Um den Wissensbereich bzw. die Domain eines Musikers zu beschreiben, benötigt man folgende Begrifflichkeiten, die in der Musiker-Ontologie festgelegt wurden: Musiker, Instrument, Alben, die der Musiker produziert hat, Events und Fans. Um dem Aspekt der Konzeptualisierung gerecht zu werden, wurden die Beziehungen zwischen den Begriffen auf die wichtigsten reduziert (Gašević, Djurić, & Devedžić, 2009, S. 48).

Je nach Zweck und Abstraktionsstufe kann man die Ontologie in unterschiedlichen Weisen darstellen. Den Autoren zufolge kann man eine Ontologie anhand von deklarativen Statements in natürlicher Sprache ausdrücken. Basierend auf einer natürlichsprachlichen Darstellung lässt sich die Ontologie grafisch wiedergeben (Abbildung 6-2).

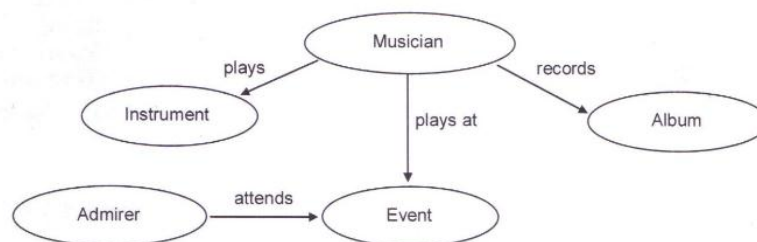


Abbildung 6-2: Musician ontology visualized as a semantic network (Gašević, Djurić, & Devedžić, 2009)

Die Abbildung 6-2 zeigt eine sehr vereinfachte Sicht auf die Musiker-Ontologie, ohne auf die Attribute der Objekte oder auf die Eigenschaften der Beziehungen zwischen den Objekten einzugehen. Da diese Art der grafischen Darstellung von Ontologien nicht formal, sondern natürlichsprachlich ist, ist sie laut Gašević, Djurić, & Devedžić für einen Computer nur schwer zu interpretieren (2009, S. 49). Möchte man eine detaillierte grafische Sicht auf die Ontologie erreichen, kann man diese beispielsweise in UML

modellieren und dabei Attribute der Objekte, der Beziehungen etc. näher beschreiben. Diese Darstellung ist etwas formaler als die der Abbildung 6-2 und ist aufgrund ihrer grafischen Form (UML-Diagramm) trotzdem für den Menschen verständlich.

Um den formalen Aspekt zu gewährleisten, ist die Ontologie im Computer in einer formalen Sprache abgebildet (Abbildung 6-3). Diese auf XML-basierende OWL-Sprache ist maschinenlesbar und wird in Kapitelabschnitt 6.2.2.2 näher betrachtet.

```
<owl:Class rdf:ID="Event"/>
<owl:Class rdf:ID="Album"/>
<owl:Class rdf:ID="Instrument"/>
<owl:Class rdf:ID="Musician"/>
<owl:Class rdf:ID="Admirer"/>
<owl:ObjectProperty rdf:ID="author">
  <owl:inverseOf>
    <owl:ObjectProperty rdf:ID="opus"/>
  </owl:inverseOf>
  <rdfs:domain rdf:resource="#Album"/>
  <rdfs:range rdf:resource="#Musician"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="player">
  <rdfs:range rdf:resource="#Musician"/>
  <rdfs:domain rdf:resource="#Instrument"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="loudness">
  <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  <rdfs:domain rdf:resource="#Instrument"/>
</owl:ObjectProperty>
...
```

**Abbildung 6-3: The Musician ontology represented in OWL [excerpt]
(Gašević, Djurić, & Devedžić, 2009)**

6.2.2.2 Formale Sprachen für Ontologien

Ontologiesprachen sind formale Sprachen zur Beschreibung von Ontologien. Laut Grimm et al. beschreiben diese nicht nur die Eigenschaften von Ontologien im Sinne der Wissensrepräsentation, sondern stellen notwendige Web-Funktionen wie die Identifikation von Ressourcen mittels URIs zur Verfügung (2011).

Im Rahmen der Semantic-Web-Forschung wurden in den letzten Jahren verschiedene Ontologiesprachen entwickelt, die sich unter anderem in ihrer Ausdrucksstärke unterscheiden. Trotzdem haben diese laut Euzenat & Shvaiko gemeinsame Bestandteile und sind ähnlich aufgebaut (2007, S. 36). Ontologiesprachen bestehen meistens aus folgenden Entitäten:

Die Entitäten einer Domain werden als **Klassen** dargestellt. Im Falle der Forschungslandkarte sind die Klassen die Entitäten, die auf der Karte angezeigt werden, also Institute, Personen, Projekte etc.

Individuen sind die Instanzen einer Entität innerhalb einer Domain. Die Individuen der Forschungslandkarte sind beispielsweise die Universität Koblenz oder eine bestimmte Person, Projekt, Konferenz oder Publikation.

Ein weiterer Bestandteil von Ontologiesprachen sind **Beziehungen**, die zwischen den Instanzen der Domain bestehen. Hiermit können Beziehungen wie Person „arbeitet in“ Institut oder Institut „beteiligt an“ Projekt modelliert werden.

Darüber hinaus können mit einer formalen Sprache **Datentypen**, wie Integer oder String und ihre zugehörigen **Werte** beschrieben werden.

Laut Kashyap, Bussler, & Moran existieren mehrere unterschiedliche Standards für die Wissensrepräsentation. Die am meisten verbreiteten Ontologiesprachen sind RDF Schema, Web Ontology Language (OWL) und die Web Services Modeling Language (WSML) (Kashyap, Bussler, & Moran, 2008, S. 84).

Die Web Ontology Language ist die bekannteste Repräsentationssprache für Ontologien im Internet und hat drei unterschiedliche Versionen: OWL-Lite, OWL-DL und OWL-Full. Diese unterscheiden sich in ihrer Expressivität und sind für die Nutzung in verschiedenen Communities entwickelt worden (Kashyap, Bussler, & Moran, 2008, S. 100).

6.2.2.3 Funktionen von Ontologien

Ontologien werden heutzutage aufgrund der Vorteile für intelligente Systeme, die sie mit sich bringen, immer öfter in der Informationstechnik eingesetzt. Dabei können Ontologien verschiedene Funktionen erfüllen:

Wie bereits erwähnt bestimmt eine Ontologie laut Euzenat & Shvaiko das **Vokabular** einer bestimmten Domain (2007, S. 29). Nicht nur werden die Begriffe bzw. die Namen der Objekte innerhalb der Domain festgelegt, sondern auch deren Beziehung zueinander beschrieben. Darüber hinaus wird laut Gašević, Djurić, & Devedžić anhand von Regeln und logischen Aussagen festgelegt, wie die Objekte miteinander kombiniert werden können, sodass das Vokabular erweitert werden kann (2009, S. 51).

Eine Ontologie beschreibt anhand eines Vokabulars eine bestimmte Domain und stellt somit einer Applikation oder einem Agenten das Wissen über die Domain zur Verfügung. Eine wichtige Funktion, die mit Hilfe von Ontologien realisiert wird, ist das **Teilen und Wiederverwenden von Wissen** durch unterschiedliche Applikationen oder Agenten. Wird das Wissen zwischen Applikationen oder intelligenten Agenten geteilt, entstehen laut Gašević, Djurić, & Devedžić „shared Ontologies“. Diese verteilten Ontologien bilden eine einheitliche Rahmenstruktur, mit der spezifische Situationen beschrieben und modelliert werden können (Gašević, Djurić, & Devedžić, 2009, S. 54). Allerdings ist es ein recht komplexes Unterfangen, Wissen über verschiedene Applikationen oder Agenten zu verteilen. Verschiedene Systeme verfügen oft über unterschiedliche Arten der Wissensrepräsentation wie zum Beispiel verschiedene formale Sprachen oder Modellierungsmethoden. Diese Heterogenität macht Wissen über unterschiedliche und sogar dieselben Domains inkompatibel und erschwert somit den Prozess der Wissensteilung und Wiederverwendung. Den Autoren zufolge kann diese Heterogenität mit Hilfe von Ontologien reduziert werden, was den Wissensteilungsprozess enorm vereinfacht. Basierend auf Ontologien kann nämlich eine auf einer formalen Sprache basierenden Beschreibung eines Wissensbereichs entwickelt werden. Dieses formale Domainwissen kann an andere Applikationen und Agenten weitergegeben und als eine Rahmenstruktur wiederverwendet werden. Dies würde den Aufbau einer verwandten Ontologie aus diesem Wissensbereich erleichtern, da die Rahmenstruktur bereits das Grundvokabular und Informationen über die Beziehungen der Objekte enthält. Außerdem wird eine einheitliche Syntax zur Verfügung gestellt, um Wissen über diese Domain zu modellieren (Gašević, Djurić, & Devedžić, 2009, S. 55).

Darüber hinaus können Ontologien, die auf der vorgestellten verteilten Rahmenstruktur basieren, leichter verknüpft werden. Auf diese Weise wird laut Grimm et al. die **Integration von Informationen und Wissen** aus fremden Quellen ermöglicht (2011).

Durch den formalen Aufbau von Ontologien sind intelligente Systeme in der Lage, explizites Wissen über eine Domain zu interpretieren und daraus logische Schlussfolgerungen zu ziehen. Somit ist das System in der Lage, implizites Wissen aus explizitem herzuleiten. Diesen Prozess nennt man **Reasoning**. Anhand dessen kann laut Grimm et al. ein gewisser Grad an Automatisierung erreicht, neue Funktionalitäten

aus dem verfügbaren Wissen gewonnen und durch den Benutzer generiertes Wissen überprüft und verifiziert werden (2011).

Ontologien unterstützen Systeme bei der Suche von Informationen im Internet. So können anhand von Ontologien, Metadaten und anderen semantischen Technologien, Informationen aus einer oder mehreren Quellen effektiv gefunden werden. Somit bilden Ontologien die Grundlage für eine erfolgreiche **Sematische Suche** (Grimm et al. 2011).

6.2.3 Zusammenfassung

Ontologien beschreiben einen Wissensbereich und legen für diesen ein Vokabular fest. Im Falle der Forschungslandkarte soll eine Ontologie als Grundlage entwickelt werden, um die Entitäten und deren Beziehungen zueinander zu modellieren und dieses Wissen anhand einer formalen Sprache maschinenlesbar zu gestalten. Basierend darauf kann die temporale Modellierung der Entitäten durchgeführt werden, was die Historisierung der Inhalte ermöglichen würde.

Darüber hinaus können bestimmte optionale Funktionen ermöglicht werden, die ohne Ontologien und Semantisches Wissen nicht umgesetzt werden können. Anhand von Ontologien kann beispielsweise eine semantische Suche realisiert werden, um die Inhalte der Forschungslandkarte effizient zu finden. Der formale Aufbau einer Ontologie gilt außerdem als Voraussetzung für ein automatisiertes Handeln durch das System, beispielsweise um neues Wissen zu integrieren oder den Wissensbereich der Forschungslandkarte mit anderen bestehenden Ontologien zu verbinden.

Nachdem die Grundlagen für die Historisierung der Inhalte der Forschungslandkarte gezeigt wurden, wird im nächsten Teil der Arbeit auf das Problem der Historisierung eingegangen und ein Lösungskonzept entwickelt. Zuerst werden die Anforderungen an das Content Management System erfasst und das CMS Drupal als Grundlage für das System vorgeschlagen.

Teil 3: Herausforderungen und Lösungskonzept für die Historisierung der Forschungslandkarte

7 Content Management Systeme als Grundlage für Forschungslandkarte

7.1 Anforderungen an Content Management Systeme im Kontext der Forschungslandkarte E-Government

Im Kapitel 3 wurden anhand der wichtigsten Begrifflichkeiten Content Management Systeme und Web Content Management Systeme in ihrer allgemeinen Form vorgestellt. Nun sollen die Anforderungen an ein solches System dargestellt werden, die durch den Kontext der Forschungslandkarte E-Government entstehen. Da es sich bei der Forschungslandkarte um eine Online-Plattform handelt, muss diese mit einem Web Content Management System realisiert und verwaltet werden. Um die Wahl auf einen bestimmten WCMS treffen zu können, muss erst erörtert werden, welche Funktionen durch das System unterstützt werden müssen bzw. sollen, denn nur so kann gewährleistet werden, dass das System die benötigten Funktionalitäten aufweist.

Als allerwichtigste Ressource der Forschungslandkarte gelten die Inhalte, also die einzelnen Instanzen innerhalb des Systems. Diese sollen möglichst durch die Nutzer selbst erstellt werden. Nutzer müssen also in der Lage sein, Instanzen zu den Entitäten aus dem Klassendiagramm aus Kapitelabschnitt 2.2 erstellen zu können. Die **Instanzerstellung** muss über Formulare oder Masken abgewickelt werden, die den Nutzern erlauben, die Informationen über die Instanz anhand der festgelegten Attribute festzulegen. Die Projektmaske bietet beispielsweise anhand von Pflicht- und Optionalfeldern die Möglichkeit, den Namen, eine kleine Beschreibung, Kontaktdaten sowie Start- und Enddatum des Projekts einzugeben. Darüber hinaus soll ein Eingabefeld „**Tagging**“ angeboten werden, über das Nutzer Tags bei der Instanzerstellung an die Objekte anhängen können. Nach dem Bestätigen wird die Instanz als Objekt im System gespeichert. Abgesehen von der Instanzerstellung muss das System die Modellierung der Beziehungen zwischen den verschiedenen Instanzen erlauben. So muss es möglich sein, die einzelnen **Instanzen durch Verlinkungen zu verknüpfen**. Es sollen dabei nur die durch das Klassendiagramm aus 2.2

vorgesehenen Verlinkungen möglich sein. Eine Verknüpfung zwischen einem Projekt und einer Konferenz beispielsweise ist nicht vorgesehen und darf vom System nicht erlaubt werden.

Eine weitere Funktionalität, die unbedingt durch das System unterstützt werden muss, ist die **Rollen- bzw. Nutzerverwaltung**. Dies ist unbedingt notwendig, da es verschiedene Arten von Nutzern gibt, die unterschiedliche Rechte haben. So wird zwischen Gastnutzer, angemeldetem Nutzer, verifiziertem Nutzer und Administratoren unterschieden. Um dies zu ermöglichen, müssen die Benutzer in der Lage sein, sich im System zu registrieren und ein Konto mit ihren Daten zu erstellen. Daraufhin ist es möglich, anhand von Rollen zwischen den verschiedenen Nutzern zu unterscheiden. Diese Rollen werden den einzelnen Nutzern zugeordnet und fassen die Rechte zusammen, die die Nutzer innerhalb des Systems haben. Gäste, also diejenigen Nutzer, die nicht im System registriert sind, haben nur die Möglichkeit, die Inhalte der Forschungslandkarte anzusehen, diese aber nicht zu manipulieren oder selbst Instanzen zu erstellen. Darüber hinaus haben sie das Recht, die Forschungslandkarte über die Such-/Filterfunktion und die erweiterte Suchfunktion durchzusuchen, Instanzen mit Tags zu versehen und sich im System zu registrieren. Angemeldeten Nutzern stehen, abgesehen von den zuvor genannten Basisfunktionen, noch weitere Funktionalitäten zur Verfügung. So haben sie die Möglichkeit ihr Profil zu bearbeiten, eigene Instanzen zu erstellen und diese zu bearbeiten, sich mit Instanzen zu verlinken, Projekte zu kommentieren und Fehler oder rechtswidrige Inhalte zu melden. Darüber hinaus können sie sich als Angehörige eines Instituts verifizieren lassen. Diese verifizierten Nutzer haben dann die Möglichkeit, die Informationen des Instituts, welchem sie angehören, zu bearbeiten. Um die Korrektheit der Inhalte zu kontrollieren und zu gewährleisten, muss eine weitere Art von Nutzern existieren - Administratoren. Diese haben, abgesehen von allen zuvor genannten Funktionalitäten, darüber hinaus die Möglichkeit, fremde Instanzen zu bearbeiten und gegebenenfalls auch zu entfernen. Diese Möglichkeiten, plus das Manipulieren weiterer administrativen Einstellungen aus dem Editorial System, soll nur wenigen berechtigten Personen bzw. Mitarbeitern zur Verfügung stehen.

Des Weiteren stellen die erwünschten Funktionalitäten der Forschungslandkarte weitere Anforderungen an das System dar. So muss den Nutzern die Möglichkeit gegeben werden, anhand einer **Filterfunktion** die Inhalte bzw. die gewünschten Entitäten auf der Forschungslandkarte zu selektieren. Auch eine **Volltext-Suche und die erweiterte**

Suche anhand bestimmter Attribute wie z.B. Forschungsfelder, muss vom System unterstützt werden.

Während der Konzeptionierung der Forschungslandkarte wurde die Möglichkeit besprochen, ein **Forum** für Diskussionen bereitzustellen. Dies soll gegebenenfalls durch das System umgesetzt werden können.

Optional könnten auch Maßnahmen zur Datensicherung getroffen werden. Dies würde bedeuten, dass eine **Rollbackfunktion**, basierend auf der **Versionierung** von Inhalten, durch das Web Content Management System geboten wird.

Tabelle 1 zeigt noch einmal die Anforderungen der Forschungslandkarte an das Web Content Management System. Diese werden darüber hinaus in Pflicht- und optionale Anforderungen unterteilt.

Anforderung/Funktionalität	Pflicht/Optional
Instanzerstellung anhand von Masken	Pflicht
Umsetzung der Beziehungen von Entitäten durch Verlinkungen	Pflicht
Tagging	Pflicht
Rollen- bzw. Nutzerverwaltung	Pflicht
Filterfunktion, Volltext-Suche und Erweiterte Suche anhand bestimmter Attribute	Pflicht
Forum	Optional
Versionierung und Rollbackfunktion	Optional

Tabelle 1: Anforderungen an das Web Content Management System im Kontext der Forschungslandkarte E-Government

7.2 Eignung von Drupal für Forschungslandkarte

In Kapitelabschnitt 7.1 wurde eine Liste der Anforderungen aufgestellt, die das Web Content Management System berücksichtigen muss, um die Funktionalitäten der Forschungslandkarte E-Government realisieren zu können. Wie bereits erwähnt wurde während der Konzeptionierung des Projekts das WCMS Drupal als erste Wahl für die Implementierung der Forschungslandkarte festgelegt. Hier soll das System Drupal kurz

vorgestellt und die wichtigsten Gründe dargelegt werden, die zu dieser Entscheidung geführt haben. Da dies bereits während der Konzeptionierung der Forschungslandkarte ausführlich ausgearbeitet wurde und eher zweitrangig in Bezug auf das Thema dieser Arbeit ist, werden die wichtigsten Kriterien, die für Drupal sprechen, kurz zusammengefasst.

Drupal ist ein Content Management System, das zur Erstellung und Verwaltung von Webseiten dient. Dabei wird Drupal laut Van Dyk als ein Open-Source-Framework verstanden (2009, S. 33). Somit besteht das System aus „freie[r] Software, entwickelt durch die Drupal Community, die sich aus tausenden Entwicklern und Firmen zusammensetzt“⁷. Aufgrund der Community, die sich um Drupal gebildet hat, ist es möglich, kommerziellen sowie nicht kommerziellen Support zu erhalten.

Die Basisversion von Drupal besteht aus dem Core. Dieser enthält die Grundfunktionalitäten des Systems (Abbildung 7-1).

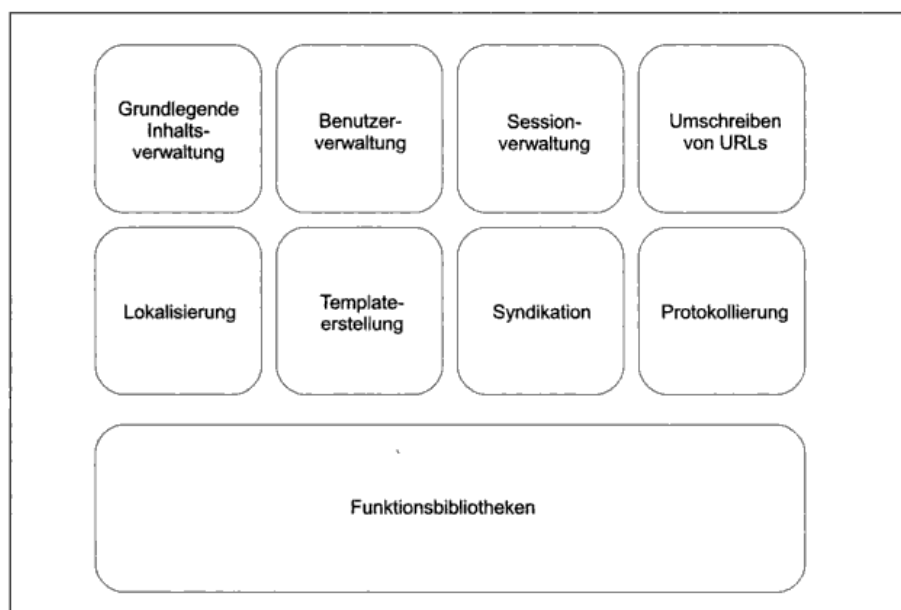


Abbildung 7-1: Überblick über den Drupal-Core (Van Dyk, 2009)

Weitere benötigte Funktionalitäten werden über sogenannte Module realisiert. Diese können beliebig aktiviert und deaktiviert werden und sind integriert oder werden von Drittanbietern angeboten. So lässt sich laut Van Dyk das System erweitern, ohne Änderungen am Code des Cores vornehmen zu müssen (2009, S. 33). Diese Modularität hat den Vorteil, dass einfache Webseiten sich mit einem relativ schlanken

⁷ www.drupal.de

Framework realisieren lassen. Sind weitere Funktionalitäten nötig, kann man diese ohne Probleme hinzufügen (vgl. Van Dyk, 2009, S. 36).

Um die Benutzbarkeit des Systems zu wahren, wird ein besonderes Augenmerk auf sauberen Code und die Etablierung von Standards gesetzt, des Weiteren wird die Sicherheit des Systems ständig durch ein Team überprüft.

Nun soll überprüft werden, inwieweit die Anforderungen bzw. Funktionalitäten aus Kapitelabschnitt 7.1 durch Drupal unterstützt werden.

Instanzerstellung und Verlinkung:

Inhalte (oder „Nodes“) werden in Drupal anhand von Inhaltstypen unterteilt. Im Core sind verschiedene Typen wie z. B. Artikel, Blog, Buch oder Seite enthalten (Graf, 2008, S. 135). Die Unterteilung in Typen hat unter anderem den Vorteil, dass man verschiedenen Benutzern unterschiedliche Rechte an verschiedenen Inhaltstypen zuordnen kann. So kann zum Beispiel erreicht werden, dass Nutzer der Forschungslandkarte nur eigen erstellte Instanzen bearbeiten können (vgl. Graf, 2008, S. 135). Durch die Erweiterung Content Construction Kit (CCK-Modul) ist es dann möglich, Inhaltstypen mit beliebigen weiteren Feldern zu versehen, wodurch sich die Entitäten und deren Attribute unproblematisch erstellen lassen (Graf, 2008, S. 136). Die Beziehungen zwischen den Instanzen lassen sich anhand von Referenzen realisieren. Anhand der Typen „Benutzerreferenz“ und „Beitragsreferenz“ kann man Objekte und Benutzer miteinander verknüpfen (Stöbel, 2010, S. 53). Es ist möglich, Verlinkungen zwischen bestimmten Typen auszuschließen, wodurch sich unvorhergesehene Verknüpfungen vermeiden lassen.

Tagging:

Das Tagging wird in Drupal durch eine Erweiterung des Taxonomie Moduls umgesetzt, mit dem sich Vokabulare erstellen lassen. Diese werden genutzt, um Inhaltstypen zu kategorisieren. Diese Erweiterung erlaubt das Erstellen eines solchen Vokabulars über Free Tagging⁸, also das Taggen der Instanzen bzw. Nodes durch die Benutzer.

⁸ <http://drupal.org/node/19697>

Nutzer- und Rollenverwaltung:

Wie in den meisten Web Content Management Systemen bietet Drupal für Benutzer die Möglichkeit, sich zu registrieren. Dabei erstellen die Benutzer ein eigenes Konto mit einem Benutzernamen, einer E-Mail-Adresse und einem Passwort. Das Konto wird dann durch den Benutzer anhand eines Bestätigungslinks aktiviert (vgl. Graf, 2008, S. 113).

Darüber hinaus erlaubt Drupal das Erstellen von Rollen wie z.B. Gast, Admin und so weiter. Es können dabei beliebig viele Rollen erstellt werden. Man sollte allerdings darauf achten, nur so viele Rollen wie nötig anzulegen, da man sonst schnell den Überblick verliert (Graf, 2008, S. 119). In einem weiteren Schritt werden laut Graf den Rollen Berechtigungen zugeordnet, die festlegen, welche Aktionen die einzelnen Benutzer ausführen dürfen (2008, S. 118). So kann z. B. vermieden werden, dass Gäste Instanzen bearbeiten, angemeldete Nutzer Änderungen an fremden Instanzen vornehmen oder diese löschen.

Filterfunktion, Volltext-Suche und Erweiterte Suche anhand bestimmter Attribute:

Drupal besitzt eine integrierte Suchfunktion, mit der sich alle gewünschten Funktionalitäten der Forschungslandkarte realisieren lassen. Über das Suchmodul (Abbildung 7-2) können Nutzer die Inhalte anhand von Schlüsselwörtern durchsuchen, was der Funktionalität der Volltext-Suche entspricht. Darüber hinaus bietet Drupal eine erweiterte Suchfunktion. Hier soll es möglich sein, Inhalte anhand von attributabhängigen Suchkriterien zu durchsuchen. Insbesondere soll es möglich sein, Instanzen anhand der Forschungsfelder zu finden. Forschungsfelder werden an den Instanzen in Form von Tags angehängt (Kapitelabschnitt 5.5 beschäftigt sich tiefgründig mit diesem Thema). Diese Tags stehen in der erweiterten Suche als Kategorien zur Auswahl. So kann man sich Inhalte anzeigen lassen, die zu diesen Kategorien und somit zu den gewünschten Forschungsfeldern gehören.

Durch die erweiterte Suche sind Nutzer auch in der Lage, einzelne Inhaltstypen zu selektieren. Auf diese Weise kann die Suche so gefiltert werden, dass nur bestimmte Entitäten in der Suche berücksichtigt und angezeigt werden.

The image shows a search interface titled "Suchen". At the top, there are radio buttons for "Inhalt" and "Benutzer". Below this is a text input field labeled "Zu suchende Schlüsselwörter:" with a "Suchen" button to its right. A section titled "Erweiterte Suche" is expanded, showing three input fields: "Beinhaltet eines der Wörter:", "Beinhaltet den Satz:", and "Beinhaltet keines der Wörter:". To the right of these fields is a dropdown menu labeled "Nur in der Kategorie/den Kategorien:" with a list of categories: "Foren", "Foren zum Thema Drupal Rezepte", and "Vorschläge für Drupal Rezepte". Below the search options is a section "Nur vom Typ/von den Typen:" with a list of content types: "Artikel", "Blogbeitrag", "Book page", "Forenbeitrag", "Seite", and "Umfrage", each with a checkbox. At the bottom of the search options is a button labeled "Erweiterte Suche".

Abbildung 7-2: Suchformular (Graf, Drupal 6: Websites entwickeln und verwalten mit dem Open Source-CMS, 2008)

Forum:

Diskussionsforen können in Drupal mit dem „forum-Modul“ erstellt werden. Dabei handelt es sich laut Graf bei einem Forum um einen neuen Inhaltstyp, bestehend aus einem oder mehreren Forumseinträgen und beliebig vielen Antworten zu diesem Eintrag (2006, S. 189).

Versionierung und Rollbackfunktion:

Mit der Funktion „Neue Revision“, die bereits im Core enthalten ist, kann man beim Speichern des Inhalts eine neue Version erstellen. Über einen Reiter ist es dann möglich, alte Versionen des Inhalts wiederherzustellen (Graf, 2008, S. 137).

Das CMS Drupal stellt einige temporale Funktionen wie die Versionierung von Inhalten zur Verfügung. Die Historisierung der Inhalte wird allerdings nicht komplett unterstützt. Im folgenden Kapitel werden die Problemen und Herausforderungen der Historisierung und Aktualisierung näher beschrieben und voneinander abgegrenzt.

8 Herausforderungen im Kontext der E-Government- Forschungslandkarte: Historisierung und Aktualisierung

In diesem Kapitel wird die Problemstellung der Ausarbeitung näher beschrieben. Konkret handelt es sich dabei um die Historisierung und Aktualisierung der Inhalte der Forschungslandkarte E-Government. In den folgenden Abschnitten werden beide Teilprobleme einzeln beschrieben und voneinander abgegrenzt. Dabei soll die Notwendigkeit und Zweck der Historisierung und Aktualisierung im Kontext der Forschungslandkarte aufgezeigt werden. Anhand einer Gewichtung beider Teilprobleme soll die Entscheidung getroffen werden, inwieweit diese in der vorliegenden Ausarbeitung berücksichtigt werden sollen.

8.1 Historisierung der Inhalte

Der Zweck der Forschungslandkarte E-Government ist es, einen gesamtheitlichen Überblick über die unterschiedlichen Akteure und deren Publikationen, Projekte etc. zu bekommen, die sich mit dem Thema E-Government befassen. Dabei spielen nicht nur aktuelle Sachverhalte eine wichtige Rolle. Dem Nutzer muss die Möglichkeit gegeben werden, im Rahmen der Recherche auf Informationen zuzugreifen, die in der Vergangenheit gültig waren. Dies ist die Aufgabe der Historisierung. Wie in Kapitel 4 ausgeführt, ist das Ziel der Historisierung oder temporalen Datenhaltung, Änderungen an Daten zu dokumentieren. Auf diese Weise wird eine Historie erschaffen, die aufzeigt, welche Werte zu einem bestimmten Zeitpunkt in der Vergangenheit gültig waren.

Speziell für eine Forschungslandkarte spielt die Historisierung der Inhalte eine wichtige Rolle. Durch die Historisierung von Inhalten wird ein Datenbestand aufgebaut, indem dokumentiert ist, wie sich die Inhalte der Forschungslandkarte mit der Zeit geändert haben. Dieser Datenbestand bildet die Grundlage für die Recherchefunktion der Forschungslandkarte. Nutzer sind dann in der Lage, sich über eine Filterfunktion Inhalte auf der Karte anzeigen zu lassen, welche zu einem frei wählbaren Zeitpunkt in der Vergangenheit gültig waren. So kann man beispielsweise auf bereits beendete Projekte zugreifen und sich alle Informationen über diese wie z. B. Start- und Enddatum, Beschreibung etc. anzeigen lassen.

Eine Historisierung ist darüber hinaus notwendig, wenn eine Verknüpfung zwischen zwei Instanzen der Forschungslandkarte aktualisiert und somit überschrieben wird. Besteht

Herausforderungen im Kontext der E-Government-Forschungslandkarte: Historisierung und Aktualisierung

ein Interesse, die Beziehung zwischen zwei Instanzen auch nach der Aktualisierung festzuhalten, muss die Verknüpfung historisiert werden. Dies ist beispielsweise der Fall, wenn eine Person das Institut, in dem diese arbeitet, wechselt. Ohne temporale Datenhaltung würde man die Institutsangehörigkeit „überschreiben“, indem man die Verknüpfung zwischen der Person und dem vorherigen Institut löscht und eine neue Verknüpfung zum neuen Institut erstellt. Dies führte dazu, dass die alte Zuordnung zum vorherigen Institut verloren gehen würde. Aus Gründen der Vollständigkeit und um die Recherchefunktion zu ermöglichen, soll allerdings dokumentiert werden, wo die Person früher tätig war. Durch die Historisierung der alten Zuordnungen sind die Nutzer in der Lage, die Historie der Person zu verfolgen. Somit kann festgehalten werden, wo die Person in der Vergangenheit gearbeitet hat, in welchen Forschungsfeldern sie tätig war, an welchen Konferenzen diese teilgenommen hat, an welchen Projekten sie mitgearbeitet hat und an welchen Publikationen diese beteiligt war.

Historisiert man beispielsweise alle Verknüpfungen zu einem Projekt, nachdem es beendet wurde, kann man festhalten, welche Personen an diesem mitgearbeitet haben, welche Institute daran beteiligt waren und welche Themenfelder bearbeitet wurden.

Des Weiteren besteht die Anforderung, einzelne relevante Attribute von Instanzen zu historisieren. Dabei wird nicht wie im Fall eines Projekts die ganze Instanz samt ihrer Attribute historisiert, sondern nur einzelne Attribute. Ein Beispiel hierfür ist das Attribut Forschungsfelder der Instanz Forschungsinstitut. Eine vollständige Historisierung des Instanztyps Forschungsinstitut macht wenig Sinn, da Institute sehr selten als ganze Instanz aktualisiert werden. Außer bei der Löschung eines Instituts, werden meistens nur einzelne Attribute geändert. Deshalb sollen nur die aktualisierten Attribute wie Forschungsfelder historisiert werden. Auf diese Weise wird die redundante Datenhaltung vermieden, die entstehen würde, wenn alle Attribute bei der Änderung eines Attributs historisiert würden.

Zusammenfassend ermöglicht die Historisierung mittels temporaler Datenhaltung die Suche nach Informationen, die in der Vergangenheit gültig bzw. aktuell waren, aber trotzdem einen informationellen Wert für die Nutzer der Forschungslandkarte in der Gegenwart haben können. Ohne Historisierung könnten nur aktuelle Informationen auf der Forschungslandkarte angezeigt werden, was das Ziel der Plattform, nämlich einen

vollständigen Überblick der E-Government-Forschung, heutzutage wie auch in den letzten Jahren, zu erschaffen, verfehlen würde.

8.2 Aktualisierung der Inhalte

Die zweite Herausforderung, mit der sich diese Ausarbeitung beschäftigt, ist die Thematik der Aktualisierung der Inhalte der Forschungslandkarte. Das Ziel dieser Thematik ist ein Konzept zu entwickeln, mit dem es möglich ist, die Inhalte der Forschungslandkarte automatisch aktuell zu halten. Unter automatisch soll ein **dezentraler Prozess** der Aktualisierung verstanden werden. Das bedeutet, dass die Datenpflege nicht zentral durch die Administratoren der Forschungslandkarte, in diesem Fall das Institut für Wirtschafts- und Verwaltungsinformatik der Universität Koblenz, sondern automatisch durchgeführt werden soll. Der Begriff automatisch bezieht sich dabei nicht unbedingt auf ein autonomes Handeln durch das System, sondern kann beispielsweise auch durch das Verwalten der Inhalte durch die Nutzer realisiert werden. Ziel ist es, den Verwaltungsaufwand für die Projektverantwortlichen zu minimieren und so Kosten und Zeit zu sparen. Auf der anderen Seite kann nur durch eine automatische und dynamische Verwaltung die Aktualität der Inhalte effizient gewährleistet werden.

Das Aktualisierungsproblem der Forschungslandkarte E-Government hat verschiedene Ebenen bzw. Fragestellungen, die getrennt betrachtet werden sollen:

Es muss die Frage geklärt werden, wie die Daten aktuell gehalten werden können. Das umfasst sowohl das Einpflegen neuer Daten in das System als auch die laufende Aktualisierung bereits bestehender Daten.

Integration neuer Daten:

Während der Konzeptionierungsphase des Projekts wurden bereits verschiedene Methoden diskutiert, um neue Informationen bzw. neue Instanzen in das System zu integrieren. Beispielsweise könnte man anhand eines Crawlers die Informationen aus den Institutsseiten, Projektportalen etc. einlesen und so den Datenbestand dynamisch ergänzen. Dies wäre allerdings mit einem komplexen und teuren Implementierungsprozess verbunden. Einen derartigen Crawler zu entwickeln, erscheint auf den ersten Blick aufgrund der unterschiedlich aufgebauten und unstrukturierten Institutsseiten, Projekt- und Konferenzportalen etc. als eine hoch komplexe Aufgabe. Eine Alternative dazu wäre das manuelle Einpflegen der Daten durch die Nutzer. Sind

Herausforderungen im Kontext der E-Government-Forschungslandkarte: Historisierung und Aktualisierung

die Nutzer bereit aktiv zu partizipieren, könnte durch die manuelle Pflege der Inhalte durch die Nutzer, der Datenbestand aktuell und akkurat gehalten werden. Während der Konzeptionierungsphase des Projekts wurde zugunsten der manuellen Einpflege der Daten seitens der Nutzer entschieden. Da die Integration neuer Daten nicht Gegenstand dieser Ausarbeitung ist, soll diese Entscheidung ihre Gültigkeit behalten und das Thema der Einpflege neuer Daten nicht weiter berücksichtigt werden.

Ursprüngliche Aktualisierungen (erste Aktualisierungsebene):

Wie bereits erwähnt, müssen nicht nur neue Daten in das System integriert, sondern bereits bestehende Daten laufend aktualisiert werden. So müssen idealerweise alle Änderungen der realen Welt erfasst und im System durchgeführt werden. Mögliche Änderungen sind der Institutswechsel einer Person, Änderungen in den Forschungsfeldern einer Person oder eines Instituts oder einfache Attributsänderungen wie Namens- oder Anschriftsänderungen etc.. Diese Aktualisierungen entstehen aufgrund von Änderungen in der realen Welt und sollen in dieser Ausarbeitung als „ursprüngliche Aktualisierungen“ bezeichnet werden. Hierbei gilt es, die automatische Erfassung der Änderungen aus der realen Welt und ihre Umsetzung innerhalb der Forschungslandkarte auf ihre Machbarkeit hin zu untersuchen.

Ausgelöste Aktualisierungen (zweite Aktualisierungsebene):

Die zweite Aktualisierungsebene der Forschungslandkarte resultiert aus den ursprünglichen Aktualisierungen der bestehenden Informationen, die im letzten Abschnitt erwähnt wurden, unabhängig ob diese durch das System oder durch die Nutzer durchgeführt werden. Aufgrund der Beziehungen, die zwischen den einzelnen Instanzen bestehen können (siehe UML-Diagramm in Kapitelabschnitt 2.2), kann im schlimmsten Fall die Änderung eines Attributs oder einer Verknüpfung zu wasserfallartigen Änderungen bei den verknüpften Instanzen führen. Wird im Falle eines Institutswechsels einer Person beispielsweise eine Verknüpfung zwischen einer Person und einem Institut gelöscht und durch eine neue ersetzt, hat das weitere Folgen für diese und andere Instanzen. So muss die E-mail-Adresse der Person geändert werden, falls diese mit der Institutsadresse im System registriert ist. Darüber hinaus löst der Wechsel des Instituts Änderungen beim Institut aus. So müssen die Forschungsfelder des vorherigen Instituts aktualisiert werden, denn der Wechsel der Person kann unter Umständen dazu führen, dass die Forschung in diesen

Herausforderungen im Kontext der E-Government-Forschungslandkarte: Historisierung und Aktualisierung

Forschungsfeldern nicht mehr betrieben wird. Soll die Person im neuen Institut Forschung in diesen Feldern betreiben, muss eine Verknüpfung zwischen den Forschungsfeldern und dem neuen Institut erstellt werden. Außerdem muss das Attribut Kontaktperson des vorherigen Instituts aktualisiert werden, falls die wechselnde Person die Kontaktperson war. Diese Kette von Aktualisierungen soll in der Ausarbeitung „ausgelöste Aktualisierungen“ genannt werden, da diese als Folge der ursprünglichen Aktualisierungen resultiert. Auch hier ist es das Ziel zu entscheiden, ob eine automatisierte Umsetzung dieser Aktualisierungsketten technisch umsetzbar und sinnvoll ist oder ob diese manuell durch die Nutzer der Forschungslandkarte durchgeführt werden muss.

8.3 Abgrenzung und Gewichtung der Historisierung und Aktualisierung

Die Herausforderungen der Historisierung und Aktualisierung der Inhalte der Forschungslandkarte können nicht vollständig getrennt voneinander betrachtet werden. Die Historisierung der Inhalte ist notwendig, wenn sich Inhalte in der realen Welt ändern und somit auch innerhalb der Forschungslandkarte aktualisiert werden müssen. Ändert sich also eine Instanz in der realen Welt, führt dies dazu, dass relevante Zuordnungen dokumentiert (Historisierung) und weitere Änderungen an der Instanz oder an den verknüpften Instanzen durchgeführt werden müssen (Aktualisierung). Das Problem der automatischen Aktualisierung der Inhalte, also das Erfassen der Änderungen in der realen Welt und deren Umsetzung sowie die Umsetzung der „ausgelösten Aktualisierungen“ (siehe Kapitelabschnitt 8.2) kann vorläufig durch Datenpflege seitens der Nutzer gelöst werden, sodass in dieser Ausarbeitung nur kurz darauf eingegangen werden soll. Dies kommt für die Historisierung allerdings nicht in Frage, da Nutzer nicht über die technischen Fähigkeiten oder Mittel verfügen, um die Inhalte der Forschungslandkarte in dem gewünschten Maß zu historisieren.

Die Historisierung muss vom System gewährleistet werden. Ein Konzept hierzu, das als Basis für die Implementierung des Systems dient, soll im nächsten Kapitel erarbeitet werden.

9 Struktur und Lösungskonzept für die Historisierung der Forschungslandkarte

In diesem Kapitel wird ein Konzept für die Historisierung der Inhalte der Forschungslandkarte vorgelegt. In einem ersten Schritt soll die Datenstruktur der Landkarte modelliert werden, wobei temporale Aspekte mitberücksichtigt werden müssen. Dabei wird festgelegt, welche Instanzen, Attribute und Assoziationen historisiert werden sollen und welche nicht.

In einem zweiten Schritt werden Empfehlungen über welche Metadaten den Nutzern zur Verfügung gestellt werden sollen, gegeben, damit diese wichtige Informationen über temporale Aspekte der Inhalte erhalten.

9.1 Datenmodellierung

Wie in Kapitel 4 ausgeführt, müssen temporale Aspekte, die für die Historisierung relevant sind, bereits bei der konzeptionellen Ebene der Forschungslandkarte berücksichtigt und modelliert werden. Nur so wird gewährleistet, dass die Historisierung der Inhalte bei der Implementierung korrekt umgesetzt werden kann.

Eine wichtige Aufgabe bei der Konzeptionierung einer Online-Plattform ist die Modellierung der Daten. Das Ziel der Datenmodellierung ist nach Hopf⁹ den für die Online-Plattform relevanten Ausschnitt der realen Welt zu beschreiben. Dabei werden die zu dem Ausschnitt gehörenden Objekte, ihre Attribute sowie die Beziehungen zwischen den Objekten modelliert. Um komplexe Datenstrukturen zu vermeiden, muss bei der Datenmodellierung das Prinzip der Abstraktion angewendet werden. Es sollen laut Vossen also nur relevante Sachverhalte modelliert werden. Objekte oder Attribute, die für den Kontext der Online-Plattform nicht von Bedeutung sind, werden nicht berücksichtigt (Vossen, 2000, S. 79).

Anhand von abstrakten Datenmodellen soll ein Überblick über die Informationszusammenhänge und relevanten Sachverhalte geschaffen werden, die als Implementierungsgrundlage für die Datenbank und letztendlich für das zu implementierende System dienen sollen. Für diesen Zweck existieren verschiedene Datenmodellierungssprachen. Für die Forschungslandkarte soll die Datenmodellierung anhand von Entity-Relationship-Modellen nach Chen (1976) beschrieben werden. In

⁹ Hopf, H. G. (2004). *Semantische Datenmodellierung, Teil 1: Grundlagen*
<http://www.ohm-hochschule.de/fileadmin/Rechenzentrum/Virtuohm/Vorlesungen/Hopf/SDM-Teil1.pdf>

einem ersten Schritt soll mittels eines ER-Modell die Datenschicht der Forschungslandkarte beschrieben werden. Anschließend soll dieses Modell mit temporalen Aspekten erweitert werden, die die Historisierung der Inhalte ermöglichen.

9.1.1 Datenmodellierung der Forschungslandkarte

In diesem Kapitelabschnitt soll die Datenstruktur der Forschungslandkarte modelliert werden. Dabei werden alle relevanten Entitäten und ihre Attribute beschrieben. Darüber hinaus müssen die Beziehungen zwischen den Entitäten mit den jeweiligen Kardinalitäten modelliert werden. Dies wurde während der ersten Konzeptionierungsphase zwar schon anhand eines UML-Diagramms gemacht (siehe Kapitelabschnitt 2.2), nun soll dieses jedoch in Form eines Entity-Relationship-Modells übertragen werden. Dieser Schritt ist notwendig, da das ER-Modell anschließend mit den temporalen Aspekten für die Historisierung erweitert werden soll.

Abbildung 9-1 zeigt das Entity-Relationship-Modell der Forschungslandkarte ohne temporale Erweiterung. Es gelten die gleichen Entitäten und Beziehungen, die in Kapitelabschnitt 2.2 durch das UML-Diagramm festgelegt wurden. Auch die Attribute wurden übernommen, allerdings durch weitere ergänzt. Wie in der Abbildung 8.1 zu sehen ist, wurde an den Entitäten Person, Forschungsinstitut, Konferenz und Projekt jeweils ein Attribut hinzugefügt. Hierbei handelt es sich um zeitunabhängige Schlüsselattribute, die den Primärschlüssel einer Entität festlegen. Der Primärschlüssel ist ein Konzept der Datenbanktheorie und identifiziert eine Entität eindeutig (Sauer, 2002, S. 34). Die Schlüsselattribute sind im Diagramm durch einen schwarzen Pfeilkopf markiert. Der zeitunabhängige Schlüssel einer Person beispielsweise ist die Personnummer (Pe#). Diese ist eine beliebige Nummer, die die Person identifiziert. Das bedeutet, dass mehrere Personen nicht die gleiche Pe# haben dürfen. Das Hinzufügen der Schlüsselattribute ist notwendig für die Historisierung der Entitäten. Es wurden nur Schlüsselattribute den Entitäten angehängt, die historisiert werden sollen. Die Bedeutung der Schlüsselattribute für die Historisierung wird im Kapitelabschnitt 9.1.2 weiter ausgeführt.

Für alle Beziehungen gilt die Kardinalität (0, N). Jede Instanz kann sich mit beliebig vielen Instanzen verknüpfen, sofern die Verknüpfung erlaubt ist. Ein Projekt kann beispielsweise mit beliebig vielen Themenfeldern, Instituten oder Personen verknüpft werden, allerdings nicht mit einer Konferenz oder einer Publikation.

Nur die Entität Themenfeld kann mit allen anderen Entitäten verknüpft werden. Dies hängt damit zusammen, dass die Nutzer der Forschungslandkarte anhand der Taggingfunktion beliebig viele Themenfelder als Tags an alle Instanzen anhängen können.

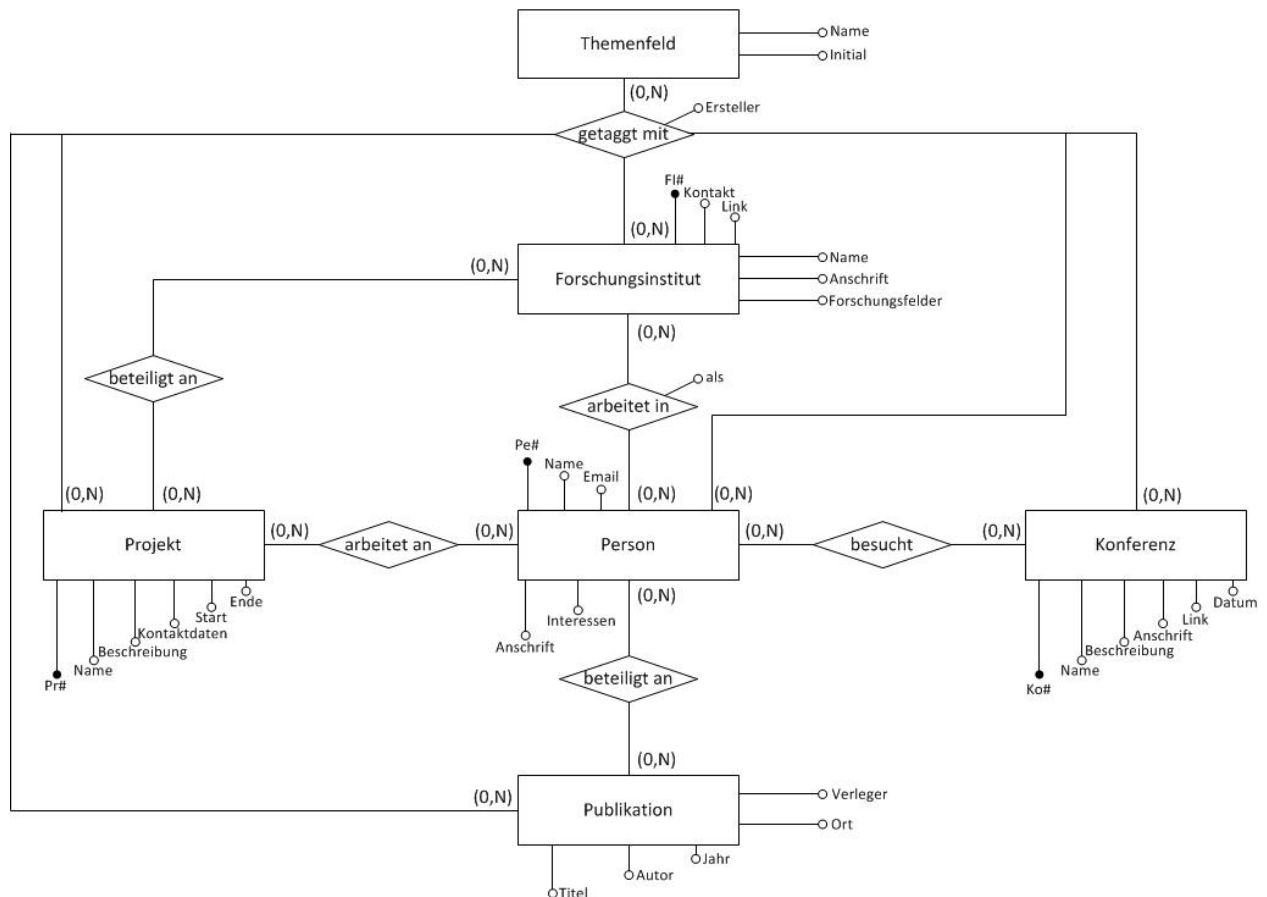


Abbildung 9-1: ER-Modell der Forschungslandkarte E-Government

9.1.2 Temporale Datenmodellierung der Forschungslandkarte

In Kapitelabschnitt 4.4 wurde das Modell RETTE nach Kaiser vorgestellt. Nun soll das RETTE Modell der Forschungslandkarte beschrieben werden. Hierbei wird das Entity-Relationship-Modell aus Kapitel 9.1.1 nach der RETTE-Notation um temporale Aspekte erweitert. Auf diese Weise ist es möglich, für alle Entitäten, Attribute und Beziehungen anzugeben, ob diese zeitbezogen sind oder nicht. Es wird also festgelegt, welche Objekte bei der Forschungslandkarte historisiert werden müssen. Das RETTE Modell dient als Grundlage für die Datenbankmodellierung und der späteren Implementierung der Forschungslandkarte.

Abbildung 9-2 zeigt das RETTE Modell der Forschungslandkarte E-Government, welches im folgenden Abschnitt beschrieben werden soll.

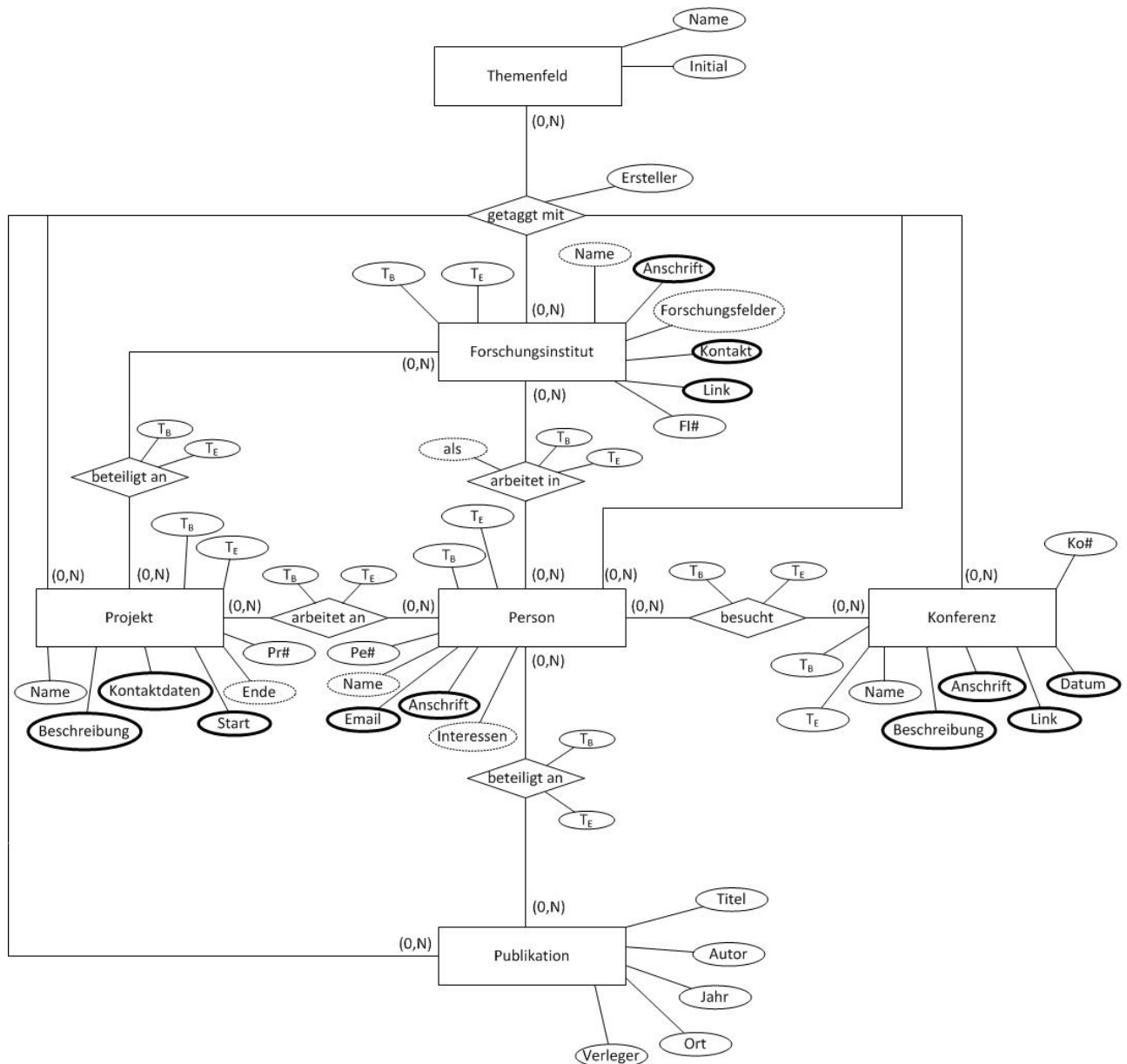


Abbildung 9-2: RETTE Modell der Forschungslandkarte E-Government

Wie bereits erwähnt, wird im RETTE Modell festgelegt, welche Entitäten, Beziehungen zwischen Entitäten sowie Attribute historisiert werden sollen. Zeitbezogene Entitäten erkennt man daran, dass an diese die Zeitstempelattribute T_B und T_E angehängt werden. Ein Beispiel hierfür ist die Entität Person.

Entitäten und ihre Attribute:

Die Zeitstempelattribute legen die Gültigkeitszeit eines Objekts vom Typ Person fest. Das zeitunabhängige Attribut Personenummer (Pe#) ist der Schlüssel der Entität Person. Der Wert dieses Attributs kann sich während der Lebensdauer einer Person nicht ändern. Darüber hinaus besitzt die Entität Person weitere zeitabhängige Attribute. Die Attribute Email und Anschrift sind zeitabhängige Attribute im weiteren Sinn, denn diese können sich während des Zeitablaufs eines Entitätstyps ändern. Allerdings muss die Historie dieser Attribute nicht festgehalten werden, da ihre früheren Werte nicht relevant sind. Lediglich der aktuelle Wert muss in der Datenbank gespeichert sein. Das Attribut Name hingegen kann sich im Rahmen einer legalen Namensänderung oder der Eheschließung ändern, die Historie der Werte muss aber nachvollziehbar sein. Wird nur der aktuelle Wert gespeichert, könnte man Autoren, bei denen sich der Name geändert hat, Publikationen, die vor der Namensänderung erschienen sind, nur schwer zuordnen. Daher ist das Attribut Name ein zeitbezogenes Attribut im engeren Sinn, was zu der Historisierung älterer Werte führt. Auch Interessen ist ein zeitbezogenes Attribut in engeren Sinn. Ändert sich der Wert, wird eine neue Version des Objekts angelegt und frühere Werte beibehalten, sodass die Historie nachvollziehbar ist.

Die Entität Forschungsinstitut ist ebenfalls zeitbezogen und besitzt Zeitstempelattribute sowie den zeitunabhängigen Schlüssel Forschungsinstitutsnummer (FI#). Weiterhin hat diese zeitabhängige Attribute, sowohl im weiteren als auch im engeren Sinn. Die Attribute Anschrift, Kontakt und Link können sich ändern, deren frühere Werte müssen aber nicht berücksichtigt werden. Anders ist es bei den Attributen Name und Forschungsfelder. Ändern sich diese Attribute, muss eine neue Version erstellt werden, wobei die früheren Werte erhalten bleiben müssen. Auf diese Weise kann der Nutzer sehen, in welchen Feldern in der Vergangenheit in einem bestimmten Institut Forschung betrieben wurde und ob sich der Name eines Instituts geändert hat und wie dieses früher hieß.

Die Entität Publikation ist eine konventionelle Entität. Das bedeutet, dass diese nicht historisiert werden muss. Die Voraussetzung hierfür ist, dass Publikationen nicht aus dem System gelöscht werden können, nachdem diese in der Datenbank aufgenommen wurden. Ausnahme ist das Löschen einer fehlerhaften Publikation durch den

Administrator. Das Historisieren fehlerhafter Publikationen ist allerdings nicht sinnvoll, sodass dieser Fall nicht weiter betrachtet werden muss.

Können Publikationen nicht durch die Nutzer gelöscht werden, ist eine Historisierung nicht notwendig, da sich die Informationen von Publikationen nach dem Erscheinen nicht ändern können. Erscheint beispielsweise eine neue Ausgabe, soll diese als eine neue Publikation behandelt werden, sodass keine Historisierung nötig ist.

Auch die Entität Themenfeld muss nicht historisiert werden, solange die Themenfelder nicht durch die Nutzer entfernt werden können.

Eine weitere zeitbezogene Entität in der Forschungslandkarte ist Konferenz. Auch hier wird die Gültigkeitszeit einer Konferenz durch Zeitstempelattribute festgelegt. Der zeitunabhängige Schlüssel Konferenznummer (Ko#) identifiziert eine Konferenz eindeutig. Der Name einer Konferenz kann sich im Normalfall nicht ändern, sodass dieser als zeitunabhängiges Attribut definiert wird. Alle anderen Attribute (Anschrift, Beschreibung, Link und Datum) können sich ändern, wobei jedoch nur der aktuelle Wert gespeichert werden soll.

Die Historie der Entität Projekt muss ebenfalls festgehalten werden. Auch hier wird das Attribut Name als zeitunabhängiges Attribut markiert, da der Name eines Projekts während der gesamten Lebensdauer eines Projekts gleich bleibt. Auch der zeitunabhängige Schlüssel Projektnummer (Pr#) bleibt im Zeitablauf unverändert. Die Attribute Beschreibung, Kontaktdaten und Start sind zeitabhängige Attribute im weiteren Sinn, da sie geändert werden können. Das Attribut Ende ist ein zeitabhängiges Attribut im engeren Sinn. Das Enddatum eines Projekts kann sich unter Umständen ändern, falls das Projekt verlängert wird. Durch die Historisierung des Attributs Ende können Nutzer nachvollziehen, ob eine Verlängerung eines Projekts stattgefunden hat.

Beziehungen und ihre Attribute:

Auch Beziehungen können im RETTE Modell als zeitbezogen definiert werden, falls diese historisiert werden sollen. Wie bei Entitäten werden hierfür die Zeitstempelattribute angehängt, die das Gültigkeitsintervall des Beziehungstyps abbilden.

Ein Beziehungstyp, der historisiert werden muss, ist die Verknüpfung arbeitet an zwischen Person und Projekt. Wird dieser Beziehungstyp historisiert, kann man nachvollziehen, welche Verknüpfungen in der Vergangenheit gültig waren. Das bedeutet, es kann angezeigt werden, welche Personen in der Vergangenheit an einem Projekt gearbeitet haben. Selbstverständlich werden auch gültige Verknüpfungen angezeigt, das heißt Personen, die momentan an einem bestimmten Projekt arbeiten, sowie auch Verknüpfungen die in der Zukunft gültig sein werden.

Analog dazu werden die Verknüpfungen beteiligt an zwischen Person und Publikation, besucht zwischen Person und Konferenz und beteiligt an zwischen Forschungsinstitut und Projekt historisiert.

Auch die Beziehung arbeitet in zwischen Person und Forschungsinstitut ist zeitbezogen und dokumentiert welche Personen (auch wenn diese nicht mehr dort arbeiten) bis zum heutigen Zeitpunkt in einem Institut beschäftigt waren. Zusätzlich ist das Attribut „als“ dieser Beziehung ein zeitabhängiges Attribut im engeren Sinn. Somit kann festgehalten werden, ob sich die Position einer Person innerhalb eines Instituts während des Zeitablaufs geändert hat und welche Position diese zu einem beliebigen Zeitpunkt in der Vergangenheit hatte.

Der Beziehungstyp getaggt mit zwischen Themenfeld und allen anderen Entitäten ist nicht zeitbezogen. Grund dafür ist, dass Nutzer wie bereits erwähnt nicht in der Lage sind, Themenfelder wieder aus dem System zu entfernen oder zu ändern, sodass diese keine Gültigkeitszeit besitzen. Somit bleibt die Verknüpfung zwischen einem Themenfeld und der jeweiligen Instanz bestehen.

9.2 Metadateninformationen in der Forschungslandkarte

In Kapitel 5 wurden die unterschiedlichen Arten von Metadaten und ihre Funktionen vorgestellt. Auch die Inhalte der Forschungslandkarte werden mit Metadaten, die die Daten beschreiben, semantisch ergänzt.

Eine wichtige Aufgabe bei der Konzeptionierung einer Online-Plattform ist laut Atzeni, Del Nostro, & Paolozzi zu entscheiden, welche Metadaten zusammen mit den zeitbezogenen Inhalten (Versionen) veröffentlicht werden sollen (2012, S. 170). Dies ist wichtig, da einige Metainformationen relevant für die Nutzer sind.

In der Publikation nach Atzeni, Del Nostro, & Paolozzi (2012) werden temporale Aspekte beim Design der Webseite festgelegt und modelliert. Bei der Modellierung ist es möglich zu entscheiden, welche Metadaten für welche Objekte angezeigt werden sollen. Die Webseite wird dann anhand der Modelle automatisch durch ein CASE Tool erzeugt. Das Paper schlägt eine konkrete Implementierung anhand eines bestimmten Tools vor. Die vorliegende Ausarbeitung soll allgemein gehalten werden und ein implementierungsunabhängiges Konzept vorstellen, sodass vom Modell nach Atzeni, Del Nostro, & Paolozzi etwas Abstand genommen wird. Aus diesem Grund werden erst einmal lediglich Empfehlungen geäußert, welche Metadaten für welche Instanzen oder Attribute angezeigt werden sollen, unabhängig davon auf welche Weise dies bei der Implementierung der Webseite realisiert wird. Eine mögliche Implementierungsoption (Dublin Core) wird als Handlungsempfehlung in Kapitel 11 vorgeschlagen.

Folgende Metainformationen können nach Atzeni, Del Nostro, & Paolozzi einen informationellen Wert für die Nutzer haben nach (2012):

Creator:

Person, die ein Inhaltsobjekt erstellt hat. Die Erstellerinformation kann sich sowohl auf die letzte und aktuelle Version oder aber auf alle Versionen eines Objekts beziehen.

Modifier:

Hiermit ist der Nutzer gemeint, der eine Änderung an einem Objekt durchgeführt hat. Da nach der Änderung eines zeitbezogenen Objekts eine neue Version erstellt wird, bezieht sich das Modifier Attribut auf eine Version des Objekts.

Description:

Beschreibt den Grund der Änderung eines zeitbezogenen Elements.

Last Modified:

Die Information Last Modified hält den Zeitpunkt fest, an dem ein zeitbezogenes Element das letzte Mal geändert wurde. Dieses ähnelt dem Prinzip der Transaktionszeit in der temporalen Datenbanktheorie und soll automatisch vom System verwaltet werden. Im Gegensatz zu der Transaktionszeit soll es laut Atzeni, Del Nostro, & Paolozzi für den Designer möglich sein zu entscheiden, für welche Elemente das Last Modified Attribut verwendet wird (2012, S. 170).

Validity Interval:

Dieses Attribut entspricht der Gültigkeitszeit und kann an jedes temporale Objekt angehängt werden.

Nun soll überlegt werden, für welche Elemente die obengenannten Metainformationen auf der Forschungslandkarte angezeigt werden sollen. Wie bereits erwähnt wird je nach dem Zeitpunkt, den der Nutzer auswählt, die zu diesem Zeitpunkt gültigen Versionen der Instanzen auf der Karte angezeigt. Darüber hinaus ist es möglich, Metainformationen über die Versionen anzuzeigen, um den Nutzern relevante Informationen über diese Versionen bereitzustellen.

Informationen über den Ersteller einer Instanz haben einen geringen informationellen Wert. Daher ist es nicht notwendig das Attribut **Creator** einer Instanz zu publizieren. Dies gilt für alle Objekte der Forschungslandkarte.

Alle Instanzen auf der Forschungslandkarte, abgesehen von Instanzen der Entität Forschungsinstitut, können nur durch den Ersteller der Instanz bearbeitet werden. Wurde eine Instanz geändert, kann man davon ausgehen, dass diese Änderung durch den Ersteller durchgeführt wurde, sodass das Attribut **Modifier** redundant wäre. Das gleiche gilt für die Verknüpfung zweier Instanzen. Anders ist es bei der Entität Forschungsinstitut. Institute können durch verifizierte Nutzer bearbeitet werden. Diese sind Nutzer, deren Angehörigkeit zu einem Forschungsinstitut durch eine zentrale Instanz verifiziert wurde. Unter Umständen könnte es interessant sein zu wissen, wer eine Instanz Forschungsinstitut aktualisiert hat. Selektiert man eine Instanz von Forschungsinstitut auf der Forschungslandkarte, soll daher das Attribut Modifier der jeweiligen Version des Instituts angezeigt werden.

Das Attribut **Description**, welches die Änderung eines Objekts begründet, ist für die Nutzer der Forschungslandkarte nicht relevant und soll deswegen nicht angezeigt werden.

Die Metadaten **Last Modified** und **Validity Interval** sind wichtige temporale Informationen, die relevant für die Historisierung der Inhalte sind und somit in folgenden Fällen sichtbar für die Nutzer sein müssen:

Die Attribute der zeitbezogenen Entität Konferenz sind alle zeitabhängige Attribute im weiteren Sinn. Diese können zwar geändert werden, die Historie der Werte muss aber nicht nachvollziehbar sein. Das bedeutet, dass bei Änderung der Attribute einer Konferenz keine neue Version angelegt wird. Daher ist es nicht nötig das Gültigkeitsintervall anzuzeigen. Sinnvoll ist es, den zeitabhängigen Attributen die Information Last Modified anzuhängen. Auf diese Weise erkennen Nutzer, dass und wann ein Attribut geändert wurde. Diese Information für die Instanz Konferenz anzuzeigen ist nicht notwendig, da die letzte Änderung über das geänderte Attribut erkennbar ist.

Bei der Entität Projekt soll für die zeitabhängigen Attribute im weiteren Sinn (Beschreibung, Kontaktdaten und Start) wie bei Konferenz, das Attribut Last Modified angezeigt werden. Das Attribut Ende ist ein zeitabhängiges Attribut im engeren Sinn, sodass ältere Werte behalten werden müssen. Daher soll in der jeweiligen Version die Metainformation Validity Interval sichtbar sein. Dadurch ist der Nutzer in der Lage zu erkennen, für welchen Zeitraum dieser Wert des Attributs Ende gültig ist oder war. Darüber hinaus ist es sinnvoll, das Gültigkeitsintervall für die jeweilige Version der Projektinstanz anzuzeigen. So kann der Nutzer erkennen, wann die Instanz in diesem Zustand in der realen Welt gültig war. Das Last Modified Attribut für Attribute im engeren Sinn anzuzeigen ist nicht notwendig, da anhand der Gültigkeitszeit der aktuellen Version die letzte Änderung erkennbar ist.

Analog zu der Entität Projekt sollen den Nutzern Metainformationen bei den Entitäten Person und Forschungsinstitut bereitgestellt werden. Zeitabhängige Attribute im weiteren Sinn sollen mit der Last Modified Information, zeitabhängige Attribute im engeren Sinn mit der Validity Interval Information versehen werden. Bei den Instanzen an sich soll die Validity Interval Information sichtbar sein.

Struktur und Lösungskonzept für die Historisierung der Forschungslandkarte

Die Entitäten Publikation und Themenfeld wurden als konventionelle Entitäten definiert, sodass keine Metainformationen über temporale Aspekte bereitgestellt werden müssen bzw. können.

Die Ergebnisse dieses Kapitels sollen im folgenden Kapitel in Form einer Anforderungsspezifikation zusammengefasst werden.

10 Anforderungsspezifikation für die Historisierung der Forschungslandkarte

In diesem Kapitel werden die funktionalen Anforderungen an die Historisierung der Forschungslandkarte dargestellt. In Tabelle 2 wird festgelegt, welche Funktionen durch die Historisierung umgesetzt werden müssen. Die Anforderungen werden in Pflicht und optional kategorisiert.

Funktionale Anforderungen	Pflicht/ Optional
<i>Historisierung der Entitäten</i>	
Die Entität Person muss historisiert werden. Für jede Version eines Objekts von Person wird die Gültigkeitszeit festgelegt	Pflicht
Die Entität Konferenz muss historisiert werden. Für jede Version eines Objekts von Konferenz wird die Gültigkeitszeit festgelegt	Pflicht
Die Entität Projekt muss historisiert werden. Für jede Version eines Objekts von Projekt wird die Gültigkeitszeit festgelegt	Pflicht
Die Entität Forschungsinstitut muss historisiert werden. Für jede Version eines Objekts von Forschungsinstitut wird die Gültigkeitszeit festgelegt	Pflicht
<i>Historisierung einzelner Attribute</i>	
Das Attribut Pe# von Person hat bei alle Versionen eines Objekts von Person den gleichen Wert und ist unveränderbar	Pflicht
Ändert sich das Attribut Name eines Objekts von Person, wird eine neue Version dieses Objekts angelegt	Pflicht
Ändert sich das Attribut Interessen eines Objekts von Person, wird eine neue Version dieses Objekts angelegt	Pflicht
Ändert sich das Attribut E-mail eines Objekts von Person, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Anschrift eines Objekts von Person, wird der Wert in der aktuellen Version überschrieben	Pflicht
Das Attribut Ko# von Konferenz hat bei alle Versionen eines Objekts von Konferenz den gleichen Wert und ist unveränderbar	Pflicht

Anforderungsspezifikation für die Historisierung der Forschungslandkarte

Das Attribut Name von Konferenz ist unveränderbar	Pflicht
Ändert sich das Attribut Beschreibung eines Objekts von Konferenz, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Anschrift eines Objekts von Konferenz, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Link eines Objekts von Konferenz, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Datum eines Objekts von Konferenz, wird der Wert in der aktuellen Version überschrieben	Pflicht
Das Attribut Pr# von Projekt hat bei alle Versionen eines Objekts von Projekt den gleichen Wert und ist unveränderbar	Pflicht
Das Attribut Name von Projekt ist unveränderbar	Pflicht
Ändert sich das Attribut Beschreibung eines Objekts von Projekt, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Kontaktdaten eines Objekts von Projekt, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Start eines Objekts von Projekt, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Ende eines Objekts von Projekt, wird eine neue Version dieses Objekts angelegt	Pflicht
Das Attribut FI# von Forschungsinstitut hat bei alle Versionen eines Objekts von Forschungsinstitut den gleichen Wert und ist unveränderbar	Pflicht
Ändert sich das Attribut Name eines Objekts von Forschungsinstitut, wird eine neue Version dieses Objekts angelegt	Pflicht
Ändert sich das Attribut Forschungsfelder eines Objekts von Forschungsinstitut, wird eine neue Version dieses Objekts angelegt	Pflicht
Ändert sich das Attribut Anschrift eines Objekts von Forschungsinstitut, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Kontakt eines Objekts von Forschungsinstitut, wird der Wert in der aktuellen Version überschrieben	Pflicht
Ändert sich das Attribut Link eines Objekts von Forschungsinstitut, wird der Wert in der aktuellen Version überschrieben	Pflicht

Ändert sich das Attribut „als“ eines Beziehungsexemplars von „arbeitet in“ zwischen Person und Forschungsinstitut, wird eine neue Version dieses Beziehungsexemplars angelegt	Pflicht
<i>Historisierung von Beziehungstypen</i>	
Der Beziehungstyp besucht zwischen Person und Konferenz muss historisiert werden. Für jedes Beziehungsexemplar von besucht wird die Gültigkeitszeit festgelegt	Pflicht
Der Beziehungstyp arbeitet an zwischen Person und Projekt muss historisiert werden. Für jedes Beziehungsexemplar von arbeitet an wird die Gültigkeitszeit festgelegt	Pflicht
Der Beziehungstyp beteiligt an zwischen Person und Publikation muss historisiert werden. Für jedes Beziehungsexemplar von beteiligt an wird die Gültigkeitszeit festgelegt	Pflicht
Der Beziehungstyp beteiligt an zwischen Forschungsinstitut und Projekt muss historisiert werden. Für jedes Beziehungsexemplar von beteiligt an wird die Gültigkeitszeit festgelegt	Pflicht
Der Beziehungstyp arbeitet in zwischen Person und Forschungsinstitut muss historisiert werden. Für jedes Beziehungsexemplar von arbeitet in wird die Gültigkeitszeit festgelegt	Pflicht
<i>Metainformationen</i>	
Wird eine Instanz der Entität Forschungsinstitut geändert, wird die Metainformation Modifier bei der jeweiligen Version angezeigt	Optional
Die Gültigkeitszeit der jeweiligen Versionen eines Objekts der Entität Person wird angezeigt	Pflicht
Die Gültigkeitszeit der jeweiligen Versionen eines Objekts der Entität Projekt wird angezeigt	Pflicht
Die Gültigkeitszeit der jeweiligen Versionen eines Objekts der Entität Forschungsinstitut wird angezeigt	Pflicht
Die Gültigkeitszeit des Attributs Name wird in der jeweiligen Version eines Objekts der Entität Person angezeigt	Pflicht
Die Gültigkeitszeit des Attributs Interessen wird in der jeweiligen Version eines Objekts der Entität Person angezeigt	Pflicht
Die Metainformation Last Modified des Attributs Email eines Objekts der	Optional

Anforderungsspezifikation für die Historisierung der Forschungslandkarte

Entität Person wird angezeigt	
Die Metainformation Last Modified des Attributs Anschrift eines Objekts der Entität Person wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Beschreibung eines Objekts der Entität Konferenz wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Anschrift eines Objekts der Entität Konferenz wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Link eines Objekts der Entität Konferenz wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Datum eines Objekts der Entität Konferenz wird angezeigt	Optional
Die Gültigkeitszeit des Attributs Ende wird in der jeweiligen Version eines Objekts der Entität Projekt angezeigt	Pflicht
Die Metainformation Last Modified des Attributs Beschreibung eines Objekts der Entität Projekt wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Kontaktdaten eines Objekts der Entität Projekt wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Start eines Objekts der Entität Projekt wird angezeigt	Optional
Die Gültigkeitszeit des Attributs Name wird in der jeweiligen Version eines Objekts der Entität Forschungsinstitut angezeigt	Pflicht
Die Gültigkeitszeit des Attributs Forschungsfelder wird in der jeweiligen Version eines Objekts der Entität Forschungsinstitut angezeigt	Pflicht
Die Metainformation Last Modified des Attributs Anschrift eines Objekts der Entität Forschungsinstitut wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Kontakt eines Objekts der Entität Forschungsinstitut wird angezeigt	Optional
Die Metainformation Last Modified des Attributs Link eines Objekts der Entität Forschungsinstitut wird angezeigt	Optional
<i>Systemfunktionen</i>	
Der Benutzer kann die Inhalte der Forschungslandkarte nach einem beliebigen (Gültigkeits-)Zeitpunkt filtern	Pflicht

Der Benutzer kann die Inhalte der Forschungslandkarte nach einem beliebigen (Gültigkeits-)Zeitintervall filtern	Pflicht
Ohne Zeitspezifizierung werden alle aktuell gültigen Instanzen auf der Karte angezeigt	Pflicht
Der Benutzer kann in der erweiterten Suche die Suchergebnisse auf einem beliebigen (Gültigkeits-)Zeitpunkt einschränken	Pflicht
Der Benutzer kann in der erweiterten Suche die Suchergebnisse auf einem beliebigen (Gültigkeits-)Zeitintervall einschränken	Pflicht

Tabelle 2: Anforderungen an Historisierung der Forschungslandkarte

Der Anforderungskatalog aus Tabelle 2 legt die Funktionen fest, die vom System umgesetzt werden sollen, um die Inhalte der Forschungslandkarte zu historisieren. Basierend darauf werden im nächsten Kapitel konkrete Handlungsempfehlungen für die Historisierung gegeben.

11 Handlungsempfehlungen für die Historisierung der Forschungslandkarte

Handlungsempfehlungen Content Management System:

In Kapitelabschnitt 3.3 wurde die Bedeutung eines Content Management Systems für die Implementierung einer Online-Plattform erörtert. Besonders für Webseiten, die eine große Anzahl von Informationen verwalten müssen, sind Content Management Systeme geeignet. Daher ist es sinnvoll, die Forschungslandkarte auf ein CMS aufzubauen, das für die Umsetzung und Durchführung der Content Management Aufgaben genutzt werden kann. In Kapitelabschnitt 7.2 wurde das Content Management System Drupal vorgestellt. Des Weiteren wurden Gründe genannt, warum sich Drupal für das Verwalten der Inhalte einer Forschungslandkarte eignet. Um die Inhalte der Forschungslandkarte effektiv zu verwalten, soll daher das **CMS Drupal als Grundlage des Systems** genutzt werden.

Eine Alternative ist es, für die Implementierung **einen CASE Tool zu nutzen, der den ganzen Erstellungsprozess der temporalen Online-Plattform unterstützt**. In der Literatur gibt es hierzu mehrere Ansätze, die auf CASE Tools basieren. Die Publikation von Atzeni, Del Nostro, & Paolozzi (2012) schlägt einen solchen Implementierungsprozess vor. Während der konzeptionellen Phase werden unterschiedliche Modelle genutzt, um die temporalen Aspekte der Inhalte der Webseite durch den Designer zu definieren. Hier ist es möglich, die Datenmodellierung der Webseite temporal zu erweitern, um zu definieren, welche Inhalte historisiert werden sollen und wie diese dem Nutzer präsentiert werden sollen. Basierend auf diesen Modellen übernimmt das CASE Tool alle notwendigen formalen Transformationen und erstellt automatisch die Webseite, sowie das dazugehörige Content Management System, um die Inhalte zu verwalten. Die Daten werden dabei in einer herkömmlichen relationalen Datenbank gespeichert. Somit unterstützt das Tool den gesamten Prozess der Webseiteentwicklung und gewährleistet die Historisierung der Inhalte. Allerdings sind solche Tools in einem relativ frühen Entwicklungsstand, sodass Probleme auftreten können. Eine weitere Gefahr ist, dass diese Tools nicht speziell für die Implementierung einer Forschungslandkarte entwickelt wurden. Hier gilt es zu prüfen, inwiefern solche CASE Tools den Anforderungen einer Forschungslandkarte genügen.

Handlungsempfehlungen Historisierung:

Eine wichtige Entscheidung, die vor der Implementierung der Forschungslandkarte getroffen werden muss, ist die Frage, wie die Historisierung auf Datenbankebene realisiert werden soll. Es gibt unterschiedliche Wege, zeitliche Dimensionen wie Transaktions- oder Gültigkeitszeit auf der Datenbank abzubilden. Wie in Kapitelabschnitt 4.2 ausgeführt gibt es sogenannte temporale Datenbanken und bitemporale Datenbanken. Diese sind Datenbanksysteme, die die temporale Datenhaltung anhand von temporalen Datentypen unterstützen. Allerdings existieren derzeit nur vereinzelte Lösungen und Prototypen, sodass der Support dieser Datenbanksysteme praktisch nicht gegeben ist. **Die beste Lösung ist es, eine konventionelle relationale Datenbank so zu erweitern, dass temporale Aspekte abgebildet werden können.** Hierbei werden temporale Attribute wie herkömmliche Attribute behandelt. Laut Atzeni, Del Nostro, & Paolozzi werden Zeitstempel genutzt, um die Gültigkeitszeit von Objekten abzubilden und die Transaktionszeit zu speichern. Für temporale Attribute werden zusätzliche Tabellen erstellt, um die zeitlichen Aktualisierungen in der Datenbank zu verwalten (Laut Atzeni, Del Nostro, & Paolozzi, 2012). Auf diese Weise können temporale Daten in einer relationalen Datenbank gespeichert werden.

Handlungsempfehlungen Metadaten:

Metadaten werden wie bereits ausgeführt unter anderem dazu genutzt, Ressourcen im Web zu beschreiben und semantisch zu ergänzen. Auch im Falle der Forschungslandkarte spielen Metainformationen eine wichtige Rolle, auch für die Historisierung. Um die Metadaten strukturiert darstellen zu können, soll im Rahmen der Forschungslandkarte ein Metadatenstandard gewählt werden. In Kapitelabschnitt 5.4 wurde das Metadaten-Schemata **Dublin Core** vorgestellt, das genutzt werden soll, um die Ressourcen der Forschungslandkarte zu beschreiben. Dublin Core ist sehr weit verbreitet, auch im E-Government Bereich, sodass bereits bestehende Ansätze übernommen und angepasst werden können.

In Kapitelabschnitt 9.2 wurden einige Metainformationen vorgeschlagen, die zusammen mit den Inhalten der Forschungslandkarte veröffentlicht werden sollen, um den Nutzern relevante Informationen über temporale Aspekte zur Verfügung zu stellen. Diese sind das „Last Modified“ Attribut und das „Validity Interval“ Attribut. Durch den qualifizierten Dublin Core, welche in Kapitelabschnitt 5.4.2 vorgestellt wurde, lässt sich laut Ebner

Handlungsempfehlungen für die Historisierung der Forschungslandkarte

die Metainformation „Date“ so verfeinern, dass die Informationen Last Modified und Validity Interval abgebildet werden können. Durch den Qualifier *Date.Valid* kann das Gültigkeitsintervall einer Ressource festgelegt werden. Das Last Modified Attribut kann durch den Qualifier *Date.Modified* zum Ausdruck gebracht werden (Ebner, 2002). Diese können dann problemlos dem Nutzer, wie in Kapitel 10 spezifiziert, zur Verfügung gestellt werden. Auch andere Metadaten wie Creator sind in Dublin Core enthalten, was für die Nutzung des Standards spricht.

12 Fazit und Ausblick

Die vorliegende Bachelorarbeit hatte als Ziel die Erörterung der Herausforderungen und Probleme für die Historisierung der Inhalte einer Forschungslandkarte. In einem weiteren Schritt sollen geeignete Technologien untersucht werden, mit denen diese Probleme gelöst werden können.

Um ein einheitliches Konzept entwickeln zu können, wurde im ersten Teil der Arbeit die Forschungslandkarte E-Government vorgestellt. Insbesondere die Entitäten und ihre Beziehungen zueinander spielen bei der Historisierung eine wichtige Rolle, sodass diese klar definiert sein müssen. Des Weiteren wurde der Bedarf einer Forschungslandkarte für den Bereich E-Government aufgezeigt. Für die Nutzer der Forschungslandkarte sind nicht nur aktuelle Informationen relevant, sondern auch solche, die in der Vergangenheit gültig waren. Um Nutzern solche Informationen bereitzustellen, müssen die Inhalte der Forschungslandkarte historisiert werden, da man sonst deren Entwicklung im Laufe der Zeit nicht nachvollziehen kann.

Im zweiten Teil der Arbeit wurde die theoretische Grundlage anhand der Literatur gelegt. Dabei wurden Technologien vorgestellt, mit denen ein Rahmen aufgebaut werden kann, ohne den die Historisierung der Inhalte nicht realisiert werden kann. Zu diesen Technologien gehören Content Management Systeme, um die Inhalte der Forschungslandkarte zu verwalten und das Konzept der temporalen Datenhaltung, um die Inhalte zu historisieren. Darüber hinaus wurden Metadaten und Ontologien vorgestellt, um die Plattform semantisch zu gestalten und um unter anderem die Interoperabilität der Seite zu gewährleisten.

Im dritten Teil der Arbeit wurde ein Konzept vorgestellt, um die Inhalte der Forschungslandkarte zu historisieren. Wie bereits erwähnt sollen die Inhalte der Forschungslandkarte mit einem CMS verwaltet werden. Hierfür wurde eine Anforderungsliste verfasst, die die notwendigen Funktionen eines CMS im Kontext einer Forschungslandkarte festlegt. Basierend darauf wurde Drupal als geeigneter CMS vorgestellt und die Gründe für eine mögliche Verwendung von Drupal aufgezeigt.

Fazit und Ausblick

Die Kapitel 8 und 9 widmen sich der Historisierung der Inhalte. Zuerst wurde das Problem der Historisierung und Aktualisierung der Forschungslandkarte beschrieben und beide Teilprobleme voneinander abgegrenzt. Dabei wurde festgestellt, dass die Aktualisierung und Historisierung nicht getrennt voneinander betrachtet werden kann. Ändert sich ein Objekt in der realen Welt, muss dieses erfasst und im System aktualisiert werden. Diese ursprünglichen Aktualisierungen führen dazu, dass weitere Aktualisierungen durchgeführt werden müssen (ausgelöste Aktualisierungen). Dabei müssen relevante Sachverhalte, die sich geändert haben, durch das System historisiert werden.

Die Historisierung der Inhalte muss von Beginn an geplant werden und nicht erst bei der Implementierung des Systems. Es ist von großer Bedeutung, relevante zeitliche Dimensionen bereits bei der konzeptionellen Ebene zu berücksichtigen. Hierzu wurde das Datenmodell der Forschungslandkarte anhand des R_{ETTE} Modell um temporale Aspekte erweitert. Auf diese Weise wurde festgelegt, welche Entitäten, Attribute und Beziehungen historisiert werden sollen. Anhand von temporalen Metadaten werden den Nutzern zusätzliche Informationen zur Verfügung gestellt. Hat ein Objekt mehrere Versionen, soll angezeigt werden, wann die jeweilige Version gültig war oder ist. Bei Attributen, bei denen nur der aktuelle Wert relevant ist, wird anhand des Attributs Last Modified angezeigt, wann das Attribut das letzte Mal geändert wurde.

Basierend auf dem temporalen Datenmodell wurde die Anforderungsspezifikation für die Historisierung erfasst. Hier wurde festgelegt, welche Objekte historisiert und welche Metadaten angezeigt werden müssen. Darüber hinaus wurden Funktionsanforderungen an das System gestellt, wie z. B. das Filtern der Inhalte nach der Gültigkeitszeit. So können Nutzer Informationen erhalten, die in der Vergangenheit gültig waren und die Forschungslandkarte zeitbezogen navigieren.

Im letzten Kapitel der Arbeit wurden konkrete Handlungsempfehlungen ausgesprochen, die bei der Implementierung des Systems berücksichtigt werden können. Dabei müssen die Anforderungen aus Kapitel 10 beachtet werden.

Wie bereits erwähnt soll die Forschungslandkarte auf einem Content Management System aufgebaut sein. Hier empfiehlt es sich, das CMS Drupal zu nutzen, um die Inhalte zu verwalten. Alternativ könnte ein CASE Tool genutzt werden, das ausgehend von Modellen die Webseite und das dazugehörige CMS automatisch erstellt.

Fazit und Ausblick

Es gibt unterschiedliche Arten, wie man temporale Daten auf der Datenbankebene verwalten kann. Hierfür wurde vom Autor der vorliegenden Arbeit die Abbildung temporaler Daten auf einer konventionellen relationalen Datenbank vorgeschlagen. Hierbei werden temporale Attribute wie herkömmliche Attribute behandelt. Zeitliche Dimensionen werden anhand von Zeitstempeln repräsentiert.

Um Metainformationen strukturiert darstellen zu können, soll der Standard Dublin Core genutzt werden. Auf diese Weise können die Inhalte mit Metadaten wie Creator, Validity Interval oder Last Modified semantisch beschrieben werden.

Mit der vorliegenden Bachelorarbeit wurde ein Konzept für die Historisierung angesetzt und Handlungsempfehlungen gegeben, wie das Konzept letztendlich realisiert werden kann. In einem weiteren Schritt sollten diese Empfehlungen genauer untersucht werden. Dabei gilt es zu prüfen, welche der Vorschläge sich am besten für die Realisierung der Historisierung eignen. Die Historisierung der Inhalte ist ein wichtiger Bestandteil der Forschungslandkarte und muss nach der Implementierung der Plattform voll funktionsfähig sein.

Außerdem sollte ein Konzept für die automatische Aktualisierung der Inhalte entwickelt werden. Dabei steht die Frage im Vordergrund, ob die Inhalte der Forschungslandkarte dynamisch und automatisiert aktuell gehalten werden können. Beispielsweise könnten anhand eines Crawlers neue Daten in das System eingespeist und bestehende Daten aktualisiert werden. Hierbei gilt es zu prüfen, ob eine solche Lösung technisch machbar und vor allem sinnvoll ist oder ob die Daten weiterhin durch die aktive Partizipation der Nutzer aktuell gehalten werden sollen.

Als letzten Schritt gilt es, die Forschungslandkarte E-Government anhand der entwickelten Konzepte zu implementieren.

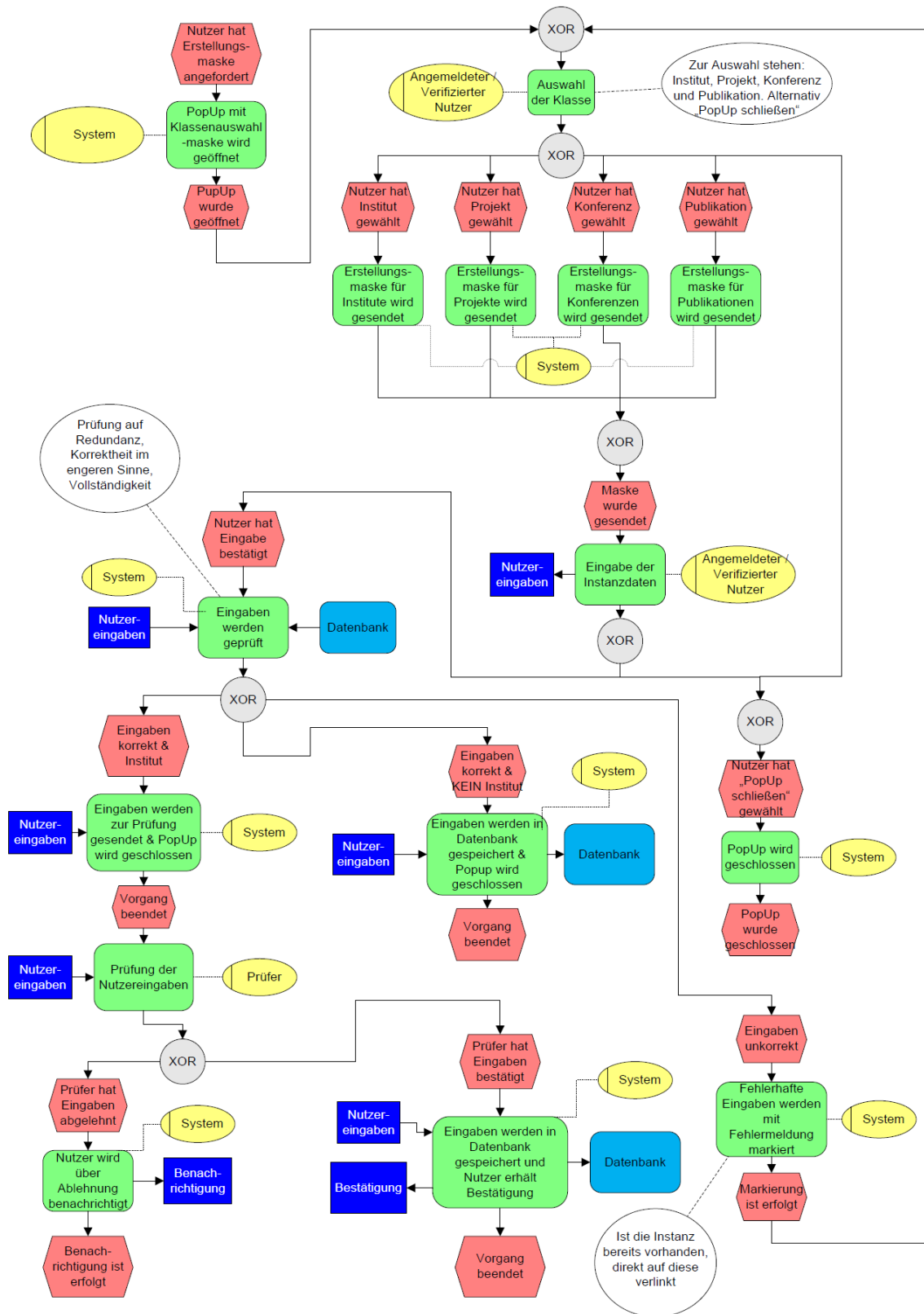
13 Literaturverzeichnis

- Alby, T. (2008). *Web 2.0 - Konzepte, Anwendungen, Technologien*. München: Carl Hanser Verlag.
- Atzeni, P., Del Nostro, P., & Paolozzi, S. (2012). Temporal Content Management and Website Modeling: Putting Them Together. In A. Hameurlain, J. Küng, & R. Wagner, *Transactions on Large-Scale Data- and Knowledge-Centered Systems V* (S. 158-182). Berlin, Heidelberg: Springer.
- Baker, T. (Oktober 2000). A Grammar of Dublin Core. *D-Lib Magazine*.
- Balzer, S. (28. November 2001). Entwicklung eines Metadatengesteuertes Historisierungswerkzeugs für die Aktualisierung von Data Warehouses. Zürich: Universität Zürich.
- Baumann, S. (2003). Ganzheitliche Unterstützung der Wertschöpfungskette in Medien-Unternehmen mit Content Management Lösungen. In F. Stahl, & W. Maass, *Content Management Handbuch* (S. 77-88). St. Gallen: NetAcademy Press =mcm institute, Universität St. Gallen.
- Bodendorf, F. (2005). *Daten- und Wissensmanagement* (2. Ausg.). Berlin, Heidelberg: Springer.
- Breslin, J. G., Passant, A., & Vrandečić, D. (2011). Social Semantic Web. In D. John, F. Dieter, & J. A. Hendler, *Handbook of Semantic Web Technologies* (S. 467-506). Berlin, Heidelberg: Springer.
- Brooks, C. H., & Montanez, N. (2006). An Analysis of the Effectiveness of Tagging in Blogs. *Computation Approaches to Analyzing Weblogs. Papers from the 2006 AAAI Spring Symposium*, (S. 9-15).
- Chen, P. P.-S. (März 1976). The Entity-Relationship Model-Toward a Unified View of Data. *ACM Transactions on Database Systems*, S. 9-36.
- Dobratz, S., Hänger, A., Huth, K., Kaiser, M., Keitel, C., Klump, J., et al. (November 2009). <http://www.langzeitarchivierung.de>. Von http://files.d-nb.de/nestor/materialien/nestor_mat_08.pdf abgerufen
- Domingue, J., Fensel, D., & Hendler, J. A. (2011). Introduction to the Semantic Web Technologies. In J. Domingue, D. Fensel, & J. A. Hendler, *Handbook of Semantic Web Technologies* (S. 5-39). Berlin, Heidelberg: Springer.
- Ebner, W. (Oktober 2002). Temporales Web-Management. Wien: Wirtschaftsuniversität Wien.
- Etzion, O., Gal, A., & Segev, A. (1998). Extended Update Functionality in Temporal Databases. In O. Etzion, S. Jajodia, & S. Sripada, *Temporal Databases: Research and Practices* (S. 56-95). Berlin, Heidelberg: Springer.
- Euzenat, J., & Shvaiko, P. (2007). *Ontology Matching*. Berlin, Heidelberg: Springer.
- Frick, R., & Najar, C. (2009). *Historisierung, nachhaltige Verfügbarkeit und Archivierung von Geoinformationen - Eine Auslegeordnung*. Basel: AG SIK-GIS.
- Gašević, D., Djurić, D., & Devedžić, V. (2009). *Model Driven Engineering and Ontology Development* (2. Ausg.). Berlin, Heidelberg: Springer.
- Gersdorf, R. (2003). Content Management für die flexible Informationswiederverwendung. In F. Stahl, & M. Wolfgang, *Content Management Handbuch* (S. 59-75). St. Gallen: NetAcademy Press =mcm institute, Universität St. Gallen.
- Goralwalla, I. A., Özsu, M. T., & Szafron, D. (1998). An Object-Oriented Framework for Temporal Data Models. In O. Etzion, S. Jajodia, & S. Sripada, *Temporal Databases: Research and Practice* (S. 1-35). Berlin, Heidelberg: Springer.
- Graf, H. (2006). *Drupal. Community-Websites entwickeln und verwalten mit dem Open Source-CMS*. München: Addison-Wesley.
- Graf, H. (2008). *Drupal 6: Websites entwickeln und verwalten mit dem Open Source-CMS*. München: Addison-Wesley Verlag.
- Gregersen, H., & Jensen, C. S. (Mai/Juni 1999). Temporal Entity-Relationship Models—A Survey. *IEEE Transactions on Knowledge and Data Engineering*, S. 464-497.
- Grimm, S., Abecker, A., Völker, J., & Rudi, S. (2011). Ontologies and the Semantic Web. In J. Domingue, D. Fensel, & J. A. Hendler, *Handbook of Semantic Web Technologies* (S. 508-537). Berlin, Heidelberg: Springer.

- Gruber, T. (2005). *tomgruber.org*. Abgerufen am 12. März 2012 von <http://tomgruber.org/writing/ontology-of-folksonomy.htm>
- Guenther, R., & Radebaugh, J. (2004). *Understanding Metadata is a revision and expansion of Metadata Made*. NISO Press.
- Hengartner, U. (2003). Effektive Suchverfahren für Content Management Systeme. In F. Stahl, & M. Wolfgang, *Content Management Handbuch* (S. 131-147). St. Gallen: NetAcademy Press =mcminstitute, Universität St. Gallen.
- Hofman, H. (1999). Herausforderungen an die digitale Erhaltung. In M. (. Wettengel, *Digitale Herausforderungen für Archive* (S. 9-13). Koblenz: Bundesarchiv.
- Kaiser, A. (1999). Eine temporale Erweiterung des Entity Relationship Modells. *Rundbrief der GI-Fachgruppe 5.10 Informationssystem-Architekturen*.
- Kampffmeyer, U., & Rogalla, J. (2000). *Grundsätze der elektronischen Archivierung*. Hamburg: VOI Verband Organisations- und Informationssysteme e.V.
- Kashyap, V., Bussler, C., & Moran, M. (2008). *The Semantic Web*. Berlin, Heidelberg: Springer.
- Koop, H. J., Jäckel, K. K., & Anja, O. (2001). *Erfolgsfaktor Content Management: Vom Web Content bis zum Knowledge Management*. Wiesbaden: Vieweg.
- Kronz, S. (2003). *Content Management: Einführung, Prozesse und Objekte*. Saarbrücken: Universität des Saarlandes.
- Kumar, A., Tsotras, V., & Faloutsos, C. (Januar/Februar 1998). Designing access methods for bitemporal databases. *IEEE Transactions on Knowledge and Data Engineering*, S. 1-20.
- Maass, W., & Stahl, F. (2003). Content Management als Teil des Kommunikations-Management. In F. Stahl, & W. Maass, *Content Management Handbuch* (S. 35-47). St. Gallen: NetAcademy Press =mcminstitute, Universität St. Gallen.
- Myrach, T. (2005). *Temporale Datenbanken in Betrieblichen Informationssystemen: Prinzipien, Konzepte, Umsetzung*. Wiesbaden: Teubner Verlag.
- Nix, M. (2005). *Web Content Management*. Software & Support Verlag.
- Odenthal, R. (2010). *Digitale Archivierung* (2. Ausg.). Heidelberg: Hüthig Jehle Rehm.
- Peters, I. (2009). *Folksonomies. Indexing And Retrieval In Web 2.0*. Berlin: Walter de Gruyter.
- Ramsch, J., & Sosna, D. (1997). *Temporale Datenmodelle und Metainformationssysteme für GIS*. Leipzig: Universität Leipzig.
- Rothfuss, G., & Ried, C. (2001). *Content Management mit XML*. Berlin, Heidelberg: Springer.
- Sauer, H. (2002). *Relationale Datenbanken - Theorie und Praxis*. München: Addison-Wesley Verlag.
- Schumann, M. (Dezember 2007). *Zur Bedeutung von Metadaten in einem umfassenden Veröffentlichungsprozess*. Leipzig: HTWK Leipzig.
- Schuster, E., & Wilhelm, S. (2000). Content Management. *Informatik Spektrum*, 23(6), 373 - 375.
- Sondermann, T. (13. März 2006). Ein datenbankgestütztes Werkzeug zur Exploration von ontologischen Daten und deren Evolution. Bonn: Rheinische Friedrich-Wilhelms-Universität Bonn.
- Spörrer, S. (2009). *Content-Management-Systeme: Begriffsstruktur und Praxisbeispiel*. Köln: Kölner Wissenschaftsverlag.
- Stahl, F., & Wolfgang, M. (2003). Marktübersicht von Content Management Systemen . In F. Stahl, & M. Wolfgang, *Content Management Handbuch* (S. 195-393). St. Gallen: NetAcademy Press =mcminstitute, Universität St. Gallen.
- Stöbel, D. (2010). Entwicklung einer generischen Plattform zur Umsetzung einer Forschungslandkarte. Universität Koblenz-Landau.
- Sugimoto, S., Baker, T., & Weibel, S. (2002). Dublin Core: Process and Principles. *ICADL '02 Proceedings of the 5th International Conference on Asian Digital* (S. 25-35). London: Springer.
- Van Dyk, J. K. (2009). *Das Drupal-Entwicklerhandbuch*. München: Addison-Wesley Verlag.
- Vossen, G. (2000). *Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme*. München: Oldenbourg Verlag.
- Weibel, S., & Koch, T. (Dezember 2000). The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions. *D-Lib Magazine*.

- Wiemer, K. (18. Juni 2008). Metadatenformate. Bonn: Fachhochschule Bonn-Rhein-Sieg.
- Zapf, S., & Myrach, T. (2000). Ein Prototyp zur Unterstützung temporaler Aspekte der Web-Administration auf Basis eines Repositories. In H. Schmidt (Hrsg.), *Modellierung betrieblicher Informationssysteme: PROCEEDINGS der MobIS-Fachtagung 2000* (S. 291-299). Siegen: Fachbereich 5 Wirtschaftsinformatik.
- Zschau, O. (2003). Web Content Management Systeme - Eine Einführung. In F. Stahl, & M. Wolfgang, *Content Management Handbuch* (S. 49-57). St. Gallen: NetAcademy Press =mcminstitute, Universität St. Gallen.

Anhang 1: Instanzerstellung bei Forschungslandkarte



Instanzerstellung bei Forschungslandkarte (Entnommen aus Projekthandbuch des Projektpraktikums Forschungslandkarte E-Government 2010)

Anhang 2: DCMI Recommended Qualifiers

DCMES Element	Element Refinement(s)	Element Encoding Scheme(s)
<u>Title</u>	<u>Alternative</u>	-
<u>Creator</u>	-	-
<u>Subject</u>	-	<u>LCSH</u> <u>MeSH</u> <u>DDC</u> <u>LCC</u> <u>UDC</u>
<u>Description</u>	<u>Table Of Contents</u> <u>Abstract</u>	-
<u>Publisher</u>	-	-
<u>Contributor</u>	-	-
<u>Date</u>	<u>Created</u> <u>Valid</u> <u>Available</u> <u>Issued</u> <u>Modified</u> <u>Date Accepted</u> <u>Date</u> <u>Copyrighted</u> <u>Date Submitted</u>	<u>DCMI Period</u> <u>W3C-DTF</u>
<u>Type</u>	-	<u>DCMI Type Vocabulary</u>
<u>Format</u>	-	<u>IMT</u>
	<u>Extent</u>	-
	<u>Medium</u>	-
<u>Identifier</u>	-	<u>URI</u>
	<u>Bibliographic Citation</u>	-

Anhang 2: DCMI Recommended Qualifiers

<u>Source</u>	-	<u>URI</u>
<u>Language</u>	-	<u>ISO 639-2RFC 3066</u>
<u>Relation</u>	<u>Is Version Of</u> <u>Has Version</u> <u>Is Replaced By</u> <u>Replaces</u> <u>Is Required By</u> <u>Requires</u> <u>Is Part Of</u> <u>Has Part</u> <u>Is Referenced By</u> <u>References</u> <u>Is Format Of</u> <u>Has Format</u> <u>Conforms To</u>	<u>URI</u>
<u>Coverage</u>	<u>Spatial</u>	<u>DCMI Point</u> <u>ISO 3166</u> <u>DCMI Box</u> <u>TGN</u>
	<u>Temporal</u>	<u>DCMI Period</u> <u>W3C-DTF</u>
<u>Rights</u>	<u>Access Rights</u>	-
	<u>License</u>	<u>URI</u>
<u>Audience</u>	<u>Mediator</u> <u>Education Level</u>	-
<u>Provenance</u>	-	-
<u>Rights Holder</u>	-	-
<u>Instructional Method</u>	-	-
<u>Accrual Method</u>	-	-
<u>Accrual Periodicity</u>	-	-

Anhang 2: DCMI Recommended Qualifiers

<u>Accrual Policy</u>	-	-
-----------------------	---	---

DCMI Recommended Qualifiers (Sugimoto, Baker, & Weibel, 2002)