Universität Konstanz
Fachbereich Informatik und Informationswissenschaft
Studienfach Information Engineering

Masterarbeit

zum Thema

Instrumente zur Metadatenerschließung im WWW am Beispiel der Online-Schlagwortnormdatei

vorgelegt von

Michael Bohner

Mai 2002

1. Gutachter: Prof. Dr. Marc Scholl

2. Gutachter: Prof. Dr. Daniel Keim

Abstract

Deutsch

Um die Fülle an Informationen und Diensten im World Wide Web (WWW) effizienter nutzen zu können, wird intensiv an Technologien geforscht, die eine Berücksichtigung des Bedeutungsgehalts von Objekten im WWW ermöglichen. Sie reichen von der Einigung auf inhaltliche Beschreibungsstandards bis zur Vision eines „Semantic Web“, welches die rechnerbasierte Verarbeitung und Interpretation von Informationen erlaubt. Allen Ansätzen ist gemeinsam, dass sie die Auszeichnung der Webressourcen mit den für die jeweilige Technologie relevanten Metadaten voraussetzen. Eine besondere Bedeutung kommt in diesem Zusammenhang kontrollierten Vokabularen zu, die eindeutige und einheitliche repräsentative Einheiten für die Metadatenbeschreibung zur Verfügung stellen. Diese Arbeit gibt einen Überblick über die bedeutendsten aktuellen Vorschläge zu einer semantischen Nutzung des WWW und stellt das Konzept der kontrollierten Vokabulare vor. Am Beispiel der Schlagwortnormdatei, einem Thesaurus aus dem deutschsprachigen, wissenschaftlichen Bibliotheksbereich, wird aufgezeigt, wie ein kontrolliertes Vokabular für die Online-Erschließung im WWW genutzt werden kann. Ferner wurde eine Anwendung implementiert, die die aktive Teilnahme neuer Nutzerkreise an Pflege und Erweiterung des Thesaurus erlaubt. Hierbei kommen gängige WWW-Sprachen und Datenbanktechnologie zum Einsatz.

Schlagwörter: Inhaltserschließung, Metadaten, World Wide Web,
Schlagwortnormdatei <Werk>, Semantic Web

English

In order to make possible an efficient use of the vast amount of information and services on the World Wide Web (WWW) intensive research continues to be undertaken on technologies which enable us to exploit the semantics of objects on the WWW. These technologies cover a variety of proposals from the agreement on content based description standards to the vision of a “Semantic Web” that allows information to be processed and interpreted by machines. All approaches have in common that they require the annotation of web resources with metadata relevant for the corresponding technology. Controlled vocabularies play an important role in this process as they provide unambiguous, standardized building blocks for the description of resources with metadata. This thesis provides an overview of the most important current proposals for the semantic use of the WWW and presents the concept of controlled vocabularies. The “Schlagwortnormdatei” (Subject Authority File, SWD), a thesaurus developed in the domain of scientific libraries in German-speaking countries, serves as an example for the use of a controlled vocabulary for the online annotation of resources in the WWW. Moreover, an application has been implemented which allows the active participation of new user groups in the maintenance and extension of the thesaurus. This application has been developed by the use of common WWW languages and database technology.

keywords: subject indexing, metadata, World Wide Web, Schlagwortnormdatei,
subject authority file, Semantic Web

Inhaltsverzeichnis

1. EINLEITUNG 1

2. BEDEUTUNG VON METADATEN FÜR DAS WORLD WIDE WEB 3

2.1 METADATEN: DATEN ÜBER DATEN 3
2.2 ERFORDERNIS VON METADATEN 4
2.3 BESONDERHEITEN DER METADATENERSCHLIEßUNG IM WWW 6

3. TECHNOLOGIEN ZUR SEMANTISCHEN NUTZUNG DES WWW 9

3.1 ALLGEMEINE METADATENELEMENTE IN HTML 9

3.2 DUBLIN CORE: KERNBAUSTEIN FÜR METADATENSTANDARDS IM WWW 11
3.2.1 Entwurf und Entwicklung der semantischen Kernelemente 11
3.2.2 Qualifizierung von Dublin Core 13
3.2.3 Sachstand und Kritik 14

3.3 XML: METASPRACHE FÜR SEMANTISCHE AUSZEICHNUNGSVOKABULARE 17
3.3.1 Ursprung als generische Auszeichnungssprache 17
3.3.2 Eigenschaften 18
3.3.3 Anwendung 21

3.4 RDF: STANDARDREPRÄSENTATIONSSPRACHE FÜR METADATEN 22
3.4.1 Motivation und Ursprung 22
3.4.2 Datenmodell 23
3.4.3 RDFS: Definition von Auszeichnungsvokabularen 25
3.4.4 Mehrwerte und Anwendung 27

3.5 ONTOLOGIESPRACHEN FÜR DAS WWW 31
3.5.1 Motivation und Definition 31
3.5.2 Ontologiesprache DAML+OIL 32
3.5.3 Mehrwerte und Anwendung 36

3.6 ARCHITEKTUR EINES SEMANTIC WEB 39
3.6.1 Schichtenmodell 39
3.6.2 Charakteristika 41
3.6.3 Anwendungsszenarien 43

4. KONTROLLIERTE VOKABULARE 45
4.1 BEDEUTUNG 45
4.2 STRUKTURIERUNGSFORMEN 45
4.2.1 Taxonomie / hierarchische Klassifikation 46
4.2.2 Thesaurus 46
4.2.3 Ontologie 47

5. ONLINE-SCHLAGWORTNORMDATEI (OSWD) 50
5.1 BIBLIOTHEKARISCHE INHALTSERSCHLIEßUNG 50
5.2 SCHLAGWORTNORMDATEI (SWD) 51
5.2.1 Merkmale 51
5.2.2 Kooperation 53
5.2.3 Qualitätssicherung 54
5.3 AUFBAU DER OSWD 55

6. PROJEKT: ERWEITERUNG DER SWD 57
6.1 MOTIVATION: AUSWEITUNG DES ANWENDUNGSBEREICHES 57
6.2 ANFORDERUNGEN 58
6.3 PLANUNG UND ENTWICKLUNG 60
6.3.1 Datenmodell und Datenfluss 60
6.3.2 Entwicklungsumgebung 61
6.3.3 Softwareergonomische Aspekte 63
6.4 SOFTWARETECHNISCHE REALISIERUNG 65
6.4.1 Architektur 65
6.4.2 Dateneingabe 66
6.4.3 Suche 67
6.4.4 Mailboxsystem 68
6.4.5 Redaktionstool 68
6.4.6 Berücksichtigung der Usability-Ziele 69

7. PILOTVERSUCH 71
7.1 ABLAUF UND BETEILIGTE 71
7.2 NUTZUNG UND PROBLEME 71
7.3 BEWERTUNG 72

8. ZUSAMMENFASSUNG 74
9. ANLAGEN 76
10. LITERATUR 92

Abbildungsverzeichnis

Abbildung 1: Inhaltsorientierte Metatags in HTML.....10

Abbildung 2: Dublin Core Element Set.....12

Abbildung 3: Beispiel für ein RDF-Statement.....23

Abbildung 4: Beispiel für einen strukturierten Wert in einem RDF-Statement.....24

Abbildung 5: Beispiel für eine Reifikation in RDF.....24

Abbildung 6: Klassensystem von RDFS. Aus: [W3C 2000].....25

Abbildung 7: Beispiel für eine RDFS-Anwendung. Aus: [W3C 2000].....26

Abbildung 8: Klassenkonstruktoren in DAML+OIL.....33

Abbildung 9: Konstruktoren für Eigenschaftselemente in DAML+OIL.....34/35

Abbildung 10: Konstruktoren für Eigenschaftsbeschränkungen in DAML+OIL.....35

Abbildung 11: Das Semantic Web Schichtenmodell. Aus: [Berners-Lee 2000].....39

Abbildung 12: Datenfluss zwischen den Projektteilnehmern.....59

Abbildung 13: Modularisierung der Datenerfassung.....65

Abbildung 14: Modularisierung des Redaktionstools.....66

Abkürzungen

API	Application Programming Interface
BSZ	Bibliotheksservice-Zentrum Baden-Württemberg
CGI	Common Gateway Interface
DAML	DARPA Agent Markup Language
DARPA	Defense Advanced Research Projects Agency
DBMS	Database Management System
DCMI	Dublin Core Metadata Initiative
DIN	Deutsches Institut für Normung
DOM	Document Object Model
DTD	Dokumenttyp-Deklaration
GKD	Gemeinsame Körperschaftsdatei
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IBW	Intelligent Bibliographic Workstation
ILTIS	Integriertes Literatur-, Tonträger und Musikalien-Informations-System
ISO	International Standard Organization
KI	Künstliche Intelligenz
MAB2	Maschinelles Austauschformat für Bibliotheken 2
OAI	Open Archives Initiative
OIL	Ontology Inference Layer
OPUS	Online Publikationsverbund der Universität Stuttgart
OSWD	Online-Schlagwortnormdatei
PND	Personennamendatei
RDF	Resource Description Framework
RDFS	RDF Schema
RSWK	Regeln für den Schlagwortkatalog
SGML	Standard Generalized Markup Language
SOAP	Simple Object Access Protocol
SQL	Structured Query Language
SWD	Schlagwortnormdatei
URI	Universal Resource Identifier
URL	Uniform Resource Locator
VBScript	Visual Basic Script
W3C	World Wide Web Consortium
WWW	World Wide Web
XSL	Extensible Stylesheet Language

1. Einleitung

Das World Wide Web (WWW) ist ein überwältigender Erfolg: innerhalb von wenigen Jahren hat es sich zu einem der bedeutendsten Medien für den Austausch von Informationen entwickelt. Der Erfolg scheint jedoch insofern auch eine Gefahr für den Nutzen des WWW darzustellen, als es wegen der angebotenen Informationsfülle und der Dynamik der Veränderungen oft schwierig ist, die gewünschte Information oder Dienstleistung zu finden und zu nutzen.

Dies beruht unter anderem darauf, dass die bisher vorherrschenden Methoden zur Recherche im WWW den Bedeutungsgehalt der angebotenen Informationen unzureichend berücksichtigen. Es wird dem Benutzer bei der Navigation durch Verfolgung von Hyperlinks vollständig selbst überlassen, den Inhalt einer Seite zu erfassen und durch Auswahl geeigneter Pfade zur gewünschten Information zu gelangen. Bei der Suche mit Hilfe von Suchmaschinen erfolgt die Selektion in der Regel auf einem rein syntaktischen Abgleich zwischen Suchanfrage und dem Volltext der indexierten Ressourcen, was zu den bekannten Problemen bei der Trennschärfe von Anfragen und der Relevanz der Treffermengen führt.

Demgegenüber existieren Verfahren die die Einbeziehung der Semantik der in Frage kommenden Objekte bei der Suche vorsehen. Sie setzen in der Regel eine vorherige inhaltliche Erschließung der Objekte in Form einer Beschreibung voraus, auf die gesondert zugegriffen werden kann. In gewisser Weise werden also Daten über die Objektdaten, sogenannte Metadaten, eingesetzt.

Durch eine solche Auszeichnung mit Metadaten ist das Problem jedoch im allgemeinen noch nicht gelöst, da eine völlig freie Beschreibung dieselben Schwierigkeiten der Mehrdeutigkeit natürlicher Sprache mit sich bringen würde, die bereits zur mangelnden Präzision der Volltextsuche führt. Der Kontrolle des bei der Beschreibung verwendbaren Vokabulars kommt bei der Metadatenerschließung daher eine große Bedeutung zu. Dabei handelt es sich um keine triviale Aufgabe, da sichergestellt werden muss, dass das Vokabular den Beschreibungsbedürfnissen zahlreicher Anbieter von Informationen gerecht wird, wobei es sich insbesondere im WWW um sehr heterogene Gruppen handeln kann.

In der vorliegenden Arbeit wird untersucht, welche Technologien zu einer semantischen Erschließung des WWW vorliegen und inwieweit diese bereits eingesetzt werden. Dabei wird zunächst der Begriff der Metadaten näher beleuchtet und besonders auf die Formen und den Einsatz von kontrollierten Vokabularen eingegangen.

Ferner wird mit der Online-Schlagwortnormdatei (OSWD) ein Vokabular vorgestellt, welches im wissenschaftlichen Bibliotheksbereich für die Sacherschließung entwickelt und gepflegt wird und nun für die Auszeichnung von Online-Ressourcen mit inhaltsbeschreibenden Metadaten zur Verfügung gestellt wurde. Während bei der Schaffung der OSWD zunächst die Beschreibung von wissenschaftlichen Online-Publikation im Vordergrund stand, wird sie als universales Vokabular zunehmend auch zur Auszeichnung von Objekten durch weitere Informationsanbieter wie etwa Museen und Archive genutzt.

In diesem Zusammenhang stellte sich die Frage, wie der neue Nutzerkreis in die bereits etablierte, technisch und organisatorisch sehr ausgereifte Entwicklung und Pflege des Vokabulars einbezogen werden kann, damit es langfristig auch seinen Beschreibungsbedürfnissen gerecht wird. Hierzu habe ich

im Rahmen eines Pilotprojektes eine Lösung implementiert, die zu einem großen Teil auf WWW-Sprachen und Datenbanktechnologie beruht, die auch anderen Gemeinschaften zur Ressourcenbeschreibung zur Verfügung stehen.

2. Bedeutung von Metadaten für das World Wide Web

2.1 Metadaten: Daten über Daten

Der Begriff *Metadaten* ist derzeit sehr häufig, in unterschiedlichen Kontexten und Schattierungen anzutreffen, wobei die genaue Herkunft unklar ist. Er scheint aus dem Gebiet der elektronischen Informations- und Datenverarbeitung der 70er Jahre zu stammen und rasch Anwendung in verschiedenen Bereichen gefunden zu haben [Lynch et al 1995].

Trotz der aktuellen Überstrapazierung soll zunächst eine begriffliche Klärung versucht werden: Neben der wortgemäßen Definition von Metadaten als „Daten über Daten“ herrscht im Webkontext offenbar Einigkeit darüber, dass es sich bei Metadaten um strukturierte Informationen handelt, die dazu dienen, Objekte im World Wide Web (WWW) inhaltlich und/oder formal zu beschreiben (vgl. beispielsweise [Tennant 1998], [Milstead, Feldman 1999]). Darüber hinaus ist eine Definition ausgehend vom Zweck dieser Beschreibung sinnvoll: Nach [Rusch-Feja 1997] S. 115 sind Metadaten „Informationen über andere Daten (Dokumente, Datensammlungen, Bilder, Server etc.), die in einer Form gehalten werden, dass sie die Recherche, das Retrieval und die Nutzung der Primärdokumente ermöglichen, erleichtern und ggf. bestimmen.“ Weitere Definitionen heben den Aspekt hervor, dass es sich bei Metadaten um ein Surrogat handelt, welches dem Benutzer Schlüsse zu dem beschriebenen Objekt erlaubt, ohne dass er umfassende vorherige Kenntnis über dessen Existenz oder Charakteristika haben muss [Milstead, Feldman 1999].

Im Sinne dieser Definitionen können alle Arten von bibliographischen Informationen, Zusammenfassungen, Indexierungstermen, Abstracts etc., die die ursprünglichen Objekte auf eine bestimmte Art beschreiben, Metadaten sein. Obwohl der Begriff der elektronischen Informations- und Datenverarbeitung entstammt, trifft er somit insbesondere auch auf Instrumente zu, die – zumindest ursprünglich – außerhalb des WWW angesiedelt waren: So sind beispielsweise Katalogkarten oder Titelaufnahmen in Bibliotheken typische Formen von Metadaten, die u. a. zur Beschreibung, Verwaltung und zum (Wieder-)Finden von Medien dienen. Die Literaturbeschreibung in Bibliothekskatalogen ist im Grunde nichts anderes als die Auszeichnung mit Metadaten, wobei bei der Formalerschließung eher nach formalen und bei der Sacherschließung nach inhaltlichen Kriterien beschrieben wird. [Hacker 2000] S. 195.

Da der Begriff der Metadaten so weit gefasst ist, ist es hilfreich, diese stets vor dem Hintergrund der Prozesse zu betrachten, die sie unterstützen oder erleichtern sollen. Danach ist in Anlehnung an [Bearman, Sochats 1995] eine Unterscheidung in folgende Arten möglich:

Metadaten zu

- Identifikations- und Nachweiszwecken (z.B. Titel, Autor)
- Nutzungs- und Beschaffungskonditionen (z.B. Nutzungsrechte, Preis)
- Strukturellen Aspekten (z.B. Art des Dokuments, Gliederung in Kapitel)
- Kontext (z.B. Querverweise, ähnliche Quellen)
- Inhalt (z.B. Schlagwörter, Abstract)
- Nutzungs- und Wirkungsgeschichte (z.B. frühere Versionen)
- Technischen Aspekten (z. B. Datenformat, Speicherbedarf)

Diese Arbeit wird sich auf inhaltsbeschreibende Metadaten konzentrieren, die in erster Linie das Auffinden von relevanten Objekten erleichtern sollen.

2.2 Erfordernis von Metadaten

Aufgrund der von der klassischen Medienverwaltung und –recherche abweichenden Bedingungen im WWW stellt sich zunächst die Frage, weshalb das – wie dargestellt – allgemeine Konzept der Metadaten in dieser Umgebung noch relevant ist. Der Suchvorgang kann im Gegensatz zur Situation bei nicht vernetzten Medien direkt auf möglicherweise relevanten Objekten ausgeführt werden. Da diese außerdem schon digitalisiert zur Verfügung stehen, können wesentliche Metainformationen automatisch aus den Objekten extrahiert bzw. berechnet werden.

Entsprechend basieren Suchmaschinen und sonstige Retrievalprozesse für das WWW überwiegend auf der Schlüsselwortsuche in Volltexten. Wie Untersuchungen zur semantischen Nutzung des Web zeigen, wurde erst relativ spät und mit bislang begrenztem Erfolg versucht, eine systematische Auszeichnung mit Metadaten zu bewirken und für die Informationssuche im WWW zu nutzen. [Rusch-Feja 1997] S. 117, [Clark 1999], [INT Media 2002]

Allerdings zeigt der Umgang mit Suchmaschinen auch mögliche Unzulänglichkeiten dieses Konzeptes: Recherchen liefern oft Ergebnisse mit zu großen Treffermengen und zu geringer Präzision. Das Matching der Suchterme mit den Volltexten der Originaldokumente, ggf. ergänzt um Algorithmen und Heuristiken, die linguistische Besonderheiten (z. B. Stammreduktionen, Stopwortlisten), Dokumentstruktur, Formatierungen, Nutzungshäufigkeiten etc. berücksichtigen, scheint klassische metadatenbasierte Recherchemethoden nicht völlig zu kompensieren.

Allgemein bietet die Nutzung von Metadaten gegenüber direkt auf den Objekten operierenden Retrievalprozessen den Vorteil, dass sie inhaltliche Beschreibungen zur Verfügung stellen können, die nicht in den gesuchten Objekte selbst enthalten sind oder nur mit hohem Aufwand daraus gewonnen werden können. Bei Multimediaobjekten wie z. B. Bildern, Audio- oder Videodateien erscheint dieser Vorteil derzeit noch wesentlich: Soweit Recherchemöglichkeiten im WWW zu diesen Objekten existieren, stützen sie sich auf Metadatenbeschreibungen; lauffähige Anwendungen, die Eigenschaften aus der internen Repräsentation der Objekte extrahieren und zu Retrievalzwecken auswerten, sind in diesem Bereich noch nicht zu finden.

Aber auch im Zusammenhang mit Textdokumenten spielt diese Eigenschaft von Metadaten eine Rolle: Bei der Schlüsselwortsuche werden Treffer weitgehend auf syntaktischer Ebene, d. h. anhand der Übereinstimmung von Zeichenfolgen, ausgewählt. Sprachliche Mehrdeutigkeiten führen dabei regelmäßig dazu, dass relevante Objekte nicht gefunden werden, wenn im fraglichen Dokument beispielsweise ein Synonym für den entscheidenden Begriff der Suchanfrage verwendet wird. Umgekehrt besteht die Gefahr, dass zu viele Objekte für relevant befunden werden, falls Terme der Suchanfrage im gefundenen Dokument zwar vorhanden, dort aber in anderer Bedeutung verwendet werden. Es existieren zwar Versuche auf Seiten der Suchmaschinen, anhand des gemeinsamen Vorkommens von mehreren Termen den Kontext und somit die Semantik eines einzelnen Termes in einem Dokument zu bestimmen. Bislang schien dies jedoch nur in Teilbereichen effektiv zu gelingen. Es ist fraglich, ob damit die

Trennschärfe einer intellektuellen Verschlagwortung unter Zuhilfenahme eines kontrollierten Vokabulars erreicht werden kann. [Milstead, Feldman 1999]

Daneben sprechen folgende Gründe dafür, dass Metadaten auch im WWW eine Bedeutung zukommt:

- **Abdeckungsgrad**
  Mithilfe von Metadaten können auch Objekte beschrieben bzw. repräsentiert werden, die selbst nicht über das Netz zugänglich sind wie z. B. nur als Printausgabe existierende Werke, Objekte in Drittspeichern.

- **Performanz**
  Metadatenbeschreibungen sind in der Regel kleiner als die Objekte, auf die sie sich beziehen. Sie nehmen daher weniger Ressourcen bei Übertragung, Suche und Speicherung in Anspruch.

- **Architektur**
  Viele Retrievalprotokolle ermöglichen es nicht, nur auf einen Teil der Zielobjekte zuzugreifen. Indem die relevanten Daten extrahiert und in getrennten Metadatenbeschreibungen gespeichert werden, kann direkt und ausschließlich auf diesen Datenteil zugegriffen werden.

- **Informationelle und wirtschaftliche Aspekte**
  Metadaten ermöglichen die Kenntnis von einem Objekt und dessen Beurteilung bevor der eigentliche Zugriff erfolgt. Dies kann wesentlich für die Bekanntmachung von Produkten und sinnvolle Kaufentscheidungen bei gleichzeitiger Gewährleistung des Urheberrechtsschutzes sein.

- **Reduzierte Hilfe- und Orientierungsmöglichkeiten**
  Im Netz entfallen für den Informationssuchenden Hilfe- und Orientierungsmöglichkeiten, die in der Realwelt zur Verfügung stehen: So finden sich z. B. in einer Bibliothek oder einem Museum Ansprechpartner, die differenziert Auskünfte erteilen können; Objekte wie Bücher sind für den Informationssuchenden oft konkret greifbar, was weitere Rückschlüsse auf den Inhalt erlaubt. Dies spricht dafür, im WWW zusätzliche Metadaten einzusetzen, um diesen Wegfall von Hilfsmöglichkeiten zu kompensieren. Ein Beispiel hierfür sind Rezensionen in Online-Shops.

  [Lynch et al 1995]

Diese Zusammenstellung lässt erkennen, dass einige Gründe für eine Verwendung von Metadaten im Webkontext sprechen und ihre Bedeutung in diesem Bereich eventuell sogar zugenommen hat. [Weibel 1997] vertritt die Auffassung, dass die Vernetzung zwar die Bedeutung der Lokalität als Hindernis für die Kommunikation erheblich vermindert, das Erfordernis einheitlicher Metadaten aber eher verstärkt hat.
Diese Sichtweise wird auch von [Dempsey et al 1997] unterstützt:
“It is recognised that in an indefinitely large resource space, effective management of networked information will increasingly rely on effective management of metadata. The need for metadata services is already clear in the current Internet environment. As the Internet develops into a mixed information economy deploying multiple application protocols and formats this need will be all the greater. Metadata is not only key to discovery, it will also be fundamental to effective use of found resources (by establishing the technical or business frameworks in which they can be used) and to interoperability across protocol domains.”

Mithin ist parallel zur Entwicklung des WWW auch eine Fülle von neuen Metadatenstandards entstanden. Einen Eindruck in Form von Übersichten vermitteln beispielsweise [Dempsey et al 1997] und [IFLA 2002].
Einige wesentliche Standards wie beispielsweise Dublin Core werden in Kapitel 2 näher besprochen. Ein zentraler Grund für das Entstehen der Vielzahl von Metadatenformaten dürfte die Tatsache sein, dass kein Standard für alle Zwecke geeignet ist. Es ergeben sich von einander abweichende Anforderungen in bezug auf unterschiedliche Medientypen, verschiedene wissenschaftliche und wirtschaftliche Gebiete, verschiedene Benutzergruppen etc. Die dezentrale Struktur des WWW, verbunden mit einfachen Protokollen und geringen Zugangskosten, fördert darüber hinaus die Schaffung von proprietären Lösungen. Daher kann davon ausgegangen werden, dass die Bedeutung von Abbildungen zur Gewährleistung von Interoperabilität zwischen verschiedenen Standards und Systemen und die Definition von auf Benutzergruppen zugeschnittenen Vokabularen zunehmen wird [Milstead, Feldman 1999]. Zu diesem Zweck werden u. a. sogenannte *Metadaten Registries* geschaffen: Über das WWW zugängliche Datenbanken, die über vorhandene Formate informieren und den Vergleich mit Hilfe von Abbildungen ermöglichen. Ein Beispiel hierfür ist die „Meta-Form“-Datenbank der Niedersächsischen Staats- und Universitätsbibliothek [Becker et al 2000].

2.3 Besonderheiten der Metadatenerschließung im WWW

Trotz der Unterschiede zwischen verschiedenen Standards können einige gemeinsame Merkmale ermittelt werden, die die Auszeichnung mit Metadaten im Web generell von der klassischen Erschließung unterscheiden.

Wer zeichnet aus?

Abgesehen von vereinzelten Versuchen auch im WWW traditionell zu katalogisieren, scheint Einigkeit darüber zu herrschen, dass eine generelle Erschließung der Internetobjekte und entsprechende Auszeichnung mit Metadaten durch Bibliotheken oder andere Dokumentationsstellen aufgrund der Größe des WWW und des starken, eigendynamischen Wachstums nicht möglich ist. [Milstead, Feldman 1999] [Capelleveen 1997]

Besondere Bedeutung kommt im Rahmen der Erschließung im WWW daher den Autoren zu: Sie sollen die von ihnen erstellten Objekte bereits selbst mit aussagekräftigen Metadaten auszeichnen. Diese Verschiebung der Erschließungsleistung bewirkt neben der Kostenersparnis, dass die Autoren Inhalt und Zweck der von ihnen erstellten Dokumente und Objekte am besten kennen und somit grundsätzlich gut beschreiben können. Kritiker dieses Verfahrens machen hingegen geltend, dass die Autoren in der Regel nicht über Fachwissen und Erfahrung auf dem Gebiet der Indexierung und Katalogisierung verfügen, welches notwendig sei, um gute Metadatensätze zu erstellen. Ferner mangele es ihnen häufig an Objektivität in bezug auf ihr Werk, wodurch die Qualität der Metadaten erheblich beeinträchtigt würde. Im Extremfall kann dies zu bewussten Falscheintragungen führen, wie nicht zuletzt der weit verbreitete Missbrauch bei HTML-Meta-Tags zur Erzielung eines guten Suchmaschinenrankings bzw. einer hohen Besuchsfrequenz gezeigt hat.
Unabhängig davon stellt das Erfordernis der Auszeichnung durch die Autoren einen zentralen Unterschied der Erschließung im WWW im Vergleich zur klassischen Erschließung durch unabhängige

Institutionen dar. Dies wirkt sich direkt auf die verwendeten Metadatenstandards und Auszeichnungshilfsmittel aus.

Komplexität der Metadaten

Die Urheber von Ressourcen im WWW sind in der Regel keine Spezialisten auf dem Gebiet der Indexierung, Erschließung oder gar Katalogisierung. Komplexe Metadatenstandards mit umfangreichen Regelwerken wie sie z. B. in Bibliotheken und Museen zum Einsatz kommen, können daher nicht einfach auf den Webkontext übertragen werden. Metadatenformate für das Web müssen vielmehr hinreichend einfach und allgemein verständlich sein, um die Akzeptanz der Nutzer und somit die flächendeckende Anwendung zu gewährleisten. Somit müssen Kompromisse zwischen der Einfachheit und Ausdrucksmächtigkeit eines Standards eingegangen werden. [Miller 1996]

Ausrichtung auf Ressource Discovery

Die Unmöglichkeit der Übernahme komplexer formaler Metadatenstandards z. B. aus dem bibliothekarischen Bereich wird teilweise entschärft, indem Metadaten im WWW in erster Linie das Auffinden von Objekten erleichtern sollen, die den inhaltlichen Kriterien einer Anfrage genügen. Die genaue Unterscheidung einer Ressource von ähnlichen Ressourcen oder anderen Versionen einer Ressource ist von geringerer Bedeutung. Der Informationssuchende gelangt über die Metadatenbeschreibung im allgemeinen ohnehin direkt zu dem damit verlinkten Objekt, von dem er sich ein Bild machen kann. Dagegen spielt bei bibliothekarischen Metadatenstandards eine genaue deskriptive Erschließung eine wesentliche Rolle, um z. B. verschiedene Ausgaben eines Buches zu unterscheiden. Insofern können Metadaten für das Web durch ihre spezifische Ausrichtung einen Teil des Ballastes der traditionellen Erschließungsformate aufgeben. [Eversberg 1999] Kap. 10.7

Vielfalt von Ressourcen im WWW

Bis zum Aufkommen des WWW konnte sich eine Dokumentationseinrichtung in der Regel auf einen charakteristischen Medien- und Objekttyp spezialisieren: z. B. Karten einer Kartographierungsstelle, Videobänder eines Videoarchivs und überlieferungswürdige Dokumente der öffentlichen Verwaltung in staatlichen Archiven. Der Einsatz eines im Detail auf den Bedarf zugeschnittenen Metadatenstandards war möglich. Im WWW werden neue Anforderungen gestellt, da generell eine große Bandbreite von Medientypen (z. B. Text- und Bilddateien, Video- und Audiosequenzen) und innerhalb eines Typs verschiedenste Formen (z. B. abgeschlossene Textdatei versus verlinkte Artikel eines Online-Journals) existieren, die über verschiedene Wissens- und Fachbereiche hinweg genutzt werden. Außerdem sollen einheitliche Sichten auf heterogene Datenquellen mit spezifischen Objekttypen geboten werden. Metadaten für das WWW müssen daher im besonderen Maße dazu geeignet sein, eine einheitliche Beschreibung für Objekte unterschiedlichster Medientypen und –formen zu leisten.

Besonderheiten der vernetzten Umgebung

Der Einsatz der Metadaten in einer vernetzten Umgebung erhöht generell den potentiellen Grad an Wiederverwendung: Die Daten werden nicht mehr nur für den zunächst grundsätzlich isolierten Katalog einer Einrichtung erstellt, sondern können von anderen Metadatenanbietern geladen, mit weiteren Daten angereichert und erneut angeboten werden. Die Verknüpfung von Daten unterschiedlicher Wissens- und

Anwendungsgebiete bietet sich an, was eine Vereinheitlichung auf der Ebene der Metadaten voraussetzt. Die Vernetzung führt also langfristig zur Forderung nach standardisierten Metadaten, die eindeutig wiederverwendet werden können und eine problemlose Verknüpfung von Wissen aus unterschiedlichen Gebieten ermöglichen. Mögliche Ansätze sind einerseits die Schaffung weniger einfacher Standards, die als kleinster gemeinsamer Nenner dienen, und andererseits die Erstellung von Architekturen, die weitgehend automatisierte Abbildungen zwischen spezialisierten Metadatenstandards ermöglichen.

3. Technologien zur semantischen Nutzung des WWW

In Kapitel 2 wurden im Zusammenhang mit dem Erfordernis des Einsatzes von Metadaten bereits Probleme bei der derzeitigen Nutzung des WWW angerissen. Diese beruhen im wesentlichen darauf, dass die herkömmlichen Nutzungsstrukturen, also das benutzergesteuerte Browsen und die Suche mit Hilfe von Suchmaschinen angesichts der entstandenen Informationsfülle nicht mehr ausreichend sind. Browsen, welches auf die Auswertung jeder angezeigten Seite durch den menschlichen Anwender ausgerichtet ist, erweist sich als zu aufwendig; die Präzision, die Suchmaschinen mit Hilfe einer allgemeinen Schlüsselwortsuche erreichen, als zu gering. Ein weiteres Problem ist der steigende Bedarf an Datenaustausch, z. B. im Bereich der Wissenschaft und der Geschäftsbeziehungen zwischen Unternehmen. Der automatisierte, mulitlaterale Austausch wird durch die bisherigen Strukturen und Dienste im Netz jedoch nur unzulänglich unterstützt.

Wesentliche Bemühungen, diesen Problemen zu begegnen, zielen darauf ab, die Semantik der Informationen im Web stärker durch Anwendung geeigneter Sprachen und Konzepte zu nutzen. Nachfolgend wird eine Auswahl zentraler Ansätze in diesem Zusammenhang vorgestellt, die mit der Vorstellung einer möglichen Architektur eines „Semantic Web“ und potentiellen Anwendungsszenarien abschließt. Es wird darauf hingewiesen, dass diese Auswahl keinesfalls vollständig ist: Es sind häufig alternative Konzepte und Variationen vorhanden, deren vollständige Erfassung jedoch den Rahmen dieser Arbeit sprengen würde.

3.1 Allgemeine Metadatenelemente in HTML

Eine grundlegende Form der Erschließung von Semantik von Ressourcen ist die Auszeichnung mit inhaltsorientierten Metadaten. Das vornehmlich als Dokumentenstruktursprache konzipierte HTML sah von Beginn an auch Elemente vor, die auf eine inhaltliche Beschreibung des gesamten Dokumentes abzielten. So erwähnt eine erste Beschreibung von HTML aus dem Jahr 1992 das TITLE-Tag als Attribut eines Hypertextknotens, welches den Inhalt des Knotens mit Hilfe einer Textzeile beschreiben soll. Ein Vorschlag zur Erweiterung von HTML aus dem Jahr 1993 enthält darüber hinaus erstmals sogenannte *Metatags*, die im Header eines HTML-Dokumentes untergebracht werden und dessen Inhalt in Form von Attribut-Wert-Paaren nach dem folgenden Muster charakterisieren:

```
<head>
<meta name="keywords" content="Metadaten, WWW, Semantic Web">
<meta name="date" content="2002-03-04T08:49:37+00:00">
  <!-- ... weitere Angaben des Dateikopfs ... --!>
</head>
```

Die möglichen Attribute werden jedoch nicht vorgegeben, der Vorschlag zählt nur *keywords*, *created* und *owner* als Beispiele auf.
[W3C 2002c]

Diese Form der Metatags wurde in den HTML 2.0 Standard vom November 1995 aufgenommen, wobei ausdrücklich erwähnt wurde, dass es sich um einen „erweiterbaren Container für Meta-Informationen“ handelt, der zur „Dokumentation des Inhalts, der Qualität und der Eigenschaften eines Datensatzes dient

und seine Eignung für den Gebrauch indiziert“. [Berners-Lee, Connolly 1995] Ferner wird darauf hingewiesen, dass es der Implementation des Servers bzw. Browsers überlassen ist, ob und wie sie die Informationen in den Metatags interpretieren.

Abgesehen von der Freiheit bei der Wahl der möglichen Metatags ist auch die Semantik der einzelnen tags nicht näher definiert. Erst der HTML 4.0 Standard vom Dezember 1997 sieht vor, dass „die Bedeutung einer Eigenschaft und die Menge der dieser Eigenschaft zugehörigen Werte in einem Referenzlexikon, das als *profile* bezeichnet wird,“ definiert werden sollte. Dies geschieht, indem der Universal Resource Identifier (URI) des Profils im Header des HTML-Dokuments referenziert wird und dieses somit Gültigkeit für das gesamt Dokument erhält. Daneben wird mit der Einführung des Attributs *scheme* innerhalb eines Metatags die Möglichkeit der Angabe eines Kontextes für Metatag-Werte geschaffen, der die korrekte Interpretation der Werte erleichtern soll. So soll im nachstehenden Beispiel die Angabe von „ISBN“ als scheme-Wert die interpretierende Anwendung dahingehend unterstützen, als diese als Wert des Metatags „identifier“ nun eine International Standard Bibliograhic Number (ISBN) erwarten kann.

```
<meta scheme="ISBN"  name="identifier" content="0-8230-2355-9">
```

[W3C 1998]

Nach [Münz 2001] sind die folgenden inhaltsorientierten Metatags am weitesten verbreitet und werden von den bedeutenden Suchmaschinen in der Regel verarbeitet:

Metatag	Bedeutung
author	Autorenschaft
description	Kurzbeschreibung des Inhalts
keywords	charakteristische Schlagwörter
date	Publikationsdatum

Abbildung 1: Inhaltsorientierte Metatags in HTML

Dies sagt jedoch noch nichts über die Gewichtung der Metadaten im Vergleich zu den sonstigen Relevanzkriterien wie Vorkommen der Schlüsselworte im Volltext, Nutzungshäufigkeit einer Webseite etc. aus. Ferner ist die Auswertung der verschiedenen Metadatenfelder sehr unterschiedlich: Die Metatags *keywords* und *description* werden u. a. von den Suchmaschinen Altavista, HotBot und Infoseek verwendet. Excite beispielsweise konzentriert sich nur auf die *keywords*, während WebCrawler die *keywords* ignoriert und nur *description* verwendet. Daneben nutzen einige Suchmaschinen noch weitere Metatags, die ihren spezifischen Such- und Indexierungskriterien entsprechen. [Sonnenreich, Macinta 1998] Kapitel 2

3.2 Dublin Core: Kernbaustein für Metadatenstandards im WWW

3.2.1 Entwurf und Entwicklung der semantischen Kernelemente

Ein Problem der allgemeinen Metadatenelemente in HTML ist, dass sie zwar die Möglichkeit schaffen, HTML-Dokumente mit Metadaten auszuzeichnen, der Nutzen dieser Metadaten jedoch dadurch relativiert wird, dass weder die Menge der möglichen Metadatenelemente noch deren Semantik festgelegt ist. Der Ansatz der Dublin Core Metadata Initiative (DCMI) geht bei diesen Punkten jeweils einen Schritt weiter.

Ziel der Initiative war es zunächst, eine Kernmenge semantischer Beschreibungselemente (Dublin Core Metadata Element Set, kurz Dublin Core) für webbasierte Ressourcen zu schaffen, die allgemein die Auszeichnung im Web und damit Suche und Retrieval erleichtern sollte. Hierzu wurde im März 1995 in Dublin, Ohio, ein erster Workshop abgehalten, an dem vor allem Teilnehmer aus dem Bereich der Bibliotheken, Archive, Geisteswissenschaften und Geographie teilnahmen. [DCMI 2002] [Weibel 1995]

Es zeichnete sich jedoch schon nach kurzer Zeit ab, dass diese einfache Menge von ursprünglich 13 Beschreibungselementen den unterschiedlichsten Anforderungen, die verschiedene Objekttypen und Communities an einen Metadatenstandard stellen, nicht entsprechen und somit nicht als ausschließlicher Standard für das Web dienen kann.

Ein Hauptziel von Dublin Core ist es daher mittlerweile den gemeinsamen semantischen Kern, der in den verschiedenen Disziplinen fortexistierenden spezifischen Metadatenstandards zu repräsentieren und eine übergreifende Suche zu ermöglichen. Dublin Core soll als der kleinste gemeinsame Nenner verschiedener Standards dienen, der eine semantische Interoperabilität zwischen ihnen ermöglicht. [Weibel 1997]

Folgende weitere Kriterien waren bei Entwurf und Entwicklung von Dublin Core maßgebend:

- **Einfache Handhabung**
  Die Beschreibungselemente wurden bewusst einfach gehalten, um eine größere Akzeptanz im Web zu ermöglichen. Es sollte grundsätzlich jedermann auszeichnen können, ohne Fachwissen oder eine spezielle Ausbildung zu benötigen.

- **Ausrichtung auf Ressource Discovery**
  Hauptziel ist die Suche nach Objekten im WWW zu erleichtern. Dublin Core Metadaten dienen daher in erster Linie der Identifikation und der Inhaltsbeschreibung. Andere Aspekte von Metadaten im Sinne der Aufzählung unter Kapitel 2.1 wie strukturelle Aspekte, Kontext etc. werden nicht oder nur teilweise unterstützt.

- **Beschreibung von *Document-Like Objects***

  Dublin Core wurde primär zur Beschreibung sogenannter Document-Like Objects (DLOs) entwickelt, da man davon ausging, dass dies der vorherrschende Typ von Objekten ist, nach denen im Internet gesucht wird. Nach [Weibel 1995] wurde der Begriff „nicht eindeutig definiert, sondern anhand von Beispielen verdeutlicht. Die elektronische Ausgabe eines Zeitungsartikels oder eines Lexikons ist ein DLO, eine unkommentierte Sammlung von Folien aber nicht.(...) DLOs können beliebig komplex sein, weil sie aus Text mit Verweisen auf Bilder, Audio- oder

Videoclips oder andere Hypertextdokumente bestehen können.“ Eine weitergehende Abgrenzung des Begriffs wurde nur insofern getroffen, als DLOs primär Text enthalten und ihre Metadaten somit den traditionellen Metadaten für gedruckte Texte ähneln sollten.

- **Erweiterbarkeit**
  Da die Kernelemente bewusst einfach gehalten wurden und nur eine partielle Beschreibung der Objekte darstellten, wurde Dublin Core von Beginn an so konzipiert, dass eine Erweiterung in Anwendungsgebieten, die mehr Attribute benötigen, möglich sein sollte. So wurde beispielsweise für das Gebiet Ausbildung inzwischen das zusätzliche Beschreibungselement „audience“ von der Arbeitsgruppe Education der Dublin Core Initiative eingeführt, da bei der Beschreibung von Lernmitteln die Zielgruppe wesentlich erscheint.

- **Internationaler Konsens**
  Die Entwicklung erfolgt unter der Beteiligung einer breiten internationalen Community. Der Kernstandard wurde bereits in zahlreiche Sprachen übersetzt.
  [Weibel 1995]

Das Dublin Core Element Set umfasst derzeit (Version 1.1 vom 07.02.99) folgende Kernelemente:

Dublin Core Element Identificator	Bedeutung / Verwendung
Title	Name der Ressource
Creator	Für die Ressource verantwortlicher Autor (Person oder Organisation)
Subject	Thema der Ressource, typischerweise Schlagwörter, Klassifikationscodes
Description	Inhaltszusammenfassung oder Abstract.
Publisher	Verleger bzw. Herausgeber des Dokuments.
Contributers	An der Entstehung der Ressource beteiligte Person(en) oder Organisation(en).
Date	Datum für einen bestimmten Bearbeitungsstand, typischerweise Erstellung oder Bereitstellung, Verwendung eines Notationsschemas empfohlen.
Type	Art oder Genre der Ressource.
Format	Form der physischen oder digitalen Ausprägung der Ressource (Format, Dateityp)
Identifier	Eine eindeutige Identifikation für die Ressource (URL, ISBN etc.).
Source	Quelle, von der die Ressource abgeleitet ist
Language	Sprache des intellektuellen Inhalts, Verwendung eines Notationsschemas wird empfohlen.
Relation	Referenz zu verwandten Ressourcen
Coverage	Von der Ressource erfasster geographischer oder zeitlicher Bereich.
Rights	Rechtliche Aspekte bezogen auf die Ressource und deren Inhalt.

Abbildung 2: Kernelemente des Dublin Core Element Sets

[DCMI 1999]

Der Dublin Core Standard hatte sich zunächst über die Syntax zur Kodierung dieser Attribute und ihrer zugehörigen Werte ausgeschwiegen. Erst im 4. Workshop im März 1997 wurden Vorschläge dazu gemacht, wie Dublin Core Metadaten in HTML-Dokumenten kodiert werden könnten. Bezogen auf HTML 4.0 sehen diese Vorschläge derzeit vor, dass grundsätzlich die in Kapitel 3.1 besprochenen Metatags verwendet werden, wobei dem Attributnamen jeweils das Präfix „DC.“ vorangestellt wird, um

zu kennzeichnen, dass es sich bei dem Attribut um ein Dublin Core Element handelt. Ein vorangestellter Verweis auf die URL bzw. URN mit den Definitionen des Element Sets stellt außerdem die Verbindung zwischen dem Präfix und dem semantischen Vokabular her, welches es repräsentiert. Ein entsprechend erstellter HTML-Header könnte demnach wie folgt aussehen:

```
<head>
  <title> Knowledge Representation Book </title>
  <link rel     = "schema.DC"
        href    = "http://dublincore.org/documents/dces/">
  <meta name    = "DC.Title"
        content = "Knowledge Representation: Logical, Philosophical
                   and Computational Foundations">
  <meta name    = "DC.Creator"
        content = "Sowa, John F.">
  <meta name    = "DC.Type"
        content = "book">
  <meta name    = "DC.Date"
        content = "1999-08-16">
  <meta name    = "DC.Format"
        content = "text/html">
  <meta name    = "DC.Language"
        content = "en">
</head>
```

[Kunze 1999]

Neben der physischen Unterbringung der Dublin Core Metadaten im Header der HTML-Seite ist es auch möglich, dort mit Hilfe des LINK-REL-tags nur einen Verweis auf die getrennt gespeicherten Metadaten unterzubringen.

Ferner liegen inzwischen weitere Vorschläge vor, wie Dublin Core – auch völlig unabhängig von HTML - kodiert werden kann. Eine Möglichkeit ist beispielsweise die Verwendung von XML unter Benutzung des RDF-Datenmodells. [Kokkelink, Schwänzl 2001]

3.2.2 Qualifizierung von Dublin Core

Die praktische Anwendung von Dublin Core zeigte bald, dass die Kernelemente häufig keine ausreichende Tiefe der Beschreibung ermöglichen. So lässt z. B. das Element *Relation* offen, welche Art von Beziehung zu einer anderen Ressource vorliegt. Nach heftigen Diskussionen darüber, ob der Beibehaltung der Einfachheit oder zusätzlicher Aussagekräftigkeit der Vorrang zu geben sein, wurde mit Hilfe von sogenannten *Qualifiers* die Möglichkeit geschaffen, die Semantik einzelner Elemente näher zu qualifizieren bzw. diese in einen Interpretationskontext zu stellen.

Derzeit werden zwei Klassen von Qualifiers unterschieden:

- **Element Refinement**
  Diese Qualifiers spezifizieren die Bedeutung eines Elements genauer. So kann beispielsweise für das Element *Date* durch Angabe der Qualifiers *Created*, *Available*, *Modified* etc. näher bestimmt werden, welcher Stand des Objekts mit der Datumsangabe beschrieben wird.

- **Encoding Scheme**
  Diese Qualifiers geben Schemata an, die die Interpretation eines Elementwertes erleichtern. Dabei kann es sich um kontrollierte Vokabulare, formale Notationen oder Parsing-Regeln handeln. Der Wert eines *Subject*-Elementes kann z. B. mit der Angabe von *LCSH* qualifiziert sein, was bedeutet, dass die verwendeten Schlagwörter aus dem kontrollierten Schlagwortvokabular der Library of Congress, den Library of Congress Subject Headings (LCSH) stammen.

Eine Liste der momentan empfohlenen Qualifiers findet sich unter [DCMI 2000].

Ein wesentliches Prinzip der Qualifiers in Dublin Core ist, dass sie die Grundbedeutung des Elementes, in dem sie verwendet werden, nicht verändern sollen. Das qualifizierte Element soll in seinem ursprünglichen Sinn verstanden werden können, auch wenn der Qualifier nicht ausgewertet werden kann. Qualifiers können die Semantik eines Elementes folglich nur beschränken oder verfeinern, aber nicht so verändern, dass es nicht mehr konsistent mit seiner ursprünglichen Definition ist.
[DCMI 2000]

Mit diesem Prinzip wird versucht dem Spannungsverhältnis zwischen einfacher Handhabung und hoher Interoperabilität einerseits und Ausdrucksmächtigkeit andererseits Rechnung zu tragen: Ein nur aus unspezifizierten Kernelementen bestehendes Dublin Core ist nicht nur für die Autoren einfach zu handhaben, sondern stellt auch geringe Anforderungen an die verarbeitenden Systeme und erleichtert somit den Datenaustausch zwischen verschiedenen Domänen. Befürworter einer Qualifizierung machen hingegen geltend, dass die Suchpräzision in Metadaten, die sich auf die Kernelemente beschränken, kaum über die einer Volltextsuche hinausgeht. Ferner können solche Metadaten selten sinnvoll wiederverwendet und weiterverarbeitet werden. Der Aufwand der Auszeichnung mit unqualifiziertem Dublin Core wäre daher aufgrund seiner geringen Ausdrucksstärke kaum gerechtfertigt.

Um Mehrdeutigkeiten durch die Verwendung von Qualifiers zu reduzieren schlägt die Dublin Core Metadata Initiative außerdem die Etablierung von *usage boards* in den jeweiligen Anwendungscommunities vor, die die Verwendung der Qualifiers überwachen und deren Interoperabilität in sogenannten *registries* festhalten. [Lagoze 2000] S. 15

3.2.3 Sachstand und Kritik

Dublin Core ist es mittlerweile gelungen, sich als semantisches Grundmodell für beschreibende Metadaten vorwiegend textueller Ressourcen zu etablieren. Bei der Entwicklung neuer Metadatenstandards für Fachgemeinschaften wird häufig explizit auf Dublin Core aufgebaut bzw. auf entsprechende Abbildungsmöglichkeiten geachtet. Ferner verwenden zahlreiche Server für Online-Publikationen Dublin Core Metadaten für die Erfassung neuer Werke und für die Suche im Publikationsbestand. In Australien, Dänemark, Finnland und Großbritannien ist die Verwendung von Dublin Core zur Beschreibung von elektronisch veröffentlichten Regierungsinformationen zwingend vorgeschrieben [Weibel 1999]. Einen Überblick über die Vielzahl der Anwendungen und Projekte, die Dublin Core verwenden, geben beispielsweise [CEN 2000], [DCMI 2002b].

Zum ausschließlichen allgemeinen Standard für die Auszeichnung und Suche im Web ist Dublin Core hingegen nicht geworden. So zählt die Initiative selbst nur fünf eher unbekannte Suchmaschinen auf, die ihr Element Set derzeit auswerten [DCMI 2002 c] Nr. 22. An der erwähnten beschränkten und uneinheitlichen Berücksichtigung von Metadaten im Suchmaschinenkontext hat sich darüber hinaus auch durch die Einführung von Dublin Core nichts geändert [Rusch-Feja 1997] S. 117, [Clark 1999], [INT Media 2002].

Einer der Hauptkritikpunkte an Dublin Core ist, dass es nicht als eigenständiger Erschließungsstandard geeignet sei. Dies ist zum einen darauf zurückzuführen, dass die Semantik hinter den sehr breiten Definitionen für die Kernelemente nicht genau genug bestimmt ist und diese in verschiedenen Anwendungen und Kontexten daher äußerst unterschiedlich verstanden und benutzt werden. Der Versuch dies durch Qualifizierung zu beheben, ohne die erwünschte vermittelnde Eigenschaft des gemeinsamen semantischen Modells zu verlieren, kann nur begrenzt gelingen, wie [Lagoze 2000] ausführlich darstellt.

Eine weitere Schwachstelle ist dadurch gegeben, dass in den Attributwerten grundsätzlich keine Struktur vorhanden ist. So wurden zunächst weder Formate für die Elementwerte, noch irgendwelche Regeln oder Referenzen für deren mögliche Inhalte vorgegeben. Durch die entsprechende Beliebigkeit der Einträge wird die Suchpräzision erheblich verschlechtert. Etablierte Metadatenstandards machen an dieser Stelle bewusst von kontrollierten Vokabularen und detaillierten Regelwerken Gebrauch, um ihre Datenbestände konsistent zu halten und eine hohe Trennschärfe zu bewahren. [Lynch 1998], [Eversberg 1999] Kap. 10.7

Diese Schwachpunkte sind u. a. daran ersichtlich, dass mittlerweile zwar zahlreiche bibliothekarische Einrichtungen Dublin Core Metadaten für Publikationen aus dem Webkontext übernehmen, jedoch in der Regel gezwungen sind, diese Daten nachzubearbeiten um sie auf das Niveau ihrer Kataloge zu bringen („Hochkatalogisierung“), bevor sie weiterverwendet werden können. Entsprechend wurde in einer Studie der American Library Association festgestellt, dass Metadaten in der Form von Dublin Core Katalogformate nicht ersetzen können. [Eversberg 1999] Kap. 10.7

Weitere Kritikpunkte sind die starke Ausrichtung auf Document-Like Objects, wodurch anderen Ressourcen, wie z. B. Audio- und Videofiles ungenügend Rechnung getragen wird. Ferner wird die Beschreibung von Objekte, die ihrerseits Sammlungen von Objekten darstellen, und der Austausch der Metadaten kaum von Dublin Core unterstützt.

Die Dublin Core Metadata Initiative reagiert auf diese Probleme insofern, als sie zunehmend den Einsatz von kontrollierten Vokabularen und Schemas im Zusammenhang mit den Attributwerten empfiehlt (vgl. entsprechenden Hinweise zum aktuellen Element Set [DCMI 1999]). Ferner definiert sie ihre Zielsetzung nun in stärkerem Maße in Richtung „Erleichterung der Entwicklung von community- und fachbereichsbereichsspezifischen Metadatenstandards (...)“ und deren Interoperation [Dekkers, Weibel 2002].

Zusammenfassend darf wohl angenommen werden, dass dies tatsächlich die zentrale Errungenschaft und zukünftige Perspektive von Dublin Core ist: Der gemeinsame semantische Grundkonsens, auf den spezialisierte Metadatenstandards aufbauen und mit dessen Hilfe sie auf kleinster gemeinsamer Basis integriert werden können. Dem Anspruch, im Zusammenhang mit Resource Discovery alles mit einer

einheitlichen und zudem sehr einfachen Grammatik zu beschreiben, ohne dabei wesentlich an Ausdrucksmächtigkeit zu verlieren, kann Dublin Core ebenso wenig gerecht werden wie vergleichbare frühere Versuche.

3.3 XML: Metasprache für semantische Auszeichnungsvokabulare

3.3.1 Ursprung als generische Auszeichnungssprache

Einen weiteren wesentlichen Baustein auf dem Weg zur semantischen Nutzung des WWW stellt die Metasprache *Extensible Markup Language (XML)* dar. Im Gegensatz zu Dublin Core definiert XML keine semantischen Beschreibungselemente, sondern gibt dem Benutzer die Möglichkeit, diese anwendungsspezifisch selbst zu definieren, wobei die entstehenden Datenstrukturen leicht zu erweitern, auszutauschen und zu bearbeiten sind.

XML ist eine Untermenge der *Standard Generalized Markup Language (SGML)*, die 1986 zum ISO Standard wurde und den vorläufigen Abschluss in der Entwicklung generischer Auszeichnungssprachen seit Ende der 60er Jahre darstellte. Leitgedanke dieser Entwicklung ist die Schaffung von anwendungsspezifischen Sprachen, mit denen Texte unabhängig vom Layout in logische und inhaltliche Informationseinheiten codiert werden können. Dadurch ist eine Trennung der jeweiligen Informationsstruktur vom optischen Erscheinungsbild möglich. SGML ist folglich eine Metasprache zur Generierung von Auszeichnungssprachen. [Gennusa 1999] S. 29 f, [Behme 2000] S. 34ff

XML als Sprache wurde seit 1996 in einer Arbeitsgruppe des World Wide Web Consortiums (W3C) unter Beteiligung von SGML- und Internet-Experten entwickelt. Im November 1996 wurde der erste Entwurf vorgestellt; seit Februar 1998 hat XML 1.0 den Status einer W3C-Empfehlung. Hauptmotivation für den Entwurf der neuen Sprache war die negative Entwicklung von HTML, welches aufgrund der firmenspezifischen Erweiterungen durch die Browserhersteller die gemeinsame Basis zu verlieren und in mehrere proprietäre Teile zu zerfallen drohte. Andererseits erschien SGML zu komplex und aufwendig in Implementation und Unterhaltung, um im WWW Aussicht auf Erfolg zu haben. Mit XML wurde daher eine schlanke Version von SGML geschaffen, die einfach und verständlich bei gleichzeitiger Wahrung von ausreichender Funktionalität sein sollte, um für eine große Zahl von Anwendungen in Frage zu kommen. Im Gegensatz zu HTML, welches eine SGML-Anwendung darstellt, ist XML wie erwähnt eine Untermenge von SGML. D. h. HTML ist eine konkrete Auszeichnungssprache dessen Vokabular mit SGML festgelegt wurde. XML dient dagegen selbst zur Definition von Auszeichnungssprachen. Neben dieser Bedeutung von XML als Metasprache im engeren Sinne wird der Begriff inzwischen häufig stellvertretend für einen Vielzahl von Konzepten und Sprachen verwendet, die mit XML definiert wurden oder dieses ergänzen. [Gennusa 1999] S. 34, [Behme 2000] S. 41

Die Praxis im Webkontext zeigt, dass sich XML auch sehr gut zur Beschreibung semistrukturierter Daten eignet. Hierbei kann die grundlegende Datenstruktur von XML als geordneter Baum verstanden werden, wobei jedes Auszeichnungselement (Tag) einem beschrifteten Knoten entspricht. Der Zugriff kann über die einheitlichen Schnittstellen *SAX (Simple API for XML)* oder *DOM (Document Object Model)* erfolgen. Bei der letzteren handelt es sich um einen umfangreichen objektorientierten Co-Standard des W3C zu XML.
[Abiteboul et al 1999] S. 32f, 49f

Aufgrund dieser Eigenschaft wird XML im WWW nicht nur zur Definition von Dokumenttypen im Sinne generischer Auszeichnungssprachen, sondern auch zur Speicherung und insbesondere zum Austausch von Daten verwendet. Der Datenteil des XML-Dokumentes wird aus Datenbanksicht damit gewissermaßen zur Datenbasis. Es wird die Auffassung vertreten, dass die Unterscheidung zwischen Dokument und

Daten in diesem Zusammenhang verschwinde, da sie „tatsächlich zwei Zustände derselben Information“ ([Goldfarb 1999] S. 15) seien bzw. die Vision von XML sei, „die Sprache als Basis für jede Art von Daten zu etablieren“ ([Pott, Wielage 1999] S. 18). Im weiteren werden beide Nutzungsformen von XML berücksichtigt werden, soweit sie vorliegend von Interesse sind.

3.3.2 Eigenschaften

Eine genaue Beschreibung des Sprachumfangs und der Syntax von XML würde den Rahmen dieses Überblicks sprengen. Diesbezüglich muss auf die Spezifikation bei [W3C 1998a] verwiesen werden. Nachfolgend sollen jedoch die Eigenschaften von XML näher beleuchtet werden, die besonders zu seiner Bedeutung für die semantische Nutzung des WWW beitragen.

Einfachheit

Ebenso wie für Metadatenstandards gilt auch für Sprachstandards im WWW, dass sie hinreichend einfach sein müssen, um akzeptiert zu werden. XML erfüllt diese Voraussetzung, indem es im Datenteil in der einfachsten Form nur Element-Wert-Paare vorsieht, die optional geschachtelt, wiederholt und mit Attributen versehen werden können.

Beispiel:

```
<person idnr="112145086">
        <name>
                <firstname>John</firstname>
                <lastname>Sowa</lastname>
        </name>
        <title>Conceptual structures: Information processing in mind and machine</title>
        <title>Knowledge Representation: Logical, Philosophical and Computational Foundations</title>
</person>
```

Der Vorteil dieses einfachen Aufbaus ist u. a., dass er sowohl für den menschlichen Leser leicht verständlich als auch für einen Parser interpretierbar ist. XML kann in der erweiterten Backus-Naur-Form (EBF) ausgedrückt und somit effizient geparst werden [W3C 1998a]. Ferner wird die Akzeptanz im WWW dadurch gefördert, dass die für die Auszeichnung verwendete Tagsyntax zahlreichen Anwendern bereits durch HTML bekannt ist. Außerdem stellt die Textorientierung sicher, dass neben ausgefeilteren Hilfsmitteln auch eine Bearbeitung mit einem einfachen Editor möglich ist.

Semantische Auszeichnung

XML gibt dem Anwender die Möglichkeit ein semantisches Vokabular für die in seinem Anwendungsgebiet benötigten Daten- und Dokumenttypen zusammenzustellen und diese entsprechend auszuzeichnen. Informationen können nun qualifiziert, d. h. anhand der semantischen Bezeichner gesucht, gefiltert, extrahiert, referenziert und verarbeitet werden. Im obigen Beispiel wäre also eine gezielte Suche nach den Werken des Autors anhand des <title>-Tags, anstelle einer Schlüsselwortsuche im gesamten

Text möglich. Auch das bei HTML angewandte „Screen scraping“, also der Versuch, mit Hilfe von Layoutinformationen Rückschlüsse auf den Inhalt zu ziehen, entfällt.

Ausdrucksmächtigkeit und Erweiterbarkeit

Das zur Verfügung stehende Vokabular ist nicht vorgegeben oder beschränkt. Alles was mit Hilfe einer Grammatik beschrieben werden kann, kann in XML kodiert werden. Somit können Auszeichnungssprachen beliebigen Umfangs und für nahezu beliebige Anwendungen geschaffen werden. Durch die Möglichkeit der Schachtelung und Verknüpfung können insbesondere auch komplexe Objekte beschrieben werden. Die Erweiterung eines bereits bestehenden Vokabulars ist aufgrund der flexiblen Datenstruktur leicht möglich. Sofern erforderlich können die zugehörige Dokumenttyp-Deklaration rasch editiert und Datenteile entsprechend angepasst werden. Dies stellt auch einen der Vorteile von XML für die Bereitstellung sich häufig ändernder Daten für das WWW gegenüber den eher starren Strukturen dar, die in Datenbanken nach Festlegung der Schemata gegeben sind. [Pott, Wielage 1999] S. 18

Syntaktische Interoperabilität

Trotz der universellen Ausdrucksmächtigkeit stellt die Struktur, die die Metagrammatik von XML vorgibt, sicher, dass ein allgemeiner XML-Parser Dokumente beliebige XML-Daten parsen kann. Somit sind XML-Parser unabhängig von der Anwendungsdomäne wiederverwendbar, wodurch der Aufwand zur Gewinnung von anwendungsspezifischen Datenrepräsentationen aus XML-Dokumenten gering gehalten wird. Dadurch wird eine hohe syntaktische Interoperabilität gewährleistet.

Flexibilität und Kostenreduktion durch Trennung von Inhalt und Form

Da mit XML die Möglichkeit zur inhaltsorientierten Auszeichnung besteht, muss bei Speicherung, Verwaltung und Austausch der Dokumente bzw. Daten zunächst keine Rücksicht mehr auf Darstellungsformate genommen werden. Daten sind die tatsächlichen Inhalte, die bei Bedarf z. B. mittels Baumtransformationssprachen wie *XSL (e**X**tensible **S**tylesheet **L**anguage)* in die gewünschte Darstellungsform gebracht werden können. Dadurch sind flexible Konvertierungen in verschiedene Ausgabeformate, wie beispielsweise unterschiedliche Dokumentenformate, aber auch medienspezifische Ausgaben wie Bildschirmanzeige, Druck, Blindenschrift etc. möglich. Ferner können Daten kostengünstig - ohne Darstellungsoverhead – ausgetauscht werden. In einer vernetzten Umgebung kann dies genutzt werden, um den Aufwand für die Präsentation der Daten zum Client zu verschieben. Über das Netz werden nur noch die Nettoinformationen übertragen. Wenn Daten mehrfach wiederverwendet bzw. in mehrstufigen Prozessen verarbeitet werden, wie dies oft bei Metadaten für Webobjekte der Fall ist, verringert sich ferner die Gefahr von Fehlern, da das Layout bis zum Ende der Prozesskette vom Inhalt getrennt bleibt.
[Pott, Wielage 1999] S. 18, [Goldfarb 1999] S. 13f

Plattformunabhängigkeit und Internationalität

Die Trennung von Inhalt und Form trägt in Verbindung mit der Textorientierung dazu bei, dass XML-Dokumente weitgehend unabhängig von bestimmten Systemplattformen und anwendungsspezifischen

Speicherungsformen sind. Basis sind die bereits vorhandenen Internettechnologien; der Einsatz proprietärer Software ist grundsätzlich nicht erforderlich. Da XML darüber hinaus als offener Standard vom W3C gepflegt wird, sind die Voraussetzungen für eine breite Anwendung im WWW gegeben.
[Pott, Wielage 1999] S. 18

XML benutzt Unicode bzw. ISO 10646 als Standardzeichensatz und kann damit grundsätzlich die Zeichen sämtlicher derzeit bekannter Sprachen kodieren [W3C 1998a].

Validierbarkeit von XML-Dokumenten

Durch Definition der Auszeichnungselemente und ihrer Beziehungen in einem Schema (entweder Dokumenttyp-Deklaration oder XML-Schema) wird ein Dokumenttyp definiert. Wird ein konkretes Dokument mit diesem Schema verknüpft, erhält es den entsprechenden Typ zugewiesen. Ein validierender Parser kann nun überprüfen, ob ein Dokument mit dem zugewiesenen Schema konform und somit valide ist. Dadurch kann auf einfache und wirksame Weise geprüft werden, ob eine für einen Anwendungsbereich geschaffene Auszeichnungssprache formal eingehalten wird. Diese Eigenschaft von XML ist v. a. zur Gewährleistung von Konsistenz beim Datenaustausch von Bedeutung.

Erweiterte Verknüpfungsstrukturen

Mit Hilfe der Extensible *Linkage Language (XLink)*, die als Komplementärsprache zu XML entwickelt wurde, sind im Gegensatz zum HTML-Verknüpfungskonzept auch multidirektionale Verweise und Verweise mit mehreren Verweiszielen möglich. Außerdem sind sogenannte *Out-of-Line*-Verweise vorgesehen, die dadurch gekennzeichnet sind, dass die Verweise in keinem der verknüpften Dokumente physisch präsent sind. Sie stehen vielmehr in einem separaten Dokument und regeln von dort aus die Verknüpfung. In Verbindung mit der Adressierungssprache *XML Pointer Language (XPointer)*, die das gezielte Ansprechen interner Dokumentbestandteile erlaubt, ergeben sich durch dieses Verknüpfungskonzept weitreichende Möglichkeiten, Dokumente aus anderen Dokumenten zusammenzusetzen bzw. Datenquellen im WWW zu integrieren. Ferner können Ressourcen im WWW von Dritten mit Annotationen versehen oder in Zusammenhang gebracht werden, ohne sie editieren zu müssen. Weitere Möglichkeiten eröffnen sich für die Personalisierung und Selektion: Zu einem XML-Dokument können mehrere Verweisnetze aufgebaut werden, die in Abhängigkeit von Benutzerinteressen oder Applikationsanforderungen bereitgestellt werden. Die Spezifikation zu *XLink* wurde am 27.06.2001 als W3C-Empfehlung verabschiedet, für *XPointer* liegt bislang nur eine Kandidatenempfehlung vor. Ferner wurden bisher nur Teile der Konzepte in Forschungsprojekten implementiert und aktuelle Browserversionen unterstützen – wenn überhaupt – nur einfache *XLink*-Verweise.
[Lowe, Wilde 2001], [W3C 2002]

Namensräume

In XML ist es möglich, die in einem Dokument verwendeten Elementtypen und Attribute durch Verweis auf Sammlungen von eindeutigen Namen zu qualifizieren. Der Verweis erfolgt durch Angabe der URI der betreffenden Sammlung, die in der Regel auch Definitionen der Namen enthält. So könnte beispielsweise

durch einen Verweis auf die Referenzbeschreibung des Dublin Core Element Sets das in einem XML-Dokument verwendete *title*-Tag als das entsprechende Dubline Core Element identifiziert werden.

Beispiel:

```
<ht:html xmlns:ht=http://www.w3.org/TR/REC-html40
          xmlns:dc="http://dublincore.org/documents/dces/">
  <ht:head><ht:title>Suchergebnis</ht:title></ht:head>
  <ht:body>
        <dc:creator>Sowa, John F. </dc:creator>
        <dc:title>Knowledge Representation: Logical, Philosophical and Computational
                Foundations</dc:title>
        <dc:date>1975</dc:date>
  </ht:body>
</ht:html>
```

Durch Qualifikation der Namen können Mehrdeutigkeiten, die durch zufällige Übereinstimmung von Bezeichnern entstehen, aufgelöst werden. Auf diese Weise entsteht im obigen Beispiel kein Konflikt zwischen dem HTML-Tag *title* für den Dokumentitel und den im selben Dokument verwendeten Element *title* im Sinne des entsprechenden Dublin Core Elementes. Dieses Konzept ist vor allem deswegen von Bedeutung, weil es die kombinierte Nutzung von Metadatenelementen aus verschiedenen Anwendungen erlaubt. Somit können auch verschiedene Metadatenstandards in einem XML-Dokument integriert werden.
[Behme 2000]

Die zusätzlichen Einsatzgebiete von XML im Bereich der Datenhaltung und des Datenaustausches führte u. a. dazu, dass höhere Anforderungen an die mittels XML formulierbaren Grammatiken gestellt wurden. Daher wurde seit 1999 *XML Schema* als Alternative zu den Dokumenttyp-Deklarationen (DTD) entwickelt. *XML Schema* zeichnet sich gegenüber DTDs dadurch aus, dass die Grammatikregeln selbst in XML-Syntax formuliert werden und eine Vielzahl an Datentypen, sowie einen Mechanismus zur Definition neuer Datentypen vorgesehen ist. Ferner ermöglicht *XML Schema* als echter Schemamechanismus reichere grammatikalische Beziehungen wie z. B. Vererbung und die verbesserte Unterstützung von Namensraummechanismen. Für XML Schema existiert seit dem 02.05.2001 eine Spezifikation in Form einer W3C-Empfehlung.
[W3C 2002a]

3.3.3 Anwendung

Mittlerweile ist eine Vielzahl von Auszeichnungssprachen entstanden, die mit Hilfe von XML definiert wurden. Prominente Vertreter sind z. B. Mathematical Markup Language (MathML) zur Codierung mathematischer Ausdrücke, Chemical Markup Language (CML) zur Auszeichnung wissenschaftlicher Dokumente in der Chemie und Vector Markup Language (VML) für die Kodierung von Vektorinformationen. Einen Überblick über weitere Auszeichnungssprachen auf Basis von XML gibt u. a. [Cover 2002].

Daneben führte die erwähnte Eignung von XML für die Standardisierung des Datenaustausches zu einer starken Verbreitung im Bereich des E-Commerce, insbesondere im Business-to-Business-Sektor. So zitieren [Waldt, Drummond 2002] aus einer Studie der Gartner Group, derzufolge für 2003 erwartet wird,

dass 80 % der Kommunikation zwischen Anwendungen über öffentliche Netzwerke von XML gesteuert wird. Herkömmliche Datenaustauschformate des E-Commerce wie Electronic Data Interchange werden durch XML-Anwendungen ersetzt oder mit diesen in neuen Standards kombiniert. [Gennusa 1999] S. 38, [Huemer 2001] S. 13
Weitere kommerziell bedeutsame XML-Anwendungen sind die Beschreibung von Diensten im WWW und medienunabhängiges elektronisches Publizieren. Hierbei nimmt XML oft die Rolle einer Mittelschicht zwischen verschiedenen Datenquellen und dem endgültigen Darstellungsformat ein. [Goldfarb 1999] S. 14f, [Behme 2000] S. 48f

Die unmittelbare Nutzung als Dokumentenstruktursprache in dem Sinne, dass Dokumente mit Hilfe eines XML-Vokabulars in logische und inhaltliche Informationseinheiten aufgeteilt werden und somit bereits in inhaltlich erschlossener Form im WWW zur Verfügung stehen, scheint gegenüber den anderen Nutzungsformen in den Hintergrund zu treten. So spielt XML auch im Bereich der digitalen Bibliotheken eine große Rolle. Statt der Auszeichnung der Dokumente selbst werden jedoch häufig Metadatenmodelle, die die Dokumente beschreiben, mit Hilfe von XML formuliert. Dadurch können die Möglichkeiten von XML und seinen Co-Standards genutzt werden, um diese Metadaten zu speichern, zu transformieren und auszutauschen. Ein Beispiel für ein derartige Anwendung von XML ist das Datenmodell *DLmeta* [DLMI 2000] welches vom Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) dem Lehrstuhl für Organisation und Management von Informationssystemen (LOMI) der Universität Ulm und dem Zentrum für Datenverarbeitung (ZDV) der Universität Tübingen entwickelt wurde. Ferner kann das Datenformat *METS* der Digital Library Federation [LoC 2001] und das Austauschprotokoll für Metadaten der Open Archives Initiative (OAI) [OAI 2001] in diesem Zusammenhang genannt werden.

Charakteristisch für diese Metadatenmodelle ist, dass der Inhalt der einzelnen Beschreibungselemente auf einem semantischen Metadatenstandard, in der Regel Dublin Core, aufbaut. Dies scheint auch die zentrale Bedeutung von XML im Bereich der Inhaltserschließung im Web darzustellen: Es liefert eine sehr flexible, aber dennoch validierbare und praktisch verwendbare Syntax, mit der inhaltliche Beschreibungselementen, die auf konkreten semantischen Metadatenstandards beruhen, kodiert werden können. Andererseits eignet sich die Syntax aber auch zur Serialisierung von abstrakten Metadatenmodellen, wie beispielsweise dem nachfolgend beschriebenen *Resource Description Framework (RDF)*.

3.4 RDF: Standardrepräsentationssprache für Metadaten

3.4.1 Motivation und Ursprung

Wie im vorangegangenen Unterkapitel dargestellt liegt die Stärke von XML in seiner syntaktischen Interoperabilität. Voraussetzung zahlreicher Konzepte zur effektiveren Nutzung des WWW ist jedoch eine breitere gemeinsame semantische Basis. Diesbezüglich weist XML vor allem folgenden Mangel auf: Es trennt zwar den Inhalt eines Dokumentes vom Layout. Inhalt und Struktur, die durch die Grammatik des Dokumentenschemas vorgegeben wird, bleiben jedoch verbunden. Nun können verschiedene Grammatiken benutzt werden, um einen Sachverhalt zu beschreiben und die Semantik eines einzelnen Datenelementes steht in XML nicht grundsätzlich fest. Daher ist es nicht oder nur sehr eingeschränkt möglich, XML-Datenstrukturen auszuwerten, deren Struktur nicht im vorhinein bekannt ist.

[Decker 2000], [Berners-Lee 1998]

Mehr rechnerinterpretierbare Semantik ist nach Meinung des W3C mit Hilfe des *Ressource Description Framework (RDF)* zu erreichen. RDF ist im wesentlichen eine Sprache, die die Repräsentation von Semantik bei der Ressourcenbeschreibung ermöglicht und eine Infrastruktur zur Nutzung verteilter Metadatenvokabulare zur Verfügung stellt.

Die Ursprünge von RDF gehen auf Arbeiten des W3C an der *Platform for Internet Content Selection (PICS)* im Jahre 1995 zurück. Bei PICS handelt es sich um ein Rating-System, welches es ermöglichen soll, Inhalte von Ressourcen im WWW zu bewerten und diese Bewertungen – beispielsweise zum Schutz Minderjähriger beim Browsen – zu nutzen. Dabei sollten die Bewertungskriterien nicht allgemein und zentral fixiert werden, sondern durch jede Nutzergruppe individuell vergeben werden. Dadurch stellte sich bei der Arbeit an PICS die Frage der Schaffung einer allgemeinen Architektur, mit deren Hilfe Ressourcen mit Metadaten versehen werden können. Diese Problemstellung wurde in Arbeitsgruppen des W3C unter Beteiligung von kommerziellen Anbietern und Vertretern der Metadaten-Communities aufgegriffen und resultierte u. a. in der W3C-Empfehlung zu RDF vom 22.02.1999.
[Miller 1998]

3.4.2 Datenmodell

Nachstehend wird das Grunddatenmodell von RDF kurz vorgestellt. Das zentrale Element sind Ressourcen, zu deren Beschreibung RDF dient. Als Ressourcen werden in diesem Modell alle Objekte bezeichnet, die mit Hilfe einer URI eindeutig identifiziert werden können. Da dies auf alle Entitäten zutrifft, die überhaupt abstrakt bezeichnet werden können, handelt es sich um einen sehr weiten Begriff, der dem des Objektes in der Logik entspricht. Eine Ressource kann also beispielsweise eine Website, ein einzelnes XML-Datenelement, aber auch ein Gegenstand sein, der nicht direkt über das WWW zugänglich ist, wie etwa die Druckausgabe eines Buches. Mit Ressourcen werden Eigenschaften (properties) verknüpft, die bestimmte Werte (values) besitzen. Ein solches Tripel wird als *Statement* bezeichnet. Die Werte eines Statements können selbst wieder Ressourcen oder atomare Werte eines einfachen XML-Datentyps sein. Das Datenmodell kann mit Hilfe eines gerichteten, beschrifteten Graphen repräsentiert werden.

Beispiel: Darstellung der Aussage
„Autor der Ressource http://users.bestweb.net/~sowa/krbook/index.html ist John F. Sowa“. Ressourcen sind als Ovale und atomare Werte als Rechtecke dargestellt.

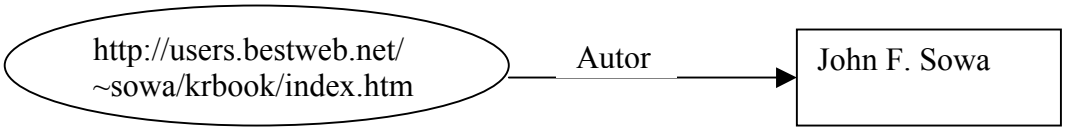


Abbildung 3: Beispiel für ein RDF-Statement

Falls der Wert einer Eigenschaft strukturiert sein soll, wird eine sogenannte anonyme Zwischenressource als Wert eingefügt, deren Eigenschaften die Bestandteile der strukturierten Eigenschaft sind.

Beispiel: Wird das vorherige statement um die Angabe der email-Adresse des Autors erweitert, so ergibt sich die Darstellung in Abbildung 4:

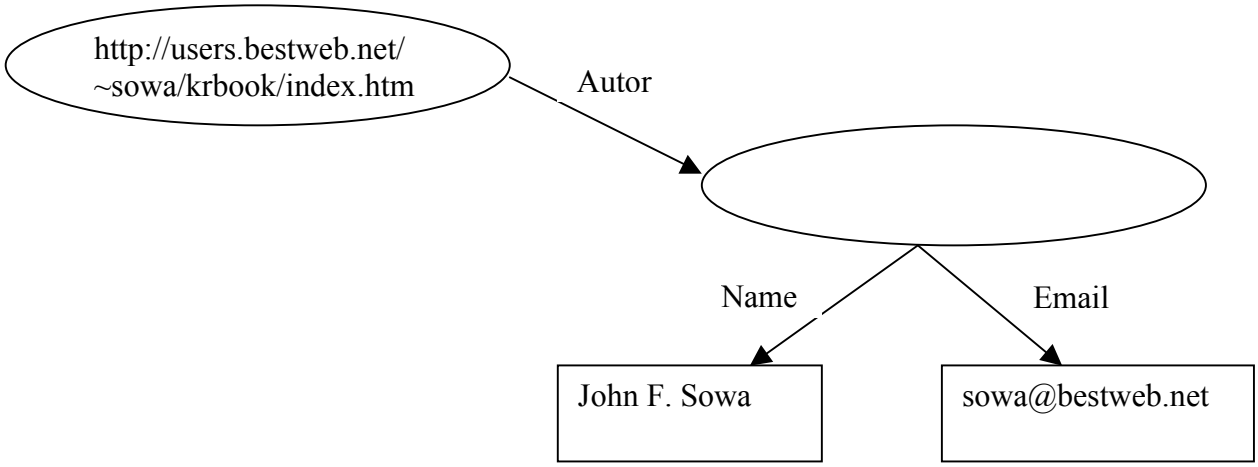


Abbildung 4: Beispiel für einen strukturierten Wert in einem RDF-Statement

Anonyme Ressourcen erhalten anwendungsintern URIs, um ihren Charakter als Ressource zu wahren und die interne Identifikation zu ermöglichen. Für eine von außen zugreifende Anwendung ist dies transparent.
RDF erlaubt Reifikation, d.h. die Umwandlung von Aussagen in Objekte, um Aussagen über Aussagen machen zu können. Hierzu wird das ursprüngliche Statement, über das eine Aussage gemacht werden soll, zur Ressource gemacht, deren Eigenschaften die Bestandteile des ursprünglichen Statements sind. Ferner kennzeichnet die zusätzliche Eigenschaft *rdf:type* mit dem Wert *rdf:statement*, dass es sich bei der neuen Ressource um ein Statement handelt.

Beispiel: Abbildung 5 stellt den Graphen zu folgender Aussage dar:
"Joe Metz sagt, John F. Sowa ist der Autor der Ressource http://users.bestweb.net/~sowa/krbook/index.htm."

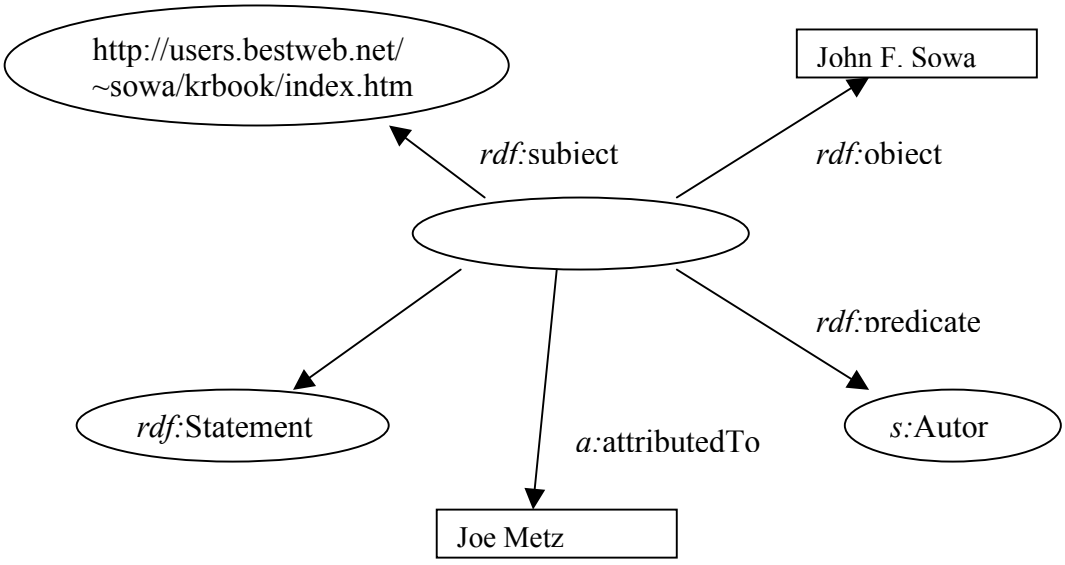


Abbildung 5: Beispiel für eine Reifikation in RDF

Weiterhin sieht RDF Sammlungen von Ressourcen oder atomaren Werten vor. Hierzu werden die „Container"-Konstrukte *bag*, *sequence* und *alternative* zur Verfügung gestellt, die für eine Multimenge, eine geordnete Liste und Alternativen zu einem Einzelwert stehen. Für nähere Angaben hierzu bzw. weitere Details zum Datenmodell wird auf die Spezifikation unter
[W3C 1999] verwiesen.

3.4.3 RDFS: Definition von Auszeichnungsvokabularen

Das vorgestellte Grunddatenmodell stellt kein Vokabular für die Beschreibung der Ressourcen zur Verfügung. Es wird nicht festgelegt, welche Attribute verwendet werden können und welche Beziehungen zwischen diesen Attributen und anderen Ressourcen bestehen. Hierfür wurde beim W3C *RDF Schema (RDFS)* entwickelt, zu dem seit März 2000 eine Spezifikation in Form einer Kandidatenempfehlung des W3C vorliegt [W3C 2000]. Bei RDFS handelt es sich um eine Schemaspezfikationssprache, die es erlaubt, Attribute und Klassen von Ressourcen, auf die die Attribute angewandt werden können, zu definieren. Außerdem können die Wertbereiche von Attributen und die Anwendbarkeit von Attributen auf bestimmte Klassen beschränkt werden. RDFS stellt somit ein grundlegendes Typsystem zur Nutzung innerhalb des RDF-Modells zur Verfügung.

Hierzu werden einige grundlegende Begriffe eingeführt. So ist z. B. *Class* der Basistyp aller im Typsystem verwendeten Klassen, mit Hilfe des Attributes *type* können Instanzen von einer oder mehreren Klassen abgeleitet werden, *subClassOf* erlaubt die Festlegung der hierarchischen Organisation solcher Klassen etc. Mit Hilfe der Konstrukte *range* und *domain* werden Attributwerte und die Anwendbarkeit von Attributen beschränkt. Insgesamt ergibt sich ein hierarchisch organisiertes Klassensystem, wie es in Abbildung 6 skizziert ist.

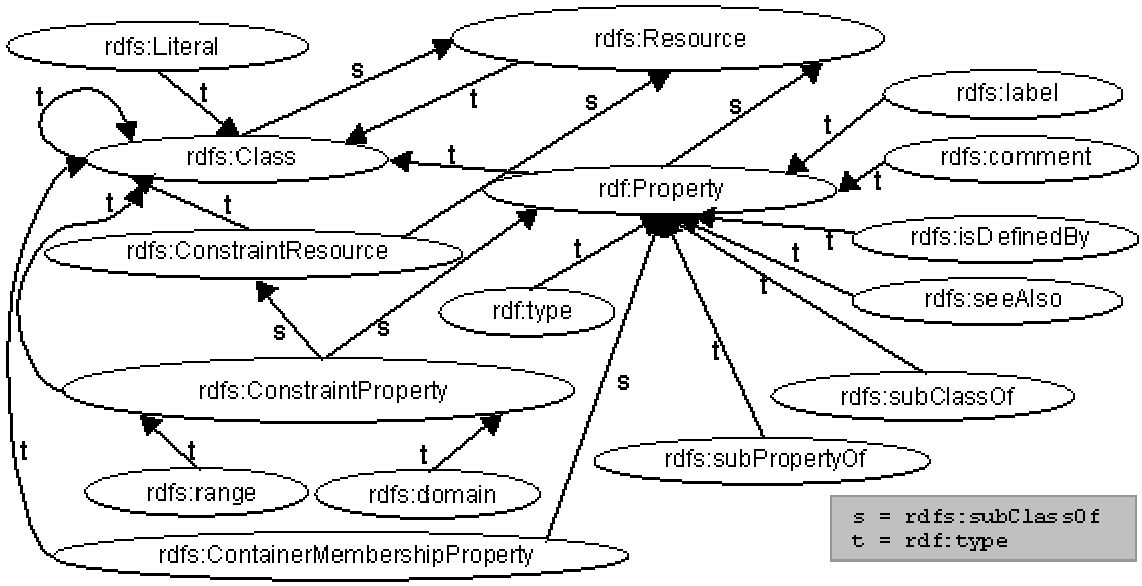


Abbildung 6: Klassenhierarchie von RDFS. Aus: [W3C 2000]

Fachgemeinschaften, die an der Beschreibung von Ressourcen interessiert sind, können nun mit Hilfe von RDF-Schema ein für ihre Domäne geeignetes Vokabular erstellen, indem sie das Klassensystem nutzen und für ihre Zwecke erweitern. Abbildung 7 stellt dazu ein einfaches Beispiel dar.

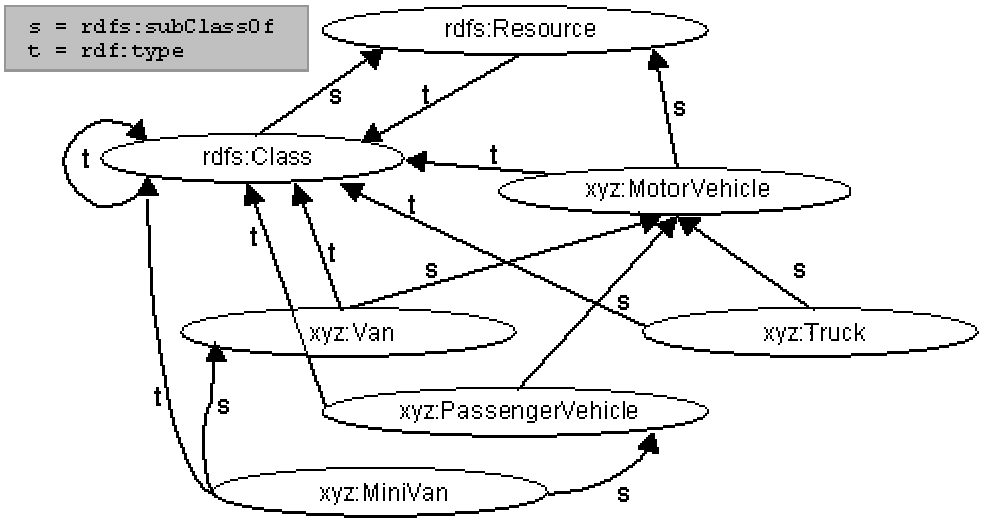


Abbildung 7: Beispiel für eine RDFS-Anwendung. Aus: [W3C 2000]

Derartige Vokabulare, die mit dem von XML bekannten Namensraum-Mechanismus eindeutig identifiziert werden, werden *Schema* genannt. In Abbildung 5 bezeichnet das Präfix „xyz“ ein solches Schema, „rdf“ und „rdfs“ in Abbildung 4 und 5 stehen für die Schemata, die die mit dem RDF-Datenmodell bzw. RDFS eingeführten Begriffe definieren.

Im Unterschied zu XML Schema bzw. DTDs, die die Ordnung und Kombination von Tags in einem XML-Dokument beschreiben, beschränkt RDFS nicht das syntaktische Aussehen einer RDF-Beschreibung. Als deklarative Repräsentationssprache ermöglicht es vielmehr, Informationen zur Interpretation von RDF-Beschreibungen zur Verfügung zu stellen.
[W3C 2000]

Das Datenmodell von RDF sieht keine spezielle Syntax vor. Die Spezifikation enthält mehrere Syntaxalternativen basierend auf XML, die teilweise jedoch noch überarbeitet werden, wie ein Arbeitspapier des W3C zur Syntax vom 18.12.2001 zeigt [W3C 2001]. Ausgehend von der XML-Serialisierungssyntax kann RDF auch als XML-Anwendung angesehen werden. Es sind jedoch andere Serialisierungen des Graphenmodells z. B. mit Hilfe des *Simple Object Access Protocols (SOAP)* denkbar.

Die XML-Serialisierung sieht in ihrer Grundform vor, dass alle Statements zu einer Ressource mit Hilfe des Elementes *description* gruppiert werden, wobei das Attribut *about* dieses Elementes die Ressource identifiziert, auf die es sich bezieht. Der Graph aus Abbildung 1 würde somit folgendes XML-Dokument ergeben:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/">
  <rdf:Description
   about="http://users.bestweb/~sowa/krbook/index.htm">
    <s:Autor>John F. Sowa</s:Autor>
  </rdf:Description>
</rdf:RDF>
```

In diesem Beispiel wurde mit s="http://description.org/schema/" ein fiktives Schema angegeben, in welchem das Attribut *Autor* definiert sein soll.

Abkürzungen der Syntax sind möglich, indem z. B. Eigenschaften, die nicht wiederholt werden und atomare Werte haben, als XML-Attribute des *description*-Elementes angegeben werden.
[W3C 1999]

3.4.4 Mehrwerte und Anwendung

Worin liegt nun der Mehrwert von RDF gegenüber XML, wenn es letztlich in XML serialisiert wird und somit eine XML-Anwendung darstellt? Folgende Charakteristika machen RDF als Standardrepräsentationssprache für Metadaten interessant:

Eindeutige Beschreibung der Semantik einer Anwendungsdomäne

Wie erwähnt ist es nur beschränkt möglich, an eine Datenbasis mit unbekannten XML-Dokumenten eine semantische Anfrage zu stellen, auch wenn alle Objekte der betreffenden Anwendungsdomäne bekannt sind. Die Semantik der Anwendungsdomäne kann in der Regel durch eine große Vielzahl von XML-Bäumen repräsentiert werden, so dass eine Anfrage die alle denkbaren Bäume berücksichtigen soll, rasch zu komplex wird. Darüberhinaus ist es möglich, dass ein Element eines XML-Baumes ein anderes qualifiziert und die Reihenfolge der Elemente bedeutsam ist. RDF hat demgegenüber den Vorteil, dass alle Objekte in der Form von Ressourcen unabhängige Entitäten sind. Mit Hilfe der Objekt-Attribut-Wert-Struktur von RDF können die Objekte und Beziehungen in einer Anwendungsdomäne natürlich und eindeutig modelliert werden. Eine Aussage, die in XML verschiedene Erscheinungsformen annehmen kann, führt bei Anwendung des RDF-Modells zu einem eindeutigen logischen Graphen.
[Miller 1998] drückt es folgendermaßen aus: "RDF is an application of XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics". Dies löst noch nicht das Problem, dass in verschiedenen Domänen unterschiedliche Bezeichnungen für Objekte der Realwelt verwendet werden und daher entsprechende Abbildungen gefunden werden müssen. Mit RDF ist jedoch ein hoher gemeinsamer Grad der Wissensrepräsentation gegeben, der entsprechende semantische Interoperabilität zwischen den Anwendungsdomänen ermöglicht.
[Decker 2000], [Berners-Lee 1998]

Förderung flexibler Kommunikation

Durch die eindeutige, deklarative Semantik von RDF ist es nicht mehr - wie bei reinen XML-Anwendungen - zwingend erforderlich, dass sich zwei Kommunikationspartner im WWW vor Aufnahme einer Kommunikation über die Struktur der auszutauschenden Daten einigen. Einer Anwendung ist es durch Interpretation der RDF-Beschreibung unter Auswertung der referenzierten Schemas grundsätzlich möglich, auch fremde Datenstrukturen zu verstehen und zu nutzen. Dadurch wird flexible Kommunikation mit häufig wechselnden Kommunikationspartnern gefördert.
[Decker 2000]

Möglichkeit der partiellen Auswertung

Die Statements in einem RDF-Graphen können von einander unabhängig betrachtet werden. Folglich ist es möglich, dass eine Anwendung nur jene Tripel einer RDF-Beschreibung nutzt, die sie versteht, und die restlichen außer acht lässt. Ein derartiges partielles Verständnis ist bei Datenstrukturen, die auf einer Ordnung beruhen, wie z. B. XML, nicht möglich.
[Berners-Lee 1998]

Infrastruktur zur Kodierung, Austausch und Wiederverwendung strukturierter Metadaten

Mit RDFS können Fachgemeinschaften das für ihre Anwendungsdomäne notwendige Metadatenvokabular erstellen bzw. vorhandene Vokabulare abbilden. Diese können aus RDF-Beschreibungen zu beliebigen Ressourcen mit Hilfe des Namensraum-Mechanismus referenziert werden, wobei die Mischung von Metadatenelementen aus verschiedenen Vokabularen möglich ist. Strukturierte Metadaten können somit – unabhängig vom Metadatenstandard – allgemein verständlich kodiert werden. Der Austausch und die Wiederverwendung wird gefördert. Dazu trägt auch die Wahl von XML als vorgeschlagener Serialisierungssyntax bei. Ferner spielt es eine Rolle, dass RDFS selbst in RDF formuliert ist: Auch wenn eine Anwendung die Semantik eines Vokabulars nicht versteht, so kann sie es dennoch einlesen und transportieren.
[Miller 1998]

Semantische Modularität

Es wird nicht der – vermutlich unmögliche – Versuch unternommen, alles mit einer gemeinsamen, universellen Wissensrepräsentationssprache zu beschreiben. Mit den verteilten Schemas existieren zahlreiche, unabhängige Metadatenrepräsentationen, die jeweils für eine Domäne adäquat sind und darüber hinaus nach Bedarf abgebildet und kombiniert werden können. Vom Einsatz von RDF wird zwar eine hohe semantische Interoperabilität erwartet, in bezug auf die konkreten Beschreibungsmittel ist der Ansatz aber dennoch nicht normierend und dezentral.
[Decker 2000]

Erweiterbarkeit

Die Eigenschaften, die eine Ressource haben kann, sind nicht beschränkt und werden auch nicht bei Definition der Klasse der Ressource mit Hilfe von RDFS festgelegt. Jeder Anwender kann neue Eigenschaften kreieren, die selbst Ressourcen darstellen und über eine URI eindeutig identifiziert werden. Da diese Eigenschaften ferner unabhängig von der Ressource gespeichert werden können, die sie beschreiben, ist die Erweiterung auch praktisch leicht durchzuführen. Damit ist gewährleistet, dass jeder Anwender zu jeder beliebigen Ressource die für seinen Anwendungskontext erforderlichen Aussagen machen kann, ohne dass die Unabhängigkeit der beschriebenen Ressource beeinträchtigt wird.

Unabhängigkeit von der Dokumentstruktur

Da RDF keine Annahmen über die Dokumentstruktur macht, kann es grundsätzlich zur Beschreibung beliebiger Datenstrukturen und Dokumenttypen dienen. Es sind flexible Verknüpfungen denkbar: So könnte die Beschreibung beispielsweise als Overhead in die Dokumentdatei selbst eingefügt werden, gemeinsam mit dem Dokument im Rahmen des http-Protokolls übertragen werden oder über das Protokoll von einem Drittanbieter, z. B. einem Katalog, bezogen werden.

Einfachheit

Durch das einfache Grunddatenmodell sind RDF-Beschreibungen grundsätzlich gut verständlich und können effizient implementiert und gespeichert werden. In Bezug auf eine große Menge von Aussagen ist eine gute Skalierbarkeit zu erwarten.

Nutzung der Reifikation

Die Möglichkeit der Reifikation in RDF erlaubt Aussagen über Aussagen. So kann z. B. beschrieben werden, von wem und wann eine Aussage gemacht wurde, es können aber auch Zweifel und Zustimmung zu Aussagen ausgedrückt werden. Diese Möglichkeiten könnten etwa für die Bewertung der Glaubwürdigkeit von Informationen und Diensten bedeutsam sein.
[Decker 2000]

RDF hat noch nicht soviel praktische Anwendung gefunden, wie dies bei reinem XML der Fall ist. Vor allem im kommerziellen Bereich ist die Verbreitung bislang eher gering. Bedeutende Anwendungen sind das bereits vorgestellte PICS und die *Platform for Privacy Preferences (P3P)*, ein Standard, der die automatisierte Darstellung und Nutzung der Datenschutzpraxis von Websites ermöglichen soll. Weitere interessante Ansätze stellen das *RDF Site Summary (RSS)*, ein Metadatenbeschreibungs- und Verteilungsformat, dessen Ursprünge in der Verbreitung von Schlagzeilen liegen, und die Angabe von Geräteeigenschaften und Nutzungspräferenzen im Rahmen des *Wireless Application Protocol (WAP)* dar. Diese Anwendungen sehen die Beschreibung von verteilten Ressourcen mit Metadaten nach dem RDF-Datenmodell vor und stellen teilweise Schemas zur Verfügung.

Ein weiterer Bereich, in dem RDF schon praktisch eingesetzt wird, sind Online-Kataloge und Archive. Systeme wie der *Cooperative Online Resource Catalog (CORC)* des Online Computer Library Centers (OCLC) verwenden RDF-konforme Metadaten zur kooperativen Erschließung von Ressourcen und zum Datenaustausch. Ein Überblick über die aktuellen Anwendungen und Projekte findet sich bei [Beckett 2002].

3.5 Ontologiesprachen für das WWW

3.5.1 Motivation und Definition

RDF und RDFS sind im Gegensatz zu XML darauf ausgerichtet, die Semantik von Beschreibungselementen durch Referenzierung von Schemas deklarativ darzustellen und damit den Aufbau eines semantischen Modells einer Domäne zu ermöglichen. Derartige semantische Darstellungen von Daten werden vielfach als wesentliche Voraussetzung für eine effiziente, rechnerbasierte Nutzung des WWW und den Einsatz von automatisierten Webdiensten angesehen (vgl. [Heflin, Hendler 2000], [van Harmelen, Fensel 1999], [Berners-Lee 1998]).

Eine Beschreibung von Ressourcen mit Hilfe von RDFS ist jedoch nur beschränkt rechnerinterpretierbar, da keine formale Semantik für die verwendeten Modellierungsprimitiven zur Verfügung gestellt wird und die Primitiven nicht ausdrucksstark genug sind, um komplexe Beziehungen zwischen Objekten darzustellen. So kann mit Hilfe von RDFS z. B. keine Äquivalenz oder Disjunktheit von Klassen ausgedrückt werden. Eigenschaften können nur unzureichend etwa hinsichtlich ihrer Transitivität beschrieben werden und die Möglichkeiten zur Formulierung von Constraints sind limitiert. Die beschränkte Ausdrucksmächtigkeit führt insbesondere auch dazu, dass keine ausreichenden logischen Schlussfolgerungen über RDFS-Daten möglich sind. Schlussfolgerungsfähigkeit stellt jedoch eine wesentliche Voraussetzung für die inhaltliche Interpretation von Daten durch Rechner, die Ableitung neuen Wissens und Konsistenz- sowie Typprüfungen dar.[Broekstra et al 2000] [W3C 2002b]

Eine Lösung dieses Problems könnten Ontologiesprachen sein, wie sie auf dem Gebiet der Künstlichen Intelligenz (KI) und hierbei insbesondere im Teilbereich Wissensrepräsentation seit Beginn der 90er Jahre entwickelt werden. Auf den Begriff der Ontologie wird im Zusammenhang mit der Darstellung von Terminologien näher eingegangen werden. Eine ausreichende Charakterisierung dürfte zunächst die Definition von [Gruber 1993] sein: „An ontology is a formal, explicit specification of a shared conceptualization“. Unter Spezifikation einer gemeinsamem Konzeptualisierung ist zu verstehen, dass eine Gruppe von Menschen oder Anwendungen einen Weltausschnitt beschreibt, indem sie ein gemeinsames Verständnis der Konzepte definiert, die für diesen Ausschnitt relevant sind. Es handelt sich um eine explizite Spezifikation, da die vorkommenden Objekte, die Beziehungen zwischen den Objekten und die für ihren Gebrauch geltenden Regeln und Beschränkungen eindeutig und intensional definiert werden. Ferner geschieht dies auf formale Weise, d. h. mit Hilfe einer formalen, logikbasierten Sprache, die rechnerinterpretierbar sein sollte. Grundsätzliches Ziel der Entwicklung von Ontologien auf dem Gebiet der KI ist die erleichterte gemeinsame Nutzung und Wiederverwendung von Wissen.
[Fensel 1999]

Von der Verwendung von Ontologiesprachen im WWW erhofft man sich eine genauere semantische Beschreibung von Ressourcen. Dies soll wiederum eine größere Interoperabilität der unabhängig entwickelten Anwendungsdomänen und die Nutzung des WWW als formale Wissensbasis ermöglichen. D. h. es sollen in einem gewissen Ausmaß logische Anfragen beantwortet, Wissen kombiniert und neues Wissen abgeleitet werden können.

Bereits vorhandene Ontologiesprachen wie z. B. CycL, KIF und Ontolingua können aufgrund der besonderen Anforderungen, die das WWW stellt, jedoch nicht ohne weiteres angewandt werden. So sollte

eine Ontologiesprache für das WWW für den menschlichen Benutzer verständlich und relativ einfach handhabbar sein. Ferner muss sie über eine ausreichende Ausdrucksfähigkeit verfügen, um die Vielfalt an Wissen im WWW darzustellen, wobei zugleich die Effizienz der Anfragebearbeitung und die Schlussfolgerungsfähigkeit gewahrt bleiben sollte. Außerdem kann im Gegensatz zur Konzeption vieler KI-Systeme nicht von einer sogenannten „geschlossenen Welt“ ausgegangen werden: Aufgrund der Größe und des dynamischen Wachstums ist es unmöglich, alle Fakten zu kennen und zu modellieren; unzuverlässige Informationen und Inkonsistenzen sind unvermeidbar. Eine weitere Anforderung ist der Aufbau auf bestehenden Metadatenstandards für das WWW, um Interoperabilität und Akzeptanz zu fördern.
[W3C 2002 b], [Heflin, Hendler 2000], [Fensel 1999]

Beispiele für Ontologiesprachen für das WWW sind *Simple HTML Ontology Extensions (SHOE)* und *Ontobroker*, die beide zunächst die Beschreibung von HTML-Dokumenten durch in die Dokumente eingebettete Sprachelemente vorsehen. Zur Definition der verwendeten Beschreibungselemente wird auf externe Ontologien verwiesen, wobei es sich im Fall von SHOE um beliebige mittels SHOE definierte Ontologien handeln kann, die durch eine URI eindeutig identifiziert werden, während Ontobroker auf einer einzigen, zentral definierten Ontologie beruht. Ein weiterer Unterschied ist, dass SHOE auf deskriptiver Logik (auch terminologische Logik genannt) basiert, eine Familie logischer Formalismen zur Wissensrepräsentation, die auf die Bereitstellung wohldefinierter Semantik bei Wahrung effizienter Schlussfolgerungsfähigkeit ausgerichtet ist. Ontobroker baut dagegen auf Frame-Logik auf, eine Sprache, die die Grundlage für objektorientierte Systeme darstellt, indem sie die entsprechenden Modellierungskonstrukte wie Klassen, Attribute, Wertebeschränkungen etc. in eine kohärente logische Architektur integriert. [Heflin, Hendler 2000], [van Harmelen, Fensel 1999] Zu SHOE existiert eine Weiterentwicklung in Form der *Ontology Markup Language (OML)*, die eine Kodierung von SHOE-Ausdrücken in XML ermöglicht. Eine weitere XML-basierte Ontologiesprache ist die ***X****ML-based* ***O****ntology Exchange* ***L****anguage (XOL)*, welche die Definition von Ontologien mit einem in einer DTD definierten Vokabular vorsieht.
[Fensel 1999]

3.5.2 Ontologiesprache DAML+OIL

Den vorstehenden Sprachen ist gemeinsam, dass sie die – teilweise parallel oder später entwickelten – Standards zur Entwicklung und Verbreitung von Metadaten im WWW nicht oder ungenügend berücksichtigen. Dies hat sich bei den späteren Ansätzen *Ontology Inference Layer (OIL)* und *DARPA Agent Markup Language (DAML)* dahingehend geändert, dass beide auf RDFS und die dafür benutzte Serialisierungssyntax XML aufbauen.
OIL wurde ab Frühjahr 2000 im Rahmen der EU-Projekte On-To-Knowledge und IBROW entwickelt und versucht die Stärken von deskriptiver Logik und Frame-Logik in einer Ontologiesprache für das Web zu vereinen. Die Sprache wurde weitestmöglich in RDFS bzw. – falls erforderlich – als Erweiterung von RDFS formuliert. DAML ist dagegen ein seit August 2000 laufendes Projekt von DARPA, der zentralen Forschungs- und Entwicklungsorganisation des US-Verteidigungsministeriums. Es zielt auf die intelligente Ermittlung und Auswertung von Informationsquellen durch Softwareagenten und die semantische Interoperabilität zwischen den Agenten ab. Als Teil des Projektes wurde mit DAML-ONT eine Ontologiesprache entwickelt, die ein Kernstück der künftigen Agentensprache DAML darstellen soll.

Abgesehen von Details in der Klassenhierarchie und einer geringeren Rückwärtskompatibilität zu RDFS weist DAML-ONT starke Ähnlichkeit zu OIL auf.
[Broekstra et al 2000], [Burke 2002]

Inzwischen wurde ein gemeinsames Komitee ins Leben gerufen, welches versucht, die Stärken beider Entwicklungen in Form der gemeinsamen Sprache DAML+OIL zu vereinen. Da dies einer der fortgeschrittensten Ansätze ist und vom W3C als Grundlage für die dortigen Arbeiten zu einer Ontologiesprache angesehen wird, werden die Grundzüge der Sprache in der Version vom März 2001 kurz vorgestellt:

DAML+OIL ermöglicht es, eine Ontologie zur Beschreibung eines Weltausschnitts zu definieren, in der die relevanten Konzepte durch Klassen mit zugehörigen Eigenschaften und durch Instanzen dieser Klassen spezifiziert werden. Im Rahmen der Beschreibung können auch Werte von XML Schema Datentypen (Datentypwerte) benutzt werden. Eine DAML+OIL-Ontologie beginnt üblicherweise mit einem Header, der Angaben zur Version und Angaben darüber enthält, ob Begriffsdefinitionen aus einer anderen DAML+OIL-Ontologie importiert werden. Anschließend folgen optionale Klassen- und Eigenschaftselemente sowie Instanzen.

Klassenelemente

In Klassenelementen wird die Klassendefinition mit dem Klassennamen, einer URI, verbunden. In der Klassendefinition für eine Klasse *C* können folgendende Konstruktoren verwendet werden:

Konstruktor	Argument	Bedeutung
rdfs: subClassOf	Klassenausdruck	Unterklasse, Mehrfachvererbung und zyklische Unterklassenbeziehungen sind möglich.
daml:disjointWith	Klassenausdruck	*C* hat keine gemeinsamen Instanzen mit Klassen des Klassenausdrucks.
daml:disjointUnionOf	Liste von Klassenausdrücken	*C* ist Vereinigung der Klassen in der Liste, welche paarweise disjunkt sind.
daml:sameClassAs	Klassenausdruck	*C* ist äquivalent zum Klassenausdruck.
daml:equivalentTo	Klassenausdruck	Bei Anwendung auf Klasse entspricht Semantik der von sameClassAs.
Boolesche Kombinationen	Klassenausdrücke	*C* muss äquivalent zur Booleschen Kombination sein.
Aufzählung	Aufzählungselemente	*C* umfasst genau die Instanzen, die aufgezählt werden.

Abbildung 8: Klassenkonstruktoren in DAML+OIL

Alle Konstruktoren sind optional und wiederholbar. Klassenausdrücke können entweder Klassennamen, Aufzählungen aller möglichen Instanzen einer Klasse, Eigenschaftsbeschränkungen (property restrictions) oder Boolesche Kombinationen von Klassenausdrücken sein. Für die Verwendung von Präfixen gelten die XML-Namensraum-Konventionen, die durch die Nutzung der RDF-Serialisierungssyntax Anwendung finden.

Mit Hilfe der Konstruktoren rdfs:subClassOf und daml:disjointWith können notwendige, aber nicht hinreichende Bedingungen für eine Klassenmitgliedschaft definiert werden, während sich mit den übrigen

Elementen notwendige und hinreichende Bedingungen formulieren lassen. Dies weist auf ein Konzept von DAML+OIL hin, welches einerseits sogenannte primitive Klassen ermöglicht, die durch die Angabe von notwendigen Eigenschaften der zugehörigen Individuen die Klasse zwar beschreiben, aber nicht ausreichend definieren. So könnte beispielsweise die Klasse „Pflanze“ wie folgt beschrieben werden:

```
<daml:Class rdf:ID="Pflanze">
  <rdfs:subClassOf rdf:resource="#Lebewesen"/>
</daml:Class>
```

Andererseits können durch Angabe von notwendigen und hinreichenden Bedingungen Definitionen von Klassen formuliert werden, wie im nachfolgenden Beispiel die Konstruktion der Klasse „Mann“:

```
<daml:Class rdf:ID="Mann">
    <daml:intersectionOf parseType="daml:collection">
        <daml:Class rdf:resource="#Erwachsener"/>
        <daml:Class rdf:resource="#männlich"/>
    </daml:intersectionOf>
</daml:Class>
```

Die Notation “#” gefolgt von einer Zeichenkette in den obigen Beispielen steht für eine Referenz auf Objekte, die in dem vorliegenden Dokument definiert und mit der Zeichenkette als ID bezeichnet wurden. Falls sich die Definition in einem anderen Dokument befände, müsste zusätzlich die entsprechende URI angegeben werden.

Eigenschaftselemente (property elements)

Eine Eigenschaft ist eine binäre Relation, die ein Objekt zu einem anderen Objekt oder zu einem Datentypwert in Beziehung setzt. Im ersten Fall handelt es sich um eine Instanz der Klasse *ObjectProperty*, im zweiten Fall um eine Instanz der Klasse *DatatypeProperty.*

Ein Eigenschaftselement *P* wird wie eine Klasse über eine URI identifiziert und kann durch die in Abbildung 9 angegebenen Elemente definiert werden:

Element	Argument	Bedeutung
rdfs: subPropertyOf	Eigenschaftsname	Eine Instanz von *P* ist auch eine Instanz der angegebenen Eigenschaft.
rdfs:domain	Klassenausdruck	*P* ist nur auf Instanzen des Klassenausdrucks anwendbar.
rdfs:range	Klassenausdruck	*P* kann nur Werte enthalten, die Instanzen des Klassenausdrucks sind.
daml:samePropertyAs	Eigenschaftsname	*P* ist äquivalent zur angegebenen Eigenschaft.
daml:equivalentTo	Eigenschaftsname	Bei Anwendung auf eine Eigenschaft entspricht Semantik der von samePropertyAs.
daml:inverseOf	Eigenschaftsname	*P* ist die inverse Relation der angegebenen Eigenschaft.

daml:TransitiveProperty		Zeigt an, dass *P* transitiv ist.
daml:UniqueProperty		Für jede Instanz von *P* kann die Eigenschaft nur einen Wert annehmen.
daml:UnambigousProperty		Ein Wert kann nur einer Instanz von *P* zugeordnet werden.

Abbildung 9: Konstruktoren für Eigenschaftselemente in DAML+OIL

Alle Elemente sind optional. Die drei letztgenannten Elemente, die eine Eigenschaft qualifizieren, sind keine Instanzen der Klassen ObjectProperty bzw. DatatypeProperty und können sinnvollerweise nur jeweils einmal verwendet werden. Die übrigen Elemente sind wiederholbar.

Eigenschaftsbeschränkungen

Eigenschaften wurden zuvor als binäre Relation bezeichnet. Eine solche binäre Relation *R* stellt eine Menge von Paaren *(x,y)* von Elementen der Mengen *A* und *B* dar, wobei *x* aus der Menge *A* und *y* aus *B* stammt. Eine Eigenschaftsbeschränkung *C* gibt Bedingungen und Einschränkungen an, die *x* und *y* erfüllen müssen, damit das Paar *(x,y)* Element von *R* mit der Eigenschaftsbeschränkung *C* ist. Eine Eigenschaftsbeschränkung definiert somit implizit eine anonyme Klasse: Die Klasse aller Objekte, die die Beschränkungen erfüllen.

Für jede Eigenschaftsbeschränkung muss mit Hilfe des *daml:onProperty*-Elementes angegeben werden, auf welche Eigenschaft *P* sie sich bezieht. Außerdem kann sie eines oder mehrere der folgenden Elemente enthalten:

Element	Argument	Bedeutung
daml:toClass	Klassenausdruck	Klasse der Objekte, bei denen alle Werte von *P* den Klassenausdruck erfüllen.(Allquantifikator)
daml:hasValue	Objekt oder Datentypwert	Klasse der Objekte bei denen mindestens ein Wert von *P* gleich dem angegebenen Objekt / Datentypwert ist.
daml:hasClass	Klassenausdruck oder Datentypreferenz	Klasse der Objekte, bei denen mindestens ein Wert von *P* den Klassenausdruck / die Typangabe erfüllt. (Existenzquantifikator).
daml:cardinality	nicht-negativer Integer-Wert *N*	Klasse der Objekte, die genau *N* unterschiedliche Ausprägungen für *P* hat.
daml:maxCardinality	nicht-negativer Integer-Wert *N*	Klasse der Objekte, die höchstens *N* unterschiedliche Ausprägungen für *P* hat.
daml:minCardinality	nicht-negativer Integer-Wert *N*	Klasse der Objekte, die mindestens *N* unterschiedliche Ausprägungen für *P* hat.

Abbildung 10: Konstruktoren für Eigenschaftsbeschränkungen in DAML+OIL

Zu den Kardinalitätsbeschränkungen existiert auch jeweils eine qualifizierte Variante, die neben der Kardinalitätsangabe ein *daml:hasClassQ*-Element enthält, welches zusätzlich den Klassenausdruck oder Datentyp angibt, dem die Ausprägungen entsprechen müssen.

So könnte beispielsweise spezifiziert werden, dass eine Person höchstens eine Arbeitsstelle haben kann, die eine Vollzeitbeschäftigung ist:

```
<daml:Class rdf:about="#Person">
  <rdfs:subClassOf>
    <daml:Restriction daml:maxCardinalityQ="1">
      <daml:onProperty rdf:resource="#hatArbeitsstelle"/>
      <daml:hasClassQ rdf:resource="#Vollzeitbeschäftigung"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>
```

Bei Angabe von mehreren Beschränkungselementen in einer Eigenschaftsbeschränkung, muss *P* alle Beschränkungen erfüllen.

Für die Ableitung von Instanzen von Klassen oder Eigenschaften wird RDF und RDFS- Syntax benutzt. Die Instanz „Asien“ könnte also beispielsweise wie folgt von der Klasse Kontinent abgeleitet werden:

```
<rdf:Description rdf:ID="Asien">
  <rdf:type>
   <rdfs:Class rdf:about="#Kontinent"/>
  </rdf:type>
</rdf:Description>
```

Zusätzlich stehen die Konstrukte *daml:sameIndividualAs* und *daml:differentIndividualFrom* zur Verfügung, um Aussagen über die Gleichheit bzw. Verschiedenheit von Objekten treffen zu können.

[van Harmelen et al 2001]

3.5.3 Mehrwerte und Anwendung

Durch folgende Charakteristika von DAML+OIL ist ein Mehrwert gegenüber den bisher vorgestellten Technologien gegeben:

Wohldefinierte Semantik

Die Bedeutung der vorgestellten Modellierungsprimitiven von DAML+OIL ist formal definiert. Es werden sowohl modelltheoretische als auch axiomatische Spezifikationen in Prädikatenlogik erster Ordnung zur Verfügung gestellt. Mit der Bereitstellung wohldefinierter Semantik ist eine wesentliche Voraussetzung für die maschinelle Interpretierbarkeit und automatische Schlussfolgerungen gegeben.
[van Harmelen et al 2001b], [Fikes, McGuinnes 2001]

Erweiterte Ausdrucksmächtigkeit

Gegenüber RDF und RDFS wurde die Ausdrucksmächtigkeit insbesondere durch folgende Spracheigenschaften erhöht:

- Möglichkeit notwendige und ausreichende Bedingungen für eine Klassenzugehörigkeit zu formulieren.
- Boolesche Kombinationen von Klassen und Eigenschaften sind z. B. bei der Konstruktion von Klassen und bei Eigenschaftsbeschränkungen möglich.
- Definition von Klassen durch Aufzählung aller Instanzen.
- Durch die Möglichkeit, die Disjunktheit von Klassen zu modellieren, Eigenschaften als transitiv, invers oder eineindeutig zu kennzeichnen, stehen zusätzliche Informationen für Inferenzmechanismen zur Verfügung. Das Äquivalenzkonstrukt erleichtert die Kombination von Informationsquellen.
- Die vorgestellten Eigenschaftsbeschränkungen stellen ein mächtiges Mittel dar, um Klassen mit Hilfe einer Menge von Ressourcen zu definieren, die durch das Vorhandensein bestimmter Eigenschaften, die Anzahl oder den Wert dieser Eigenschaften bestimmt sind. Dabei ist u. a. Existenz- und Allquantifikation möglich.
- Mit den Eigenschaftselementen sind reichhaltigere Möglichkeiten zur Definition und Beschränkung von Eigenschaften gegeben.

Aufbau auf RDF und RDFS

Eine in DAML+OIL formulierte Ontologie ist eine Menge von RDF-Statements. Dadurch kann die RDF-Architektur zur Interoperabilität von Metadaten genutzt werden. Ferner ist gewährleistet, dass Ressourcenbeschreibungen in DAML+OIL auch mit einem gewöhnlichen RDF-Parser lesbar sind. Die zusätzliche Semantik einiger Aussagen kann hierbei zwar nicht genutzt und die mit DAML eingeführten Modellierungsprimitiven können nicht inhaltlich ausgewertet werden. Die Nutzung als RDF-Datenbasis ist dadurch jedoch nicht eingeschränkt. Insbesondere können innerhalb von DAML+OIL auch beliebige sonstige Aussagen in RDF gemacht werden.

Praktische Anwendbarkeit

Die Konzeption von DAML+OIL berücksichtigt die Anwendungsbedingungen im WWW dahingehend, dass die Sprache zwar über hohe Ausdrucksmächtigkeit verfügt, aber im Gegensatz zu herkömmlichen Ontologiesprachen wie etwa Ontolingua immer noch entscheidbar ist und effiziente Schlussfolgerungsmechanismen erlaubt.

Unterstützung von verteilten Ontologien

Über den Import-Mechanismus können Ontologien mittels URI identifiziert und darin enthaltenen Spezifikationen importiert werden. Es können auch einzelne Objekte einer anderen Ontologie referenziert und Aussagen zu ihnen gemacht werden. Dies sind grundlegende Voraussetzungen, um ein Web mit einer Vielzahl von domänenspezifischen, interoperierenden Ontologien zu unterstützen.

[DAML 2002], [van Harmelen et al 2001]

Schwächen von DAML+OIL sind, dass derzeit keine beliebigen Regeln und Axiome beschrieben werden können und die Reifikation im Gegensatz zu RDF nicht unterstützt wird. Diese Möglichkeiten sollen gegebenenfalls in späteren Versionen oder in weiteren Sprachen, der DAML-Sprachfamilie, die auf DAML+OIL aufbauen, geschaffen werden. So wird momentan in Form von *DAML-Logic (DAML-L)* eine Erweiterung entworfen, die es ermöglichen soll, kompakte Repräsentationen von Regeln in Form von Horn Klauseln auszudrücken.
[DAML 2002], [McIlraith et al 2001]

Obwohl die Bedeutung von Ontologien – insbesondere für das Wissensmanagement und E-Commerce-Anwendungen - nach [Mädche et al 2001], [Das et al 2001] erkannt wurde, ist die Zahl der praktischen Anwendungen von Ontologiesprachen für das WWW noch beschränkt:

In einer Reihe von Projekten und Fallstudien wurde OIL zur Integration von Datenquellen und Unterhaltung von Wissensportalen genutzt, indem Ontologien in OIL formuliert und Ressourcen entsprechend ausgezeichnet wurden. Die ontologische Beschreibung bildet dabei jeweils die Grundlage für die intelligente, automatische Bereitstellung und Strukturierung des Domänenenwissens unabhängig von der Struktur der zugrundeliegenden Datenquellen. [Fensel et al 2001], [Decker et al 2002]
Eine vergleichbare Anwendung mit DAML-Grundlage existiert in Form des ITTALKS-Portals. Außerdem ermöglicht *WebScripter*, ein von der University of Southern California entwickeltes Tool, die Integration von Webressourcen, die mit DAML-Beschreibungen ausgezeichnet sind. Diese und weitere – eher prototypische Anwendungen – werden unter [DAML 2002a] aufgelistet.

Große Bedeutung wird Ontologiesprachen bei der Beschreibung von Webdiensten zugemessen: Die ontologische Beschreibung soll es Softwareagenten ermöglichen, Dienste zu ermitteln, zu nutzen und nach inhaltlichen Kriterien zu kombinieren, ohne dass ein strikter Standard auf Syntax- und Protokollebene durchgesetzt werden muss. Der kommerzielle Anbieter VerticalNet Solutions scheint zwar stark auf den Einsatz von Ontologien in diesem Bereich zu setzen, jedoch v. a. proprietäre Beschreibungsmöglichkeiten anzuwenden. Die Kompatibilität zu offenen Wissensrepräsentationsstandards wie DAML+OIL ist nur eines von vielen Kriterien [Das et al 2001], [McIlraith et al 2001]. Im übrigen entwickelt die DAML-Initiative mit *DAML-Services (DAML-S)* eine weitere Ontologiesprache, die auf DAML+OIL aufbaut und speziell auf Webdienste ausgerichtet ist [Ankolekar et al 2001]. Daneben sind Ontologiesprachen ein wesentlicher Baustein für viele sonstige Anwendungsszenarien eines zukünftigen „Semantic Web“. Diese sollen im Anschluss an eine mögliche Architektur des Semantic Web kurz vorgestellt werden.

3.6 Architektur eines Semantic Web

3.6.1 Schichtenmodell

Der Begriff „Semantic Web“ wurde und wird vor allem von Tim Berners-Lee, dem Vorsitzenden des World Wide Web Consortiums (W3C) geprägt, der aufgrund seiner Pionierarbeiten zum WWW beim European Laboratory for Particle Physics (CERN) häufig auch als dessen „Erfinder“ bezeichnet wird. Nach seinen Angaben handelt es sich bei dem Semantic Web um eine Vision, wonach sich das bisherige WWW, welches die Zusammenarbeit von Menschen erleichterte, auf die Zusammenarbeit von Computern erweiteren soll. „Computer werden in der Lage sein, alle Daten im Web zu analysieren – Inhalte, Links und Transaktionen zwischen Menschen und Computern.“ [Berners-Lee, Fischetti 1999] S. 229

Konkreter wird diese Vision, wenn man das Modell in Abbildung 6 betrachtet, welches Berners-Lee dem Semantic Web zugrundelegt. Die in den vorangegangenen Teilkapiteln vorgestellten Technologien werden gemeinsam mit weiteren Konzepten und Sprachen in einer Schichtenarchitektur vereinigt, welche insgesamt ein hohes Maß an automatisierter, rechnerbasierter Nutzung der Daten im WWW ermöglichen soll. Inzwischen scheint dieses Modell nicht mehr nur von Berners-Lee und dem W3C propagiert zu werden, sondern als Diskussionsgrundlage in den an der Entwicklung des WWW interessierten Communities wie z. B. Wissensrepräsentation (vgl. [Maedche, Staab 2001]) und Agentensysteme (vgl. [Hendler 2001]) zu gelten.

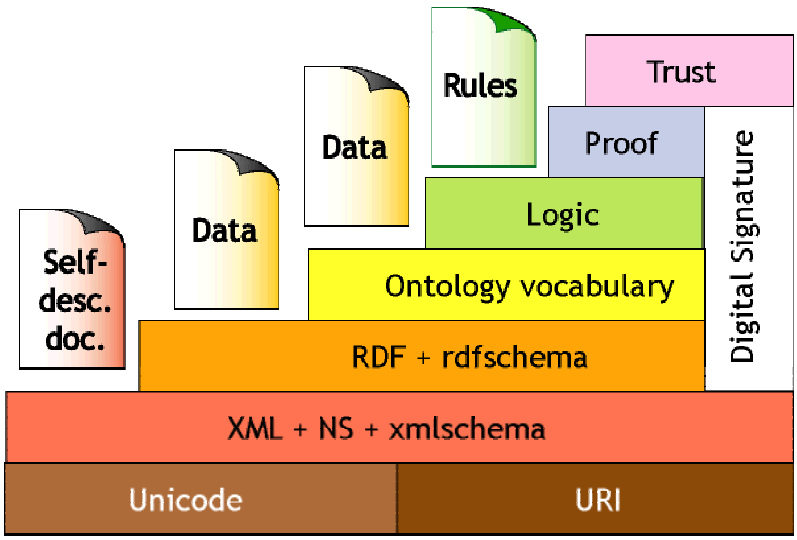


Abbildung 11: Das Semantic Web Schichtenmodell. Aus: [Berners-Lee 2000]

Den einzelnen Schichten kommt im Modell folgende Bedeutung zu:

URI und Unicode
Die Identifikation von Ressourcen erfolgt für das ganze WWW eindeutig über Uniform Resource Identifiers (URI), wobei - dem bei RDF erwähnten Konzept folgend - alles eine Ressource sein kann. Die Verwendung von URIs hat den Vorteil, dass jedermann leicht allgemeingültige Bezeichnungen schaffen

und auf vorhandene Bezeichnungen Bezug nehmen kann, wodurch komplexe Systeme aufgebaut und genutzt werden können.
Die Verwendung von Unicode auf Zeichenebene ergibt sich aus dem Einsatz von XML als Serialisierungssyntax für das gesamte Modell. Durch die eindeutige Nummerierung aller bekannten Zeichen in Unicode können diese unabhängig von Systemplattform, Anwendungsprogramm oder Sprache dargestellt werden.

Syntaxschicht XML
Durch Verwendung von XML wird dessen hohe syntaktische Interoperabilität genutzt und es kann auf eine Vielzahl bereits vorhandener Anwendungen und Tools zurückgegriffen werden. Der Namensraum-Mechanismus stellt außerdem auch für höhere Schichten eine wesentliche Komponente zur Verfügung, um Bezeichner mit ihren semantischen Definitionen zu verbinden.

RDF und RDFS
Nach [Berners-Lee 1998a] erfordert ein universales Web von semantischen Aussagen ein gemeinsames Modell von großer Allgemeinheit, um die Vielfalt der beteiligten Informationssysteme mit unterschiedlichsten internen Modellen darauf abbilden zu können. Diese Funktion wird von der RDF/RDFS-Schicht wahrgenommen.

Ontologieschicht
Sie besteht aus Ontologiesprachen und ermöglicht v. a. eine genauere semantische Beschreibung der Domänen, einschließlich verbesserter Abbildungs- und Schlussfolgerungsmöglichkeiten.

Logische Schicht
Die logische Schicht soll eine Turing-vollständige Sprache enthalten, in der man im Gegensatz zu Ontologiesprachen auch Funktionen und Regeln definieren kann, die weitere Schlüsse, etwa zur Wahl von Handlungsabläufen oder Kombination von Wissen ermöglichen. Ein allgemeiner Vorschlag für eine solche Websprache liegt noch nicht vor. Die DAML-Erweiterungen DAML-L und DAML-S weisen jedoch in diese Richtung.

Beweis- und Vertrauensschicht
Eine Beweisrepräsentationssprache soll es Agenten ermöglichen, gleichzeitig mit einer Aussage den Schlussfolgerungsweg, der zu der Aussage geführt hat, darzulegen. Dadurch ist es möglich, Aussagen zu verifizieren bzw. im Fall bewiesener Aussagen auf eine Wiederholung der Schlussfolgerungen zu verzichten. Außerdem dürfte versucht werden, die Vertrauenswürdigkeit von Informationen in Abhängigkeit von ihrem Kontext zu modellieren. D. h. eine Anwendung macht ihr Vertrauen auf eine Aussage davon abhängig, aus welcher Quelle diese stammt und wie diese Quelle zu bewerten ist. Auch die Beweis- und Vertrauensmechanismen sind noch nicht entwickelt.

Digitale Signaturen
Digitale Signaturen spielen über die ganze Bandbreite der semantischen Schichten eine große Rolle, da eine wesentliche Voraussetzung von Vertrauen in die verschiedenen Arten von Daten deren gesicherte Herkunft ist. Schlüssel, mit denen Aussagen unterschrieben werden, werden daher ebenfalls Ressourcen in einem Semantic Web darstellen. Parallel zur Verifizierung von Aussagen muss die Prüfung der zugehörigen Signaturen erfolgen.
[Berners-Lee 1998 a], [Mädche, Staab 2001], [Berners-Lee 1999]

3.6.2 Charakteristika

In Ergänzung zum vorgestellten Modell werden im wesentlichen folgende Angaben zur Konzeption und Funktionsweise des Semantic Web gemacht:

Abgrenzung zur Künstlichen Intelligenz

Der von der Semantic Web Community verfolgte Ansatz besteht nicht darin, Rechner mit Methoden der Künstlichen Intelligenz in die Lage zu versetzen, menschliche Handlungen wie das Lesen und die Analyse eines Textes durchzuführen. Statt dessen werden Ressourcen im Web, die derzeit auf die Auswertung durch den menschlichen Nutzer ausgerichtet sind, zusätzlich mit wohldefinierten, semantischen Metadaten ausgezeichnet, so dass die enthaltenen Informationen auch durch Rechner verarbeitbar sind. Methoden zur direkten Abbildung menschlicher Intelligenz, wie z. B. Natural Language Parsing, können beispielsweise bei der Schaffung von Tools eine Rolle spielen, mit deren Hilfe Metadaten möglichst automatisch aus bestehenden Ressourcen generiert werden. Sie sind jedoch kein zentrales Konzept der Semantic Web Aktivitäten.
[Berners-Lee 1998b]

Abgrenzung zu wissensbasierten Systemen

Im Gegensatz zu herkömmlichen wissensbasierten Systemen kann im Rahmen des Semantic Web nicht davon ausgegangen werden, dass alle Informationen verfügbar und konsistent sind. Es kann zu Widersprüchen, unbeantwortbaren Anfragen und nicht beweisbaren Aussagen kommen.

Andererseits kann jedoch auf die zentralen Annahmen von Wissensrepräsentationssystemen verzichtet werden, die typischerweise globale Constraints verwenden und verlangen, dass jedes Konzept vor der Verwendung allgemeingültig definiert wird. Dadurch stellt sich auch nicht das bislang ungelöste Problem, wonach in einem umfassenden System zunächst alle Aspekte der Realwelt zu modellieren sind. Es wird dagegen darauf gebaut, dass anfänglich domänenweit Modelle aufgebaut werden, die innerhalb der Domäne mit Schemas konsistent gehalten werden. Durch Verknüpfung entsteht in der Folge ein immer größer werdender Informationsraum. Dabei wird das Wachstum durch die flexiblen Linkstrukturen, die einfache Erweiterbarkeit und die Möglichkeiten von semantischen Mappings, die die verwendeten Technologien bieten, begünstigt.

[Berners-Lee 1998b] geht davon aus, dass in diesem Informationsraum eine große Anzahl praktisch relevanter Anfragen beantwortbar ist. Er vergleicht die Konzeption mit der Entwicklung des WWW, dessen Erfolg u. a. darauf zurückzuführen sein, dass man die zentralistischen Annahmen herkömmlicher Hypertextsysteme aufgegeben habe, indem jedem erlaubt werde, Dokumente und Verknüpfungen zu kreieren. Der Vorteil der erleichterten Erweiterbarkeit des Systems und der hohen Kombinierbarkeit von Dokumenten habe den Nachteil von Inkonsistenzen in Form von toten Links überwogen.
[Berners-Lee 1998b], [Berners-Lee et al 1999]

Komplexität von Anwendungen

Nicht jede Anwendung für das Semantic Web muss mit Logik höherer Ordnung umgehen können. Vielmehr sollen Anwendungen unterschiedlicher Komplexität jeweils die Schicht nutzen, die für ihre Zwecke ausreichend ist. Ferner muss nicht jede Anwendung in der Lage sein, beliebige Beweise zu erstellen, sondern für die meisten soll es ausreichen, einen Beweis von allgemeiner Form nachzuvollziehen. Viele Anwendungen werden auf die Lösung weniger Probleme mit überschaubarem Abdeckungsgrad beschränkt sein. Ein Großteil des Potentials soll sich aus der Kombinierbarkeit der modularen Anwendungen und der Wiederverwendbarkeit semantisch beschriebener Daten ergeben.
[Palmer 2001], [Berners-Lee 1998b]

Suchmaschinen

Auch für Anfragesprachen soll gelten, dass parallel mehrere Sprachen mit unterschiedlichen Möglichkeiten eingesetzt werden. Je nach verwendeter Sprache werden sich Suchmaschinen nicht nur im Abdeckungsgrad, sondern auch in ihren Schlussfolgerungsfähigkeiten und Antwortzeiten unterscheiden. Die Indexe von Suchmaschinen sollen zunehmend eindeutig beschriebene RDF-Objekte anstelle von Wortlisten enthalten, wodurch die Effektivität der Suche verbessert wird.

Ferner erlaubt das Semantic Web nach [Berners-Lee 1998a] die Vorteile von Logiksystemen und herkömmlichen Suchmaschinen zu kombinieren: So sollen nachweisbar korrekte Ergebnisse geliefert werden, wobei gleichzeitig mit großen Datenmengen umgegangen werden kann. Dennoch wird parallel zu logikbasierten Anfragen wohl weiterhin auch Schlüsselwortsuche stattfinden, da die Details vieler Seiten zu wenig strukturiert und zu komplex sind, um sie mit Metadaten zu beschreiben. Im Rahmen von Anfragen dürften sogenannte *third party indexes* eine Rolle spielen, in denen Zwischenergebnisse von Schlussfolgerungen und Begriffsabbildungen zwischen Domänen abgespeichert werden, um eine Verkürzung der Bearbeitungsprozesse zu ermöglichen.
[Berners-Lee et al 1999]

Evolution

Durch deklarative semantische Modellierung und Schlussfolgerungsfähigkeit sollen Applikationen im Semantic Web in der Lage sein, Daten zu „verstehen“, ohne dass sich Produzent und Konsument zuvor über die Struktur der Informationseinheiten geeinigt haben. Dies umfasst auch, dass eine Version *n* Implementation einer Anwendung genügend Schema- bzw. Ontologiedaten einlesen kann, um Daten vom Typ der Version *n+1* verarbeiten zu können. Somit soll grundsätzlich eine Fortentwicklung parallel zur Erweiterung der verwendeten Modelle gegeben sein.
[Berners-Lee et al 1999], [Berners-Lee 1998a]

Realisierung

Das Konzept des Semantic Web beruht auf der Auszeichnung mit zusätzlichen Metadaten und stellt damit einen Zusatzaufwand für Informationsproduzenten im WWW dar. Dieser Aufwand wird erst kompensiert, wenn interessante Dienste zur Verfügung stehen, die auf diesen Metadaten aufbauen. Andererseits lohnt

sich die Entwicklung und Bereitstellung von Semantic-Web-Diensten nur, wenn entsprechende Ressourcen zur Einbeziehung vorhanden sind. Die Semantic-Web-Community hofft diesem Anlaufproblem durch die Bereitstellung von Tools zu begegnen, die die weitgehend automatisierte Erstellung der Metadaten ermöglichen. Außerdem wird erwartet, dass in Teilbereichen, in denen sich beispielsweise Möglichkeiten für E-Commerce eröffnen, eine starke kommerzielle Entwicklungsbeteiligung stattfindet.
[Hendler 2001], [Berners-Lee 1998a]

3.6.3 Anwendungsszenarien

Es wird angenommen, dass das Semantic Web v. a. in den nachfolgend genannten Bereichen eine große Rolle spielen wird. Im Falle einer Realisierung darf allerdings analog zur dynamischen Entwicklung des WWW angenommen werden, dass auch weitere - bisher nicht in Betracht gezogene Anwendungsmöglichkeiten – auftreten werden.

Allgemeine Suche im WWW

Das Semantic Web erlaubt eine semantische bzw. konzeptuelle Suche in entsprechend ausgezeichneten Ressourcen im Gegensatz zur bisher praktizierten Schlüsselwortsuche. Es können beispielsweise Anfragen in der folgenden Art gestellt werden: „Gib die Adressen aller US-Wissenschaftler aus, die sich mit dem Thema Semantic Web beschäftigen“. Die Qualität der Antworten steigt aufgrund der Berücksichtigung von Kontext, der Integration von Quellen auf konzeptueller Basis und der Überprüfbarkeit der Resultate.
[Berners-Lee 1998a], [Maedche, Staab 2001]

Verknüpfung von Wissensgebieten

Auf der XML- bzw. RDF-Ebene werden derzeit bereits verschiedene Wissensgebiete verknüpft, indem Datenquellen mit Hilfe von einheitlichen Metadatenmodellen integriert werden. So ermöglicht die Open Archives Initiative (OAI) den integrierten Zugriff auf eine Vielzahl von wissenschaftlichen Dokumentenservern; das DFG-Projekt „Gemeinsames Portal für Bibliotheken, Archive und Museen (BAM) – Ein Online-Informationssystem“ des Bibliotheksservice-Zentrums Baden-Württemberg, der Landesarchivdirektion Stuttgart und das Landesmuseums für Technik und Arbeit in Mannheim realisiert eine übergreifende Recherche und thematische Präsentation der Objektbestände der beteiligten Institutionen. Nach den Vorstellungen von [Berners-Lee, Hendler 2001] soll das vollständig realisierte Semantic Web über diese Ansätze bewusster Integration von Datenquellen hinausgehen, indem es eine zunehmende, automatisierte Verknüpfung von Wissensgebieten auf Konzeptebene erlaubt. Dadurch soll die Verbreitung von Wissen über wissenschaftliche Disziplinen hinweg schneller und einfacher erfolgen und ein „universales Web des Wissens“ entstehen.

Dienste und Transaktionen

Das Semantic Web soll nicht nur einen verbesserten Zugriff auf Informationen, sondern auch verbesserte und neuartige Dienste im WWW ermöglichen. Mit Hilfe der bereitgestellten Sprachen können die Fähigkeiten von Diensten und ihre Nutzungslogik semantisch beschrieben werden. Programme, die mit einem gewissen Grad an künstlicher Intelligenz ausgestattet sind, sogenannte intelligente Agenten, könnten aufgrund dieser Beschreibungen in der Lage sein, geeignete Dienste auszuwählen, diese zu nutzen und zu kombinieren. Durch die Beweis-, Vertrauens- und Signaturkomponenten wird außerdem der erforderliche Sicherheitsrahmen für solche Agentensysteme geschaffen.

Diese erweiterten Dienste in einem Semantic Web könnten insbesondere auch für die Durchführung von Transaktionen im Rahmen des E-Commerce interessant sein. Ein Beispiel wäre etwa mit einem Agenten gegeben, der im Auftrag seines Nutzers eine Reise plant und organisiert, indem er die entsprechenden Online-Dienste von Hotels, Fluglinien, Autovermietungen etc. in abgestimmter Weise auswählt und einschließlich Abwicklung der Bezahlung in Anspruch nimmt. Weitere Möglichkeiten ergeben sich daraus, dass mit Semantic-Web-Sprachen auch die Fähigkeiten und Eigenschaften von elektronischen Geräten mit Schnittstelle zum WWW beschrieben und entsprechend genutzt werden können. Es wird betont, dass unabhängige semantische Beschreibungen neue Kombinations- und Nutzungsmöglichkeiten eröffnen, die Standardisierungsbemühungen auf Protokollebene nicht umfassen können, da sie gerade nicht vorhersehbar waren.[Hendler 2001], [McIlraith et al 2001]

Personalisierung und Customizing

Semantische Metadaten auf dem Niveau eines Semantic Web eignen sich auch für flexible Aussagen dazu, wie der Inhalt einer Webseite zu bewerten ist und welche Präferenzen und Möglichkeiten eine Clientanwendung hat. Von Vorteil sind dabei insbesondere die Möglichkeit, Aussagen zu jeder beliebigen Ressourcen zu treffen und die Mechanismen zur Verknüpfung mit Schemas. Daher können Bewertungs- und Spezifikationsschemen benutzergruppenspezifisch entworfen und flexibel zugeordnet werden. So eröffnen sich beispielsweise für Webbrowser neue Möglichkeiten, gezielt bestimmte Inhalte in Abhängigkeit von der Kategorisierung der Webseite anzuzeigen. Webserver können in Abhängigkeit von den Möglichkeiten und Vorzügen der Clientanwendungen selektiv Daten übertragen.

Verteilte Software

Anstelle von Informationen oder Diensten kann es sich auch um Programmteile handeln, die mit entsprechenden Metadaten ausgezeichnet werden. Durch die Unabhängigkeit von syntaktischen Standards, die die semantische Auszeichnung bietet, dürften sich neue Möglichkeiten im Hinblick auf automatisierte Ermittlung, Nutzung und Komposition von Softwarekomponenten ergeben.

Wissensportale und Kataloge

Ontologiesprachen spielen in Zukunft voraussichtliche eine große Rolle beim (halb-) automatischen Aufbau und der Unterhaltung von Wissensportalen und Katalogen. Diese domänenspezifischen Organisationsformen des WWW dürften auch von den weiteren Möglichkeiten, die ein Semantic Web zur Integration, Verknüpfung und Suche in Informationen bietet, geprägt werden.[Mädche, Staab 2001]

4. Kontrollierte Vokabulare

4.1 Bedeutung

Im Zusammenhang mit den vorgestellten Ansätzen zur semantischen Nutzung des WWW wird deutlich, dass die Festlegung eines Vokabulars für die Beschreibung von Ressourcen eine große Rolle spielt. So wird innerhalb von Dublin Core die Verwendung von Vokabularen zur erleichterten Interpretation von Elementwerten in Form von Encoding Schemes empfohlen; XML sieht mit Hilfe des Namensraum-Mechanismus vor, Elementtypen und Attribute durch Verweis auf entsprechende Sammlungen von eindeutigen Namen zu qualifizieren. RDF und Ontologiesprachen übernehmen nicht nur diese Funktion von XML, sondern stellen mit RDFS bzw. ontologischen Modellierungskonstrukten eigene Mechanismen zur Definition von semantischen Vokabularen zur Verfügung.

Mit dem Versuch, die Semantik der Daten stärker zu modellieren und zu nutzen, scheint also eine verstärkte Kontrolle des verwendbaren Vokabulars einherzugehen. Dies geht offenbar auf das Problem der Uneindeutigkeit der natürlichen Sprache zurück, welches aus den Bereichen bibliothekarische Inhaltserschließung und Information Retrieval bekannt ist. Die Zuordnungen von Begriffen und ihren sprachlichen Repräsentanten, den Bezeichnungen, sind nicht eindeutig. So können beispielsweise mehrere Bezeichnungen einen Begriff (Synonymie) oder eine Bezeichnung mehrere Begriffe repräsentieren (Polysemie / Homonymie). Es können Begriffe unterschiedlicher Spezifität verwendet werden, um eine Ressource zu beschreiben und die in der Beschreibung verwendeten Wortformen können variieren. Dieses Problem wird dadurch verstärkt, dass es sich bei den Autoren, den Nutzern und den für die Beschreibung der Ressourcen zuständigen Indexierern i. d. R. um unterschiedliche Personen handelt. Bei sehr großen Sammlungen wie dem WWW kommt hinzu, dass diese Gruppen in sich heterogen sind und folglich eine unterschiedliche Sprache verwenden. Diese Nachteile führen häufig dazu, dass relevante Ressourcen allein deswegen nicht gefunden werden können, weil die in den Metadatenbeschreibungen und der Anfrage verwendeten Bezeichnungen nicht übereinstimmen.
[Wersig 1985] S. 14f.

Um diesem Problem zu begegnen, wird häufig das Vokabular, welches bei der Auszeichnung verwendet werden soll, beschränkt und strukturiert. Damit soll einerseits die Menge der möglichen Bezeichnungen für einen Sachverhalt limitiert und vereinheitlicht werden. Andererseits steht das Vokabular idealerweise auch dem Suchenden zur Verfügung, der sich somit gezielt geeignete Bezeichner auswählen kann. Derartige Sprachsysteme werden als *kontrollierte Vokabulare* oder *Dokumentationssprachen* bezeichnet.
[Lancaster 1986] S. 3, [Wersig 1985] S. 16

4.2 Strukturierungsformen

[Lancaster 1986] S. 3 definiert *kontrollierte Vokabulare* wie folgt: „(...) a 'controlled vocabulary', that is, a limited set of terms that must be used to represent the subject matter of documents. Such a vocabulary might be a list of subject headings, a classification scheme, a thesaurus, or simply a list of 'approved' keywords or phrases."

Kontrollierte Vokabulare, die eine stärkere Struktur, wie z. B. Querverweise zwischen den Begriffen, aufweisen, bieten gegenüber schwach strukturierten eine verbesserte sprachliche Kontrolle. Der Benutzer wird mit Hilfe der Struktur zu dem Bezeichner geführt, der den auszudrückenden Begriff am besten repräsentiert. Die Bedeutung der Struktur nimmt beim Einsatz kontrollierter Vokabulare in rechnerbasierten Informationssystemen insofern zu, als sie nach Abbildung auf geeignete Daten- und Sprachkonstrukte genutzt werden kann, um den Benutzer bei der Navigation im Vokabular zu unterstützen oder gar eine semantische Interpretation durch den Rechner zu ermöglichen.

Nachstehend werden daher wesentliche Strukturierungsebenen von kontrollierten Vokabularen vorgestellt, die zugleich jeweils für einen Typ von Vokabular stehen. Vokabulare mit sehr wenig Struktur wie z. B. einfache Schlagwortlisten oder natürlichsprachige Glossare werden nicht weiter behandelt, da sie für eine Nutzung im Zusammenhang mit Informationssystemen - und somit auch dem WWW - wenig Vorteile bieten. Zwischen den vorgestellten Typen existieren zahlreiche Mischformen und ihre Bezeichnungen werden teilweise uneinheitlich und sich überschneidend verwendet. Dennoch lassen sich die folgenden Hauptformen unterscheiden:

4.2.1 Taxonomie / hierarchische Klassifikation

Eine grundlegende Strukturierungsform ist die einer Taxonomie. Die verwendbaren Bezeichnungen werden dabei in einer baumartigen Hierarchie angeordnet. Es besteht üblicherweise nur die sich aus der Hierarchie ergebende Ober- und Unterbegriffsbeziehung. Gegebenenfalls kann daraus, dass sich Objekte in einer Hierarchieebene befinden, noch eine Gleichordnung hergeleitet werden. Es wird zwischen monohierarchischen Systemen, in denen ein Begriff nur einen Oberbegriff haben kann, und polyhierarchischen, die mehrere Oberbegriffe zulassen, unterschieden. Diese Strukturierungsform findet im bibliothekarischen Bereich v. a. in Form von hierarchischen Klassifikationen Anwendung, wobei die Bezeichnungen in der Regel durch eine Notation, die Auskunft über die klassifikatorische Struktur gibt, repräsentiert werden.
[Lancaster 1986] S. 13f, [Wersig 1985] S. 32

4.2.2 Thesaurus

Der Thesaurus als kontrolliertes Vokabular wird in DIN 1463 wie folgt definiert:

„Ein Thesaurus (...) ist eine geordnete Zusammenstellung von Begriffen und ihren (vorwiegend natürlichsprachigen) Bezeichnungen, die in einem Dokumentationsgebiet zum Indexieren, Speichern und Wiederauffinden dienen. Er ist durch folgende Merkmale gekennzeichnet:

a) Begriffe und Bezeichnungen werden eindeutig aufeinander bezogen („terminologische Kontrolle“), indem
    - Synonyme möglichst vollständig erfasst werden,
    - Homonyme und Polyseme besonders gekennzeichnet werden,
    - für jeden Begriff eine Bezeichnung (Vorzugsbenennung, Begriffsnummer oder Notation) festgelegt wird, die den Begriff eindeutig vertritt,
b) Beziehungen zwischen Begriffen (repräsentiert durch ihre Bezeichnungen) werden dargestellt.“

[DIN 1988] S. 12

Ein Thesaurus wird in der Regel erstellt, indem zunächst die Domäne, die der Thesaurus abdecken soll, die Spezifität und der Sprachstil festgelegt werden. Anschließend wird das für das fragliche Gebiet relevante Vokabular aus Quellen wie Nutzerbefragungen, Fachwörterbücher, aktuelle Fachliteratur etc. zusammengestellt. Die Mehrdeutigkeiten und Unschärfen der natürlichen Sprache werden in der Folge aufgelöst, indem diese Sammlung von natürlichsprachlichen Bezeichnungen in disjunkte Äquivalenzklassen aufgeteilt wird. Dabei steht jede Äquivalenzklasse für eine begriffliche Einheit der Domäne und umfasst alle Bezeichnungen, die innerhalb dieser Domäne den Begriff repräsentieren und als etwa gleichwertig angesehen werden können.

Um diese Aufteilung in Äquivalenzklassen zu erreichen, müssen neben der Erfassung aller Synonyme insbesondere Bezeichnungen mit mehreren Bedeutungen verschiedenen Äquivalenzklassen zugeordnet werden. Dabei kann es sich um Bezeichnungen handeln, die durch sprachliche Entwicklung zur gleichen Zeichenfolge geworden sind (*Homonyme*) wie z. B. „Tenor“ oder um Bezeichnungen, die sich ausgehend von einer Bedeutung unterschiedlich entwickelt haben (*Polyseme*), wie z. B. „Leitung“. Neben der Zuordnung zu Äquivalenzklassen werden die verschiedenen Bedeutungen in der Regel unterschieden, indem die nicht eindeutigen Bezeichnungen durch Synonyme ersetzt oder durch ergänzende Zusätze qualifiziert werden. Bei Thesauri mit Vorzugsbenennung wird nach der Klassenbildung eine Bezeichnung ausgewählt, die die Klasse repräsentiert und ausschließlich für Ressourcenbeschreibung und Recherche verwendet werden soll. Diese Vorzugsbenennung wird *Deskriptor* genannt; alle sonstigen Bezeichnungen der Klasse haben den Status von *Nicht-Deskriptoren*.[Burkart 1997] S. 163ff.

Nicht-Deskriptoren sollten grundsätzlich auf den zu verwendenden Deskriptor verweisen und mit einem Deskriptor werden in der Regel auch die Nicht-Deskriptoren angezeigt, um den Bedeutungsumfang der Äquivalenzklasse darzustellen. Somit existiert eine Relation zwischen Deskriptor und Nicht-Deskriptoren, die als Äquivalenzrelation bezeichnet wird. Daneben bestehen Beziehungen zwischen den zunächst isoliert stehenden Äquivalenzklassen, die mit Hilfe weiterer Relationen zu einem Beziehungsgefüge ausgebaut werden. DIN 1463 sieht die hierarchische Relation für Ober-/ Unterbegriffsbeziehungen und die Assoziationsrelation für Beziehungen zwischen verwandten Begriffen vor. Weitere Relationen, z. B. zur Darstellung von chronologischen Beziehungen, Kausalbeziehungen, Materialbeziehungen etc. sind möglich. [Burkart 1997] S. 171, [Wersig 1985] S. 134f

Ein Thesaurus beseitigt im Gegensatz zu Taxonomien mit der beschriebenen terminologischen Kontrolle einen Teil der Ungenauigkeiten natürlicher Sprache innerhalb eines Vokabulars. Ferner wird die Semantik der Bezeichnungen durch das Beziehungsgefüge teilweise dargestellt und die Relationen erleichtern die Ermittlung des treffendsten Begriffes für einen Sachverhalt, indem sie zusätzliche Sucheinstiege in das Vokabular und eine Struktur zur Steuerung der Suche zur Verfügung stellen.

4.2.3 Ontologie

Während die oben genannten Strukturierungsformen von kontrollierten Vokabularen ihre Ursprünge v. a. im bibliothekarischen bzw. dokumentarischen Bereich haben, gehen Ontologien ursprünglich auf eine philosophische Disziplin zurück. Inzwischen wurde der Begriff von der Informatik aufgegriffen und spielt v. a. im Bereich Künstliche Intelligenz, Computerlinguistik und Datenbanktheorie eine Rolle, wobei

Ontologien eher als Objekte verstanden werden, mit deren Hilfe die Konzepte einer Domäne formal spezifiziert werden können, um Domänenwissen zu teilen und wiederzuverwenden. Auf die in Kapitel 3.5.1 im Zusammenhang mit Ontologiesprachen angegebene Definition von [Gruber 1993] und die zugehörigen Ausführungen wird hingewiesen. [Guarino, Giaretti 1995], [Mädche et al 2001]

Der Begriff wird auch innerhalb der Informatik noch sehr unterschiedlich verstanden und teilweise zur Dekoration herkömmlicher Strukturierungsformen und –methoden verwendet [Guarino 1998]. Ontologien im engeren Sinne zeichnen sich jedoch im Unterschied zu den vorangegangenen Formen dadurch aus, dass neben Relationen auch Eigenschaften für Begriffe festgelegt werden können, die deren Bedeutung spezifizieren. Ferner können Axiome angegeben werden, die die möglichen Interpretationen des verwendeten Vokabulars weiter einschränken. [Mädche et al 2001], [Gruber 1993]

Ein formales Modell einer Ontologie könnte wie folgt aussehen:

Eine abstrakte Ontologie ist ein 6-Tupel $O := (C, R, H^C, H^R, \sigma, A^O)$, dessen Komponenten wie folgt definiert sind:

- Menge C von Begriffen
- Menge R von Relationen
- C und R sind disjunkt
- Eine partielle Ordnung H^C auf C,die Begriffshierarchie oder Taxonomie genannt wird.
- Eine partielle Ordnung H^R auf R, die Relationenhierarchie genannt wird.
- Eine Funktion $\sigma: R \rightarrow C \times C$, die zwei Begriffe mit einander assoziiert, wobei $\forall R: \sigma(R) \notin H^C$
- Menge der Axiome A^O, ausgedrückt in einer Logiksprache

Nach der abstrakten Definition von Begriffen und Beziehungen einer Domäne muss festgelegt werden, durch welche Symbole diese repräsentiert werden:

Ein Lexikon für die abstrakte Ontologie $O := (C, R, H^C, H^R, \sigma, A^O)$ ist ein 4-Tupel $L := (L^C, L^R, F, G)$, das aus folgenden Komponenten besteht:

- Zwei Mengen L^C und L^R, deren Elemente Symbole (lexical entries) für Begriffe bzw. Relationen genannt werden.
- Zwei Relationen $F \subseteq L^C \times C$ und $G \subseteq L^R \times R$, die Referenzen für Begriffe bzw. Relationen genannt werden.
  Basierend auf F wird für $l \in L^C$ definiert:
  $$F(l) = \{c \in C \mid (l,c) \in F\}$$
  und für
  $$F^{-1}(c) = \{l \in L^C \mid (l,c) \in F\}.$$
  Dies gilt analog für G und G^{-1}.

Eine konkrete Ontologie ist somit ein Paar (O,L), wobei O eine abstrakte Ontologie und L ein Lexikon für O ist. [Guarino 1998]

Im Unterschied zu Taxonomien und Thesauri können Vokabulare, die mit Hilfe von Ontologien formuliert wurden, grundsätzlich auch mit logikbasierten Sprachen ausgewertet werden. Ferner können Ontologien in der Kombination mit Auszeichnungssprachen dazu genutzt werden, um inhaltliche Beziehungen innerhalb eines Dokumentes darzustellen. Insofern bieten sie Möglichkeiten zur

semantischen Repräsentation von Ressourcen, die über die Bereitstellung eines adäquaten Vokabulars hinausgehen.

5. Online-Schlagwortnormdatei (OSWD)

Nachfolgend wird mit der OSWD ein kontrolliertes Vokabular vorgestellt, dessen Erweiterung Gegenstand des in Kapitel 6 und 7 vorgestellten Pilotprojektes ist. Hierzu werden zunächst das bisherige Entstehungs- und Einsatzgebiet des Vokabulars und seine Merkmale näher beleuchtet; anschließend wird seine Nutzbarmachung und sein Einsatz für die Online-Erschließung im WWW beschrieben.

5.1 Bibliothekarische Inhaltserschließung

Eine zentrale Aufgabe der Bibliotheken ist die Erschließung von Literatur und Medien. Dabei werden die fraglichen Objekte nach formalen und inhaltlichen Kriterien beschrieben und in die Kataloge der Bibliotheken aufgenommen. Neben Katalogen, die die Objekte beispielsweise nach ihren Autoren, nach ihrem Standort oder ihrer Zugehörigkeit zu einem Wissensgebiet beschreiben, ist der *Schlagwortkatalog* ein wesentliches Instrument der Inhaltserschließung. Er soll Auskunft darüber geben, welche Objekte eine Bibliothek zu einem bestimmten Thema führt. Zur Bildung des Kataloges wird jedes Werk mit Schlagwörtern beschrieben, die seinen sachlichen Inhalt möglichst kurz, genau und vollständig wiedergeben. Da ein einzelnes Schlagwort hierfür häufig nicht ausreicht, besteht in der verbal-syntaktischen Sacherschließung die Möglichkeit, Folgen von Schlagwörtern, sogenannte *Schlagwortketten*, im Rahmen der Beschreibung zu bilden. Die Menge an Schlagwörtern bzw. Schlagwortketten mit den Verweisen auf die beschriebenen Werke wird traditionellerweise alphabetisch sortiert, um eine leichte Benutzung des Schlagwortkataloges zu ermöglichen.
[Hacker 2000] S. 170, 172

Die Aufnahme in den Schlagwortkatalog ist ebenso wie die weiteren Katalogisierungsarten ein arbeitsaufwändiger Vorgang, der grundsätzlich von jeder Bibliothek in Anschluss an den Erwerb einer Werkes durchzuführen ist. Um diesen Prozess zu rationalisieren, organisieren sich Bibliotheken in Katalogisierungs-Verbünden, die gewährleisten sollen, dass jede teilnehmende Bibliothek nur die Neuanschaffungen katalogisieren muss, die noch nicht im Verbundkatalog nachgewiesen sind. Im übrigen kann sie für ihre Neuzugänge auf Datensätze zurückgreifen, die bereits von den anderen Verbundteilnehmern eingebracht wurden. Bibliotheksverbünde bestehen in der Regel aus einer zentralen Datenbank, die online für das Einbringen von Katalogdatensätzen und die Recherche zur Verfügung stehen, und lokalen Datenbanken der angeschlossenen Bibliotheken, die lokal bedeutsame Informationen z. B. zu Erwerbung, Ausleihstatus etc. enthalten. Der Südwestdeutsche Bibliotheksverbund, dessen Zentrale das Bibliotheksservice-Zentrum Baden-Württemberg bildet, ist ein Beispiel für ein regionales Verbundsystem in Deutschland: Er umfasst wissenschaftliche Bibliotheksbestände in Baden-Württemberg, Sachsen, Saarland und Teilen von Rheinland-Pfalz mit derzeit ca. 9 Millionen Titeln und über 26 Millionen Bestandsmeldungen.
[Hacker 2000] S. 233ff., [BSZ 2002]

Neben der arbeitsteiligen kooperativen Katalogisierung spielen im Zusammenhang mit dem Schlagwortkatalog v. a. Instrumente eine Rolle, die die Vereinheitlichung und Vereinfachung der Schlagwortvergabe zum Ziel haben:
Dies sind einerseits Regelwerke wie die in Deutschland allgemein verwendeten *Regeln für den Schlagwortkatalog (RSWK)*. Um einem Teil der in Kapitel 4 erwähnten Mehrdeutigkeiten natürlicher Sprache zu begegnen, werden darin Prinzipien für die Schlagwortvergabe definiert. So ist in den RSWK beispielsweise festgelegt, dass stets das spezifischste Schlagwort anzusetzen ist, welches einen

Sachverhalt beschreibt, und bei Vorliegen von Synonymen der am häufigsten gebrauchte Ausdruck auszuwählen ist. Ferner legen die RSWK für die Bildung von Schlagwortketten eine Syntax fest, die deren Aussagekraft erhöhen soll. Danach soll in einer Kette beispielsweise das Objekt einer Handlung vor dem Subjekt stehen, ein Personenschlagwort vor einem Geographikum, dieses wiederum vor einem Sachschlagwort etc.
Als weiteres Hilfsmittel werden kontrollierte Vokabulare eingesetzt, deren Konzepte zu großen Teilen im Bereich der bibliothekarischen Sacherschließung entwickelt wurden. Ein Beispiel hierfür ist die *Schlagwortnormdatei (SWD)*, die nachfolgend vorgestellt wird.
[Hacker 2000] S. 196ff, [DBI 1990] S. 23f

5.2 Schlagwortnormdatei (SWD)

Die *Schlagwortnormdatei (SWD)* ist ein Thesaurus im Sinne der Ausführungen in Kapitel 4.2.2 und entspricht den Anforderungen nach DIN 1463. Sie wurde seit Mitte der 1980er Jahre als Instrument der kooperativen Sacherschließung im deutschsprachigen, wissenschaftlichen Bibliotheksbereich aufgebaut.

5.2.1 Merkmale

Die SWD zeichnet sich durch folgende Merkmale aus:

- **Universales Vokabular**
  Abgesehen von dem Kontext, den die Entwicklung im Bereich wissenschaftlicher Bibliotheken mit sich bringt, ist die SWD nicht auf eine bestimmte Thematik oder einen Fachbereich eingegrenzt. Entsprechende hohe Anforderungen werden an die Kontrolle des Vokabulars und das Relationsgefüge gestellt.

- **Terminologische Kontrolle**
  Neben der Erfassung der Synonyme werden Homonyme und Polyseme unterschieden, indem sie i. d. R. mit einem qualifizierenden Zusatz in Winkelklammern versehen werden. Beispiel: „Tau <Niederschlag>“ versus „Tau <Seil>“.

- **Thesaurus mit Vorzugsbezeichnung**
  Die SWD sieht eine Vorzugsbezeichnung vor, die als *Ansetzung* bezeichnet wird. Sie kann aus einem oder mehreren Wörtern bestehen. Es kann sich auch um eine Kette von Schlagwörtern *(mehrgliedriges Schlagwort)* handeln, die durch Schrägstrich (/) getrennt werden und eine innere Struktur im Sinne der angesprochenen RSWK-Syntax haben.

- **Relationen**
  In der SWD existieren im wesentlichen folgende Relationen, die auch als *Verweisungen* bezeichnet werden:
  - Synonym- bzw. Äquivalenzrelation
    Bei der Ausgabe von Synonymen zu einer Ansetzung liegt insofern ein Sonderfall vor, als nicht auf einen weiteren Deskriptor verwiesen, sondern Nicht-Deskriptoren angezeigt werden.
  - Hierarchische Relation

Verweisungen zwischen Ober- und Unterbegriffen, wobei die Beziehung im allgemeinen generisch-abstrakt ist. D. h. der Unterbegriff enthält alle Merkmale des übergeordneten und weitere spezifizierende.

- Assoziationsrelation
  Verweisungen zwischen verwandten Begriffen.
- Chronologische Relation
  Früher- / Später-Verweisungen, v. a. für geographische Gebilde und Institutionen, deren Bezeichnung sich im Laufe der Zeit geändert hat.
- Definitorische Verweisungen (Begriffskombination)
  Eine Ansetzung wird durch eine Kombination von Begriffen, die selbst Ansetzungen sind, definitorisch beschrieben. Der beschriebene Begriff kann als Schnittmenge der Äquivalenzklassen der beschreibenden Begriffe angesehen werden. Die Beschreibung hat dabei nicht den Charakter einer Ansetzung, sondern nur ihre Bestandteile. Beispiel: „Chemie / Zeitschrift / USA“ für „Chemical abstracts“

- **Regelwerk**
  Die Bildung der Ansetzungs- und Verweisungsformen der Schlagwörter richtet sich grundsätzlich nach den RSWK. So gilt beispielsweise für die Wahl der Ansetzung, dass der gebräuchlichsten Bezeichnung der Vorzug zu geben ist. Soweit erforderlich werden die RSWK durch zusätzliche Praxisregeln ergänzt.

- **Ergänzende strukturierende Elemente**
  Zusätzlich zur Verweisungsstruktur erfolgt eine Systematisierung eines Großteils der Begriffe nach einer Grobklassifikation, die 36 Hauptgruppen umfasst. Diese Hauptgruppen, die ein Wissensgebiet repräsentieren, sind überwiegend nochmals mehrfach in Untergruppen unterteilt. Zusätzlich werden Schlagwörter mit geographischem Bezug mit einem Ländercode (nach DIN EN ISO 23166) systematisiert. Daneben ist die Vergabe von Sprachen- und Zeitcodes vorgesehen.

- **Zusatzangaben**
  Ergänzende Angaben wie Quellen einer Ansetzung, Definitionen und Verwendungshinweise dienen als zusätzliche Hilfestellungen für den Gebrauch der Schlagwörter und zur Qualitätssicherung des Thesaurus.

- **Satzstruktur**
  Eine Ansetzung wird gemeinsam mit den zugehörigen Verweisungen und den Zusatzangaben in einem Datensatz gehalten, der mit einem eindeutigen Schlüssel (SWD-Identifikationsnummer) versehen ist.
- **Zeichensatz**
  Mit ISO 646 bzw. ISO 5426-1983 wird ein mächtiger Zeichensatz verwendet, der die Darstellung zahlreicher Sonderzeichen und Diakritika erlaubt.

[DDB 2000] S. 4ff, [DBI 1990] S. 20f

Die SWD enthält derzeit über 550.000 Schlagwortsätze, zu denen insgesamt mehr als 860.000 Verweisungen existieren. Pro Jahr kommen ca. 40.000 neue Datensätze hinzu. Die SWD wird als Online-Datenbank bei Der Deutschen Bibliothek geführt; die zentralen Redaktionsstellen der teilnehmenden Organisationen haben schreibenden und lesenden Zugriff auf den Datenbestand. Die Bibliotheksverbünde

führen in der Regel in ihren eigenen Datenbanken eine lokale SWD, deren Schlagwortsätze mit den im Verbundkatalog erfassten Titeln verknüpft sind. Diese lokalen Schlagwortdaten werden durch wöchentliche Updates aus der zentralen, gemeinsamen SWD aktualisiert. Daneben wird der Gesamtbestand der SWD halbjährlich auf CD-ROM und Mikrofiche von Der Deutschen Bibliothek ausgeliefert.

5.2.2 Kooperation

Für den Nutzen eines Thesaurus ist u. a. entscheidend, dass ein für die Ressourcenbeschreibung adäquates Vokabular ausgewählt und dieses Vokabular und die Thesaurusstruktur laufend der sprachlichen Entwicklung im Anwendungsgebiet angepasst wird (vgl. [Burkart 1997] S. 163, 178, [Wersig 1985] S. 274ff).

Dieser Anforderung entspricht die SWD v. a. dadurch, dass sie kooperativ entwickelt und gepflegt wird: Jeder Teilnehmer hat das Recht und die Pflicht Schlagwortsätze einzubringen und gegebenenfalls Änderungen zu verlangen. Bislang nehmen folgende Einrichtungen an der SWD-Kooperation teil:

- Die Deutsche Bibliothek
- Bibliotheksverbund Bayern (BVB)
- Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (HBZ)
- Südwestdeutscher Bibliotheksverbund (SWB)
- Gemeinsamer Bibliotheksverbund der Länder Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein und Thüringen (GBV)
- Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV)
- Verbund der wissenschaftlichen Bibliotheken Österreichs
- Schweizerische Landesbibliothek (SLB)
- Kunstbibliotheken-Fachverbund Florenz-München-Rom
- Hessisches Bibliotheks-Informationssystem (HEBIS)

[DDB 2000] S. 4

Konkret bedeutet dies, dass beispielsweise eine Bibliothek in einem der genannten Verbünde bei der Katalogisierung eines neu erworbenen Werkes feststellt, dass ein zur inhaltlichen Beschreibung benötigtes Schlagwort nicht in der SWD vorhanden ist. Sie meldet den gewünschten Schlagwortsatz daraufhin mit den erforderlichen Zusatzangaben wie Quelle, Verweisungen etc. über die Verbundzentrale zur Neuaufnahme in die SWD an.

Durch diese Vorgehensweise wird zum einen sichergestellt, dass das Vokabular nicht nur theoretischen Überlegungen eines Thesaurusentwicklers entspricht, sondern an den tatsächlich zu beschreibenden Ressourcen ausgerichtet ist. Ferner führt die Vielzahl der Teilnehmer dazu, dass eine einseitige Schwerpunktbildung vermieden wird und der Wortschatz repräsentativ für den Bedarf der allgemeinen Sacherschließung im deutschsprachigen Raum bleibt. Etablierte Verfahren der Stellungnahme aller Kooperationspartner tragen außerdem zur laufenden Kontrolle und Qualitätssicherung bei. Die Form und Organisation der Kooperation garantiert ferner bereits die laufende Anpassung; es ist keine Abhängigkeit von festen Revisions- und Pflegeterminen gegeben.

5.2.3 Qualitätssicherung

Die Vorteile einer kooperativen Erstellung und Pflege des Vokabulars bringen allerdings auch Gefahren für die Qualität mit sich. Falls die Wünsche aller Benutzer in beliebiger Form übernommen würden, wäre langfristig mit Problemen durch Einbringung von Dubletten, nicht gebräuchlichen Bezeichnern, zu spezifischen Begriffen oder Modebegriffen zu rechnen. Ferner könnten sich Inkonsistenzen durch unterschiedliche Handhabung der Relationen und uneinheitliche Ansetzungsformen ergeben. Eine derartige Qualitätsminderung würde die Verwendung des Thesaurus, dessen Nutzen gerade auf Konsistenz und Struktur des Vokabulars beruht, in absehbarer Zeit uninteressant erscheinen lassen.

Die SWD sieht daher insbesondere folgende Maßnahmen der Qualitätssicherung vor:

Nutzung eines Regelwerks

Indem die RSWK nicht nur für die Verschlagwortung verwendet werden, sondern auch weitgehend für die Bildung der Schlagwortsätze der SWD maßgeblich sind, stehen grundlegende Regeln für die Auswahl und Syntax von Schlagwörtern und die Strukturierung im Thesaurus zur Verfügung.

Referenzprinzip

Wird ein neues Schlagwort vorgeschlagen, so ist dessen Gebräuchlichkeit durch die Angabe einer Quelle nachzuweisen. Als Referenz ist jeweils die neueste deutschsprachige Allgemeinenzyklopädie oder ersatzweise ein sonstiges Nachschlagewerk nach einer festgelegten, fachlichgeprägten Liste („Liste der fachlichen Nachschlagewerke“) zu verwenden, die laufend aktualisiert wird.

Weitere Pflichtangaben

Zusätzlich zu Schlagwort und Quelle sind vom Vorschlagenden eine Reihe von Pflichtangaben zu machen, die einerseits der Erhaltung des Beziehungsgefüges und der Systematik des Thesaurus dienen und andererseits die Eingrenzung des Begriffsumfangs eines einzelnen Schlagwortes ermöglichen. So sind in der Regel Angaben zur Einordnung in die SWD-Systematik und die Zuordnung eines Ländercodes erforderlich. Bei Schlagwörtern, deren Bedeutung sich nicht aus der Bezeichnung ergibt, sollten Definitionen angegeben werden; weitere Erläuterungen wie Benutzungs- oder redaktionelle Hinweise sind möglich. Bei Personen sind zusätzlich Lebensdaten und Berufsbezeichnung erforderlich, um Unterscheidungen bei Namensgleichheit zu ermöglichen.

Redaktionssystem

Vorschläge und Korrekturwünsche eines Indexierers führen nicht unmittelbar zu Änderungen im Datenbestand. Sie werden vielmehr zunächst durch Redaktionen inhaltlich und formal überprüft, ggf. nachbearbeitet oder zurückverwiesen. Diese Redaktionsaufgaben werden in der Regel durch eine oder mehrere Fachredaktionen innerhalb eines Bibliotheksverbundes wahrgenommen. Erst wenn ein Schlagwortsatz bzw. ein Änderungswunsch den Qualitätskriterien genügt, führt die Redaktion die entsprechende Änderung auf dem Online-Datenbestand Der Deutschen Bibliothek durch. Hierbei werden

nochmals maschinelle Konsistenzprüfungen z. B. hinsichtlich des Vorhandenseins von Dubletten, der Angabe von Pflichtfeldern etc. vorgenommen.

Kommunikation und Konsensbildung

Für den erfolgreichen Aufbau und die Pflege eines Vokabulars ist eine Kommunikation innerhalb der Entwicklungs- und Nutzergemeinschaft zwingend erforderlich. Auf diesem Wege können Problemfälle diskutiert und Einigungen zu Entwicklungsfragen erzielt werden. Ferner sind den Redaktionen bei Unklarheiten Rückfragen und Mitteilungen zu häufigen Fehlern möglich. Einzelanwender können sich ihrerseits Rat bei den Redaktionen einholen. Zu diesem Zweck wurden datenbankgestützte Mailboxsysteme etabliert: Ein Schlagwortsatz in einer Verbunddatenbank kann ein Feld enthalten, in welches adressierte Mailboxnachrichten eingetragen werden können. Jeder Verbundteilnehmer kann alle zu einem Datensatz vorhandenen Nachrichten einsehen, eigene Nachrichten verfassen, an ihn gerichtete Nachrichten abrufen etc. Ein entsprechender Mechanismus existiert in der Online-Datenbank Der Deutschen Bibliothek für die Kommunikation zwischen den Verbundredaktionen und Der Deutschen Bibliothek.

5.3 Aufbau der OSWD

Im Zusammenhang mit den Technologien zur semantischen Nutzung des WWW wurden bereits Beispiele für Initiativen aus dem bibliothekarischen Bereich genannt, die auf die Erschließung von Online-Ressourcen abzielen. Ein zentrales Anliegen ist die Verknüpfung der herkömmlichen Erschließung von gedruckten Werken mit der Erschließung von Publikationen im WWW und der integrierte Nachweis, beispielsweise in Online-Katalogen.

Interessante Online-Publikationen bieten u. a. die virtuellen Forschungsbibliotheken, die in den vergangenen Jahren an Universitäts- und Landesbibliotheken entstanden sind. Diese Systeme sehen vor, dass Autoren wissenschaftliche Dokumente in elektronischer Form auf Textservern der jeweiligen Bibliotheken abspeichern können. Die Server sind über das WWW für Recherche und Abruf der Dokumenten zugänglich. Die Beschreibung der Online-Publikationen mit Metadaten hat üblicherweise durch den Autor beim Einbringen der Dokumenten zu erfolgen, wobei u. a. die Vergabe von Schlagworten vorgesehen ist. In diesem Zusammenhang stellt sich das in Kapitel 4 dargestellte Problem, wonach eine aussagekräftige inhaltliche Beschreibung in der Regel ein kontrolliertes Vokabular erfordert. Die freihändige Vergabe von Schlagworten führte auch hier zu den erwähnten Problemen einer mehrdeutigen und uneinheitlichen Auszeichnung. Die SWD, die sich aufgrund ihrer Ausrichtung auf wissenschaftliche Bibliotheken als Instrument zur Verschlagwortung anbietet, war in der bisherigen Form jedoch nicht für eine allgemeine Nutzung im WWW geeignet. Die wöchentlichen Auslieferungen Der Deutschen Bibliothek erfolgen in einem bibliothekarischen Eigenformat; die Sichten auf den SWD-Datenbestand, die über die Online-Kataloge möglich sind, sind wenig nutzerfreundlich und erfordern bibliothekarische Vorkenntnisse.

Daher wurde im Rahmen einer Diplomarbeit [Eggers 2001] und meines Individualpraktikums, welches auf die vorliegende Masterarbeit hinführte, die Online-Datenbank OSWD aufgebaut, die die SWD über das WWW recherchierbar macht. Hierzu werden die aktuellen SWD-Daten regelmäßig in ein für das WWW geeignetes Format konvertiert und in eine relationale Datenbank (Oracle) eingespielt. Die mit

Hilfe eines Apache-Webservers und der Programmiersprache Perl realisierte Schnittstelle zum WWW bietet dem Benutzer vielfältige Suchmöglichkeiten und erlaubt eine Navigation im Thesaurus anhand der Relationen.

Darüberhinaus wurde für den *Online Publikationsverbund der Universität Stuttgart (OPUS)*, ein Konzept realisiert, welches die automatisierte Übernahme von Schlagwörtern bei der Ressourcenbeschreibung ermöglicht. OPUS ist ein System für die Publikation, Archivierung und das Management von Hochschulschriften aller Art im Sinne der vorstehenden Ausführungen, welches in Deutschland weite Verbreitung gefunden hat. Das Einbringen der Dokumente auf einen Volltextserver erfolgt per File-Upload über ein HTML-Formular. Mit Ausfüllen der im Formular vorgegebenen Felder zeichnet der Autor die Publikation zugleich mit den erforderlichen Metadaten aus. Durch eine Kombination aus dynamischem Aufbau der Webseiten und Datenaustausch mittels HTML-Formularen, konnte erreicht werden, dass der Nutzer im Rahmen der Beschreibung seiner Publikation die OSWD aufrufen, geeignete Schlagwörter auswählen und diese automatisch übernehmen kann. Eine unerwünschte Modifikation der normierten Vokabulardaten ist dabei nicht möglich. Der sonstige Beschreibungsvorgang wird durch die Übernahme der Schlagwörter nicht beeinträchtigt. Eine Demonstration dieser Anwendung ist unter folgender URL zu finden:

http://www.bsz-bw.de/opus/uni/opus-eingabe.php

Die Struktur der Metadaten in OPUS basiert auf Dublin Core. Die Metadatenbeschreibungen der erfassten Publikationen werden unter anderem in die Datenbank des Südwestdeutschen Bibliotheksverbundes eingespielt und sind somit in dessen Katalog und den abgeleiteten lokalen Katalogen nutzbar. Neben Dublin Core arbeitet OPUS mit Semantic Web-Technologien auf der Ebene von XML und RDF. So wurde testweise das XML-basierte Datenformat *DLmeta* zum Datenaustausch zwischen Projekten der Digitalen Bibliothek implementiert. Ferner wird eine Integration in die *Open Archives Initiative (OAI)* angestrebt, die den Austausch von Metadaten unter Verwendung von XML und RDF realisiert.
[Stephan, Scholze 1999]

6. Projekt: Erweiterung der SWD

6.1 Motivation: Ausweitung des Anwendungsbereiches

Zentrale Methode der Technologien zur semantischen Nutzung des WWW ist die Beschreibung von Ressourcen aus unterschiedlichen Domänen und Wissensbereichen mit Metadaten, deren Semantik allgemein akzeptiert oder zumindest für Abbildungen auf eine gemeinsame Basis zugänglich ist. Dadurch wird effektives, interdisziplinäres Retrieval und die Kombination von Informationen aus verschiedenen Bereichen ermöglicht.

Dieser Vorgehensweise folgt auch das DFG-Projekt „Gemeinsames Portal für Bibliotheken, Archive und Museen (BAM) – Ein Online-Informationssystem“, an dem das Bibliotheksservice-Zentrum Baden-Württemberg, die Landesarchivdirektion Stuttgart und das Landesmuseum für Technik und Arbeit in Mannheim beteiligt sind. Ziel des Projektes ist es, die online verfügbaren, aber von einander isolierten Bestände kultureller Information in den beteiligten Einrichtungen unter einer gemeinsamen Oberfläche zu präsentieren und für eine integrierte Recherche zugänglich zu machen. Zum Einsatz kommt das Dublin Core Element Set als Grundlage für die Beschreibungselemente und XML als Auszeichnungssprache.
[Mainberger 2001]

Hierbei stellte sich das Problem, dass erhebliche Unterschiede bei der Art und der Einheitlichkeit der inhaltlichen Beschreibung von Ressourcen zwischen den verschiedenen Domänen gegeben sind. So existieren insbesondere für das Museumswesen nicht in dem Maße allgemeingültige Standards für die Dokumentation, wie sie beispielsweise im bibliothekarischen Bereich vorhanden sind. Es herrscht u. a. kein allgemeiner Konsens über die Verwendung von kontrollierten Vokabularen; ein generell akzeptierter Thesaurus wie die SWD ist nicht vorhanden. Der Aufbau eines solchen Vokabulars dürfte aufgrund des zeitlichen und materiellen Aufwands, den die Qualitätsanforderungen und die Konsensbildung mit sich bringen, nur langfristig möglich sein. Andererseits steht mit der SWD bereits ein wissenschaftlicher Universalthesaurus zur Verfügung, der über längere Zeit kooperativ entwickelt wurde und zahlreiche Maßnahmen zur Qualitätssicherung vorsieht. Diese Überlegungen, die auch im Rahmen der Einführung eines landeseinheitlichen *Museumsinformationssystems (MUSIS)* angestellt wurden, führten zur Entscheidung, die SWD zur Beschreibung von Objekten in Museen und Archiven einzusetzen.

Es muss jedoch weiterhin auf die unterschiedliche Ausrichtung der Ressourcenbeschreibung in den verschiedenen Domänen Rücksicht genommen werden. Die SWD wurde in erster Linie für die Beschlagwortung von Literatur entwickelt, d. h. Begriffe tauchen dort i. d. R. nur auf, wenn sie ein wesentlicher Gegenstand einer Publikation sind. Museen sind dagegen stärker an der Beschreibung von konkret vorhandenen Objekten interessiert, die Bestandteile einer Sammlung sind. Hierbei sind häufig sehr feine begriffliche Differenzierungen erforderlich und sorgfältig entworfene hierarchische Klassifikationen spielen bei der Dokumentation eine große Rolle. Die Tauglichkeit der SWD für den Einsatz im Museumswesen wurde daher in bisherigen Evaluationen insbesondere in Hinblick auf den mangelnden Begriffsumfang und die unzureichenden Hierarchiestrukturen bezweifelt.
[Ludewig 2002]
Dieses Dilemma könnte möglicherweise behoben werden, wenn die Museen und Archive nicht nur die Möglichkeit haben, den Thesaurus passiv zu nutzen, sondern auch in der Lage sind, ihn nach ihren Beschreibungsbedürfnissen mitzugestalten und zu erweitern. Die aktive Mitwirkung könnte letztlich sicherstellen, dass tatsächlich ein gemeinsames Vokabular entsteht, welches die integrierte semantische

Nutzung der Ressourcen aus den verschiedenen Domänen ermöglicht. Diese Fragestellung soll in einem Pilotprojekt, an dem Die Deutsche Bibliothek, das Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) und Museen und Archive aus Baden-Württemberg teilnehmen, geklärt werden. Hierzu habe ich im Rahmen der vorliegenden Arbeit ein Konzept realisiert, welches im Unterschied zur bisherigen OSWD nicht nur die Nutzung des vorhandenen Thesaurus im WWW, sondern auch das Einbringen von neuen Datensätzen und Korrekturwünschen ermöglicht.

Eine Demonstrationsversion der über das WWW zugänglichen Teile der Anwendung ist unter folgenden URLs zu finden:

Die Schlagworterfassung:
http://www.bsz-bw.de/cgi-bin/mustest_start.pl

Teile der Redaktion und eine Demonstration der Datenkonvertierung:
http://www.bsz-bw.de/cgi-bin/oswd-red_test.pl

6.2 Anforderungen

Die in Kapitel 5 kurz dargestellte SWD-Kooperation erfolgt innerhalb der Bibliotheksverbünde auf Grundlage der Verbundsysteme. Zwischen den Fachredaktionen der Verbünde und Der Deutschen Bibliothek wird das *Integrierte Literatur-, Tonträger und Musikalien-Informations-System (ILTIS)* Der Deutschen Bibliothek genutzt, dessen Komponenten mit Software des niederländischen Herstellers PICA betrieben werden. Im weiteren wird die für das Gesamtsystem gängige Bezeichnung PICA/ILTIS verwendet. Dabei handelt es sich um eine Client-Server-Architektur mit zentraler Datenbank und Katalogisierungssystem auf einem TANDEM-Großrechner bei Der Deutschen Bibliothek und Katalogisierungsclients in Form von Windowsanwendungen bei den Fachredaktionen. Die Clients werden als *Intelligent Bibliographic Workstation (IBW)* bezeichnet. Die Kommunikation erfolgt mit Hilfe eines proprietären Protokolls auf Grundlage des TCP/IP-Protocol-Stacks.

Bei Entwurf des Projektes stand fest, dass die Museen und Archive in dessen Rahmen weder an einem Verbundsystem teilnehmen, noch direkt eine IBW nutzen können. Die einzubringenden Schlagwortkandidaten sollten über das WWW an das BSZ übertragen und dort zunächst zwischengespeichert werden. Dabei sollten gängige Browser und WWW-Protokolle genutzt werden; die Entwicklung einer proprietären Lösung schied aus. Nach ggf. erforderlicher redaktioneller Nachbearbeitung sollten die Schlagwortsätze vom BSZ mit Hilfe der dort verfügbaren IBWs in den Online-Datenbestand Der Deutschen Bibliothek eingebracht werden. Mit Auslieferung der aktualisierten SWD-Daten durch Die Deutsche Bibliothek und Einspielung in die OSWD im Rahmen der regelmäßigen Updates stünden die neuen Schlagwortsätze anschließend allgemein zur Ressourcenbeschreibung zur Verfügung. Somit ergibt sich der in Abbildung 12 dargestellte Datenfluss zwischen den Objektteilnehmern.

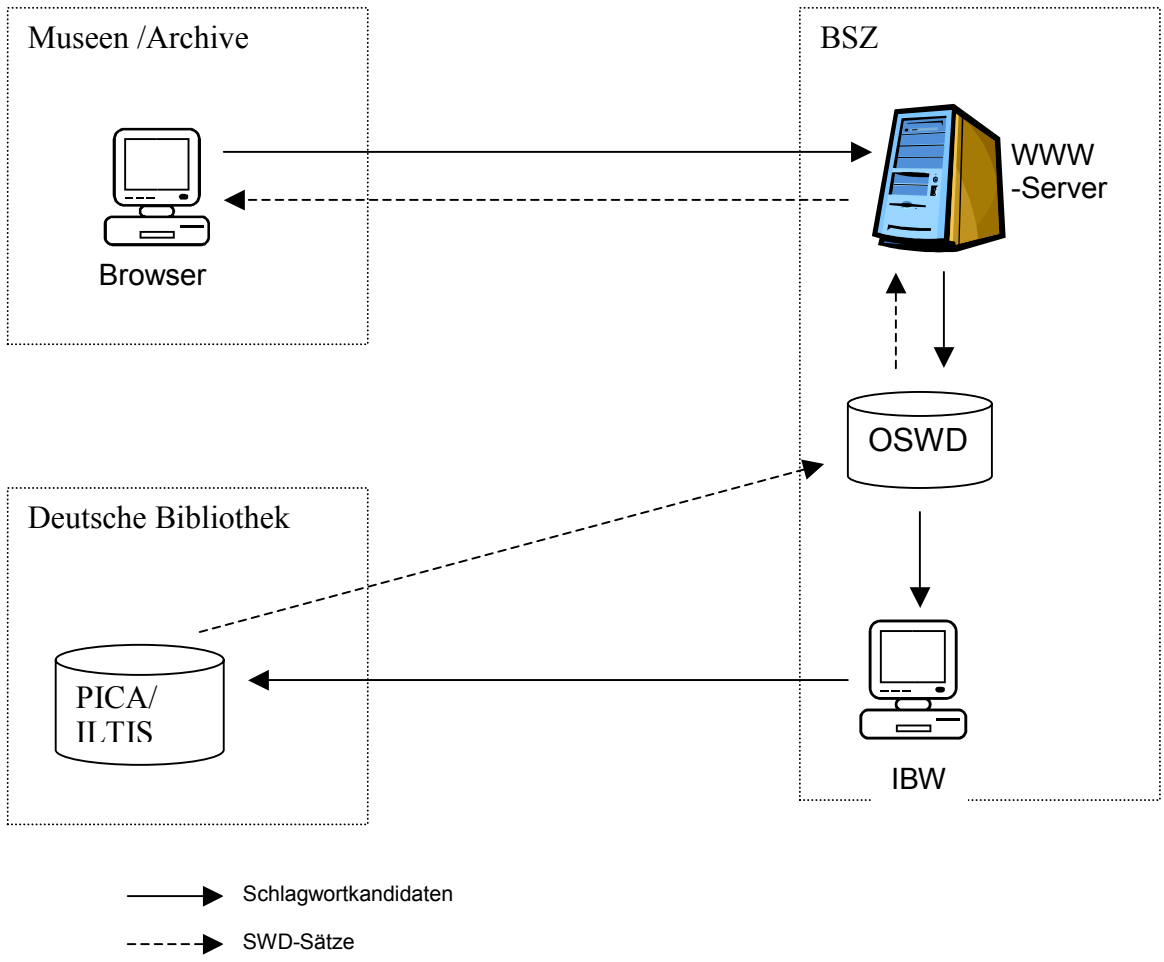


Abbildung 12: Datenfluss zwischen den Projektteilnehmern

Um die Arbeit der neuen SDW-Teilnehmer zu erleichtern, war die Speicherung der Schlagwortkandidaten in Form eines Online-Datenbestandes vorgesehen. Der Bestand an bereits eingebrachten Kandidaten sollte recherchierbar sein, wobei zu einem einzelnen Kandidaten der jeweilige Bearbeitungsstand anzuzeigen war. Weitere Funktionen sollten das Ändern und Löschen vorhandener Kandidatensätze und das Einbringen von Korrekturwünschen zu bereits in der SWD vorhandenen Schlagwortsätzen sein.

Da das Projekt die Erweiterung und Veränderung des Thesaurus zur Folge hat, spielten die Maßnahmen zur Qualitätssicherung der SWD eine große Rolle. So waren einerseits nicht nur die für die Nutzung interessanten Schlagwortdaten zu erheben, die von der bisherigen OSWD ausgegeben werden, sondern es musste eine Vielzahl von Feldern bei der Eingabe erfasst werden, die beispielsweise dem Nachweis der Gebräuchlichkeit, der Definition und der internen Strukturierung dienen. Ferner war eine eigene Redaktion zu organisieren, da das Projekt außerhalb der üblichen Arbeitsabläufe in der SWD-Kooperation abläuft und die üblichen Redaktionsverfahren nicht genutzt werden konnten. Diese Aufgabe wurde von Der Deutschen Bibliothek und dem BSZ übernommen, welches als Fachredaktion bereits über entsprechende Kenntnisse und Erfahrungen verfügt. In diesem Kontext war ein Redaktionstool zu schaffen, welches die Überwachung des Dateneingangs und gezielte Auswahl und Redaktion der Datensätze erlaubt. Ferner war eine Unterstützung der Redaktion durch automatisierte Plausibilitätsprüfungen bei der Eingabe der Kandidaten und Dublettenkontrolle wünschenswert.

Analog zur etablierten SWD-Kooperation sollten die neuen Teilnehmer ebenfalls in der Lage sein, sich über die Entwicklung des Vokabulars auszutauschen und Problemfälle der täglichen Arbeit zu diskutieren. Außerdem sollten Anfragen der Teilnehmer an die Redaktion und Rückmeldungen der Redaktion, z. B. über häufige Fehler, möglich sein. Somit war ein Kommunikationsmechanismus zu etablieren, der mit der Entwicklung des Datenbestandes verknüpft ist und u. a. einen forenartigen Austausch unterstützt.

Die Übermittlung der Datensätze in den Online-Datenbestand der DDB mit Hilfe der IBW erfordert die Einhaltung des PICA/ILTIS-Datenformats (vgl. Anlage 1). Dieses sieht Folgen von normierten Datenfeldern vor, für die jeweils detaillierte Formatvorgaben existieren. Verschiedene Felder werden durch führende Feldkennungen und ggf. Qualifikatoren unterschieden. Da die Einarbeitung in dieses komplexe Format von den Projektteilnehmern nicht erwartet werden konnte, war ein Programmmodul vorzusehen, welches die erhobenen Daten nach PICA/ILTIS konvertiert.

6.3 Planung und Entwicklung

Die Anforderungsbeschreibung und die grobe Datenflussskizze ließen erkennen, dass die Anwendung auf die Bearbeitung von Datensätze ausgerichtet sein würde, die in einer überschaubaren Folge von unterscheidbaren Schritten bearbeitet werden. Ferner waren weder intensive Interaktionen mit anderen Anwendungen noch zeitkritische Vorgänge vorgesehen. Insofern erschien eine Entwicklung, die sich an der strukturierten Analyse orientiert, wie sie z. B. bei [Balzert 1996] S. 398ff. beschrieben wird, angemessen.

6.3.1 Datenmodell und Datenfluss

Zunächst wurden im Rahmen der Produkt-Definition die relevanten Produktdaten festgelegt. Ausgangsbasis hierfür war die Feldbeschreibung des Zielformates PICA/ILTIS, wobei in Gesprächen mit den fachlichen Betreuern von BSZ und Der Deutschen Bibliothek zu klären war, welche Felder für das Projekt tatsächlich relevant sind.

Die berücksichtigten Produktdaten, sind im PICA/ILTIS-Format in Anlage 1 aufgeführt. Hierbei ist zu beachten, dass sich die Art und Form der Daten in Abhängigkeit davon unterscheidet, ob sich der einzubringende Schlagwortsatz auf eine Person, eine Organisation oder einen sonstigen Sachverhalt oder Gegenstand bezieht. Dies hängt damit zusammen, dass das unter Kapitel 5.2.1 erläuterte Prinzip von einem mit den Titeln verknüpften Online-Datenbestand auch auf weitere kooperativ entwickelte Normdateien angewandt wird, die v. a. im Rahmen der Formalerschließung eine Rolle spielen. Es handelt sich bei Personen um die *Personennamendatei (PND)* und bei Organisationen um die *Gemeinsame Körperschaftsdatei (GKD)*. Ein Schlagwortsatz, der eine Person repräsentiert, ist im Datenbestand Der Deutschen Bibliothek nur noch in der PND enthalten und wird erst bei Auslieferung eines SWD-Updates in das SWD-Format konvertiert und den sonstigen SWD-Daten beigefügt. Daher ist bei Datenlieferung das PND-Format einzuhalten, auch wenn inhaltlich ein Schlagwortsatz geliefert wird. In bezug auf die GKD ist ein solches Vorgehen derzeit noch nicht möglich, da sich die Normdateien hinsichtlich Beschreibungsregeln und -inhalten erheblich unterscheiden. Sofern eine in der SWD zu erfassende

Organisation bereits in der GKD vorhanden ist, wird lediglich ein Hinweis auf die entsprechende Identifikationsnummer in der GKD aufgenommen.

Im Anschluss an die Ermittlung der Produktdaten wurde ein Entity Relationship Modell erstellt, welches ebenfalls zwischen Personen- und sonstigem Schlagwortsatz unterscheidet. Die Mailboxfelder sind in Analogie zu den bestehenden Kommunikationsmechanismen in den Verbundsystemen und der Normdatenkooperation mit den Schlagwortsätzen verknüpft. Die Entity Relationship Diagramme sind in Anlage 2 und 3 enthalten.

Ferner wurden Datenflussdiagramme entworfen, die in Anlehnung an [Balzert 1996] S. 398 hierarchisch organisiert sind. Das Kontextdiagramm und das Diagramm für den Prozess 0 sind in Anlage 4 und 5, beispielhafte Verfeinerungen für die Kandidatenerfassung und Redaktion in Anlage 6 und 7 enthalten. Datenflüsse, die in erster Linie Bedienungsdialoge darstellen, wurden aus Gründen der Übersichtlichkeit nicht aufgenommen. Die IBW und die Übermittlung der Kandidaten zur Deutschen Bibliothek, die nicht im Rahmen dieser Anwendung realisiert wurden, werden hingegen mit angegeben, um den Datenfluss vollständig darzustellen.

6.3.2 Entwicklungsumgebung

Datenbankmanagementsystem

Wie die Ermittlung der Produktdaten und die Entity Relationship Diagramme (in Anlage 2 und 3) zeigen, sind im wesentlichen textuelle Datensätze abzubilden, deren Felder in Anzahl und Struktur variieren. Die Entitäten verfügen jedoch nicht über komplexe Eigenschaften; auch die Beziehungen zwischen ihnen sind überschaubar. Da das Zielformat PICA/ILTIS auf festen Spezifikationen beruht und deutschlandweit für die Online-Katalogisierung genutzt wird, ist ferner nicht mit häufigen Änderungen des Datenschemas zu rechnen. Daher erschien die Speicherung in einer relationalen Datenbank ausreichend hinsichtlich der Komplexität und vorteilhaft in bezug auf die Performanz des Zugriffs. Genutzt wurde das Datenbankmanagementsystem Oracle in der Version 8.1.7, da es einen leistungsfähigen Volltextindex bietet und die Mitnutzung der bereits für die OSWD bestehenden Datenbank erlaubte.

Nutzerschnittstelle

Die Vorgabe, dass die Nutzung der Anwendung mit einem Standardbrowser unter Verwendung der gängigen WWW-Protokolle und Sprache möglich sein sollte, legte die Verwendung von HTML-Formularen zur Dateneingabe und HTML-Seiten zur Darstellung von Daten und Dialoginformationen nahe. In HTML eingebettete Java Applets schieden als mögliche alternative Nutzerschnittstelle insofern aus, als die Anwendung auch für Nutzer verwendbar sein sollte, die Java in ihrem Browser deaktiviert haben oder nicht javafähige Browser verwenden. Aus demselben Grund wurde auch die Nutzung von JavaScript zur clientseitigen Dynamisierung der Interaktion mit dem Anwender ausgeschlossen.

Andererseits war offensichtlich, dass statisches HTML als Entwicklungsgrundlage nicht ausreichte: Um auf Nutzereingaben im Sinne eines Dialogs reagieren und bereits gespeicherte Daten auf Anfrage ausgeben zu können, war ein dynamischer Aufbau der anzubietenden HTML-Seiten erforderlich. Darüber

hinaus setzt eine Verarbeitung der HTML-Formulardaten im allgemeinen ohnehin den Einsatz serverseitiger Programmtechnik voraus.

Derzeit stehen im wesentlichen folgende Möglichkeiten für die Datenverarbeitung und den dynamischen Aufbau von Webseiten auf Serverseite zur Verfügung:

- **Common Gateway Interface (CGI)**
  CGI ist eine Schnittstelle des Webservers zu Programmen, die auf dem Serverrechner installiert sind. Übermittelte HTML-Formulardaten werden über die Umgebungsdaten des Webservers eingelesen, an die sogenannten CGI-Programme weitergeleitet und deren Ausgabe durch den Webserver zurückgegeben. Eine weit verbreitete Programmiersprache für CGI-Programme ist die ***P**ractical **E**xtraction and **R**eport **L**anguage (Perl)*.

- **PHP: Hypertext Preprocessor (PHP)**
  PHP ist eine interpretierte Programmiersprache, die zur CGI-Programmierung eingesetzt oder als eigenes Modul in einen Webserver eingebettet werden kann. Im Unterschied zu sonstigen CGI-Sprachen kann PHP direkt in HTML-Quellcode eingebettet werden und bietet einen komfortablen Zugriff auf Formulardaten und Datenbanken.

- **Proprietäre API-Lösungen**
  Hersteller von Webserver integrieren die Funktionalität, die CGI bietet, teilweise in ihre Server und stellen eine eigene Programmierschnittstelle zur Verfügung. Da der Programmaufruf somit durch einen prozessinternen Funktionsaufruf ersetzt werden kann, wird an Performanz gewonnen. Beispiele sind *NSAPI (Netscape Server Application Programming Interface)* von Netscape und *ISAPI (Internet Server Application Programming Interface)* von Microsoft.

- **Java Servlets**
  Bei den Java Servlets handelt es sich ebenfalls um eine API-Lösung, die im Gegensatz zu den vorgenannten unabhängig von der Serverplattform ist. Sie setzen die Installation einer *Java Virtual Machine (JVM)* auf dem Webserver voraus.

- **Server-Side-Includes (SSI)**
  Server-Side-Includes stellen auch Servererweiterungen dar, die jedoch darauf ausgerichtet sind, Dateien, Programme und Systemvariablen in HTML-Dokumente einzubinden. Der Quelltext dieser Dokumente wird nicht nur an den Client weitergegeben, sondern die eingebetteten Steuerbefehle werden zunächst ausgewertet und die Dokumente entsprechend ergänzt. Ein Beispiel für SSI sind die *Microsoft Active Server Pages (ASP)*.

[Loeser 1998], [Münz 2001]

Im Rahmen des Projektes wurde die CGI-Schnittstelle eines Apache 1.3.14 Webservers genutzt. Als Programmiersprache kam Perl in der Version 5.6.1 sowohl für die dynamische Generierung der Webseiten als auch für die Konvertierung der Daten in das PICA/ILTIS-Format zum Einsatz. Bei der Auswahl war maßgeblich, dass plattform- bzw. herstellerspezifische Komponenten vermieden werden sollten und eine entsprechende Konfiguration bereits erfolgreich für die bisherige OSWD eingesetzt wurde. Ferner bietet Perl ein performantes Laufzeitverhalten und eine gute Unterstützung von regulären Ausdrücken, die im Zusammenhang mit der Plausibilitätsprüfung der eingegebenen Daten und der

Datenkonvertierung bedeutsam sind. Der Zugriff auf die Datenbank erfolgt über die Perl-API für Datenbankzugriffe, *DBI*, die als Standardmodul zur Verfügung steht und den datenbankspezifischen Treiber *DBD::Oracle* einbindet. Als Betriebssystem wurde SunOS 5.8 der Firma SUN eingesetzt.

Import der Daten in die IBW

Zum Zweck der Weiterleitung der Schlagwortkandidaten an Die Deutsche Bibliothek waren diese in die IBW zu importieren, wobei der Importvorgang möglichst automatisiert erfolgen sollte.

Die IBW unterstützt *Visual Basic Script (VBScript)* zur Automatisierung von Katalogisierungsprozessen. Bei VBScript handelt es sich um eine von Visual Basic for Applications (VBA) abgeleitete Skriptsprache zur Dynamisierung von Webseiten, die von Microsoft ursprünglich als Antwort auf JavaScript geschaffen wurde. In Zusammenspiel mit dem Windows Scripting Host spielt sie inzwischen auch eine Rolle bei der Programmierung von Batchdateien für Windows.

VBScript ermöglicht die Anlegung einer Objektreferenz auf Microsofts XML-Parser (MSXML), der im Internet Explorer enthalten ist. Die Sammlung von DOM-basierten Komponenten von MSXML umfasst u. a. das XMLHTTP-Objekt, welches einen Zugang zu XML-Dokumenten über das HTTP-Protokoll ermöglicht. Es wird eine einfache API zur Verfügung gestellt, um XML, HTML oder binäre Daten mit Hilfe von GET- oder POST-Requests abzufragen. Diese Eigenschaft konnte genutzt werden, um die Schlagwortkandidaten im Zusammenspiel mit einer webserverbasierten Ausgabe in die IBW zu importieren.

Alternativ hätte durch eine beispielsweise in Visual Basic oder Java zu programmierende Anwendung, die einen geeigneten Treiber einbindet, direkt auf die Datenbank zugegriffen werden können. Damit wären jedoch erhebliche Schwierigkeiten beim Austausch der Daten zwischen dieser Anwendung und der IBW verbunden, da der IBW mit der Zwischenablage und der Möglichkeit des Einlesens von Textdateien nur sehr einfache Austauschinstrumente zur Verfügung stehen. Ausgefeiltere Möglichkeiten wie der Dynamic Data Exchange (DDE) unter Windows werden ebenso wenig unterstützt wie die Office-Programmierung mit VBA.

6.3.3 Softwareergonomische Aspekte

Nach [Balzert 1996] S. 459ff. sind für die softwareergonomische Gestaltung einer Anwendung die folgenden Hauptkriterien maßgeblich:

- **Aufgabenangemessenheit**
  Dieses Kriterium stellt darauf ab, ob der Anwender seine Aufgabe überhaupt mit dem System erreichen kann oder ob er zusätzlich andere Systeme oder Medien einsetzen muss. Ferner sollte die Aufgabenerledigung möglichst effizient und komfortabel erfolgen.

- **Kompetenzförderlichkeit**
  Das Software-System sollte so gestaltet sein, dass der Benutzer kompetent mit ihm umgehen kann und seine Handlungskompetenz durch den Umgang gefördert wird. D. h. der Anwender sollte die Einsatzmöglichkeiten und die Funktionsweise des Systems rasch erlernen und das Erlernte

möglichst weiterverwenden und auf andere Funktionen übertragen können. Dies kann durch Selbsterklärungskomponenten, Hilfesysteme, Funktionstransparenz etc. unterstützt werden.

- **Handlungsflexibilität**
  Dieses Kriterium ist erfüllt, wenn ein Benutzer seine Arbeit auch bei einer geänderten Aufgabenstellung noch effizient mit demselben System erledigen kann und Aufgaben in Abhängigkeit vom Kenntnisstand auf alternativen Wegen ausgeführt werden können.

Neben der grundsätzlichen Beachtung der Aufgabenangemessenheit schien bei der Gestaltung der vorliegenden Anwendung die richtige Gewichtung von Kompetenzförderlichkeit und Handlungsflexibilität eine wesentliche Rolle zu spielen. Hierbei war zwischen den Nutzern aus den Museen und Archiven und den Anwendern der Fachredaktion im BSZ zu unterscheiden. Die erste Gruppe verfügte zu Beginn des Projektes in der Regel über keine Erfahrung mit der SWD und der Online-Erfassung von Normdaten. Daher war der Schwerpunkt der Gestaltung eher auf eine leichte Erlernbarkeit und entsprechende Unterstützung durch das System legen. Die zweite Gruppe ist mit der Normdatenerfassung und dem Zielformat PICA/ILTIS sehr gut vertraut. Bei Schaffung der Redaktionsmöglichkeiten standen somit effiziente und flexible Bearbeitungsmöglichkeiten stärker im Vordergrund.

6.4 Softwaretechnische Realisierung

6.4.1 Architektur

Ausgehend von der in den Datenflussdiagrammen ermittelten Struktur wurde die Anwendung in funktionale Module aufgeteilt. Hauptkomponenten, die auch als eigene CGI-Programme nach außen sichtbar sind, sind dabei die Kandidatenerfassung, das Redaktionstool und die Suche. Ihr Kernstück ist jeweils ein Maklermodul, welches die vom Benutzer getätigten Eingaben einliest, dekodiert und benötigte weitere Module aufruft, die mit Hilfe des Perl-Includes *require* eingebunden sind. Somit ergibt sich beispielsweise für die Erfassung der Kandidaten folgender Aufbau:

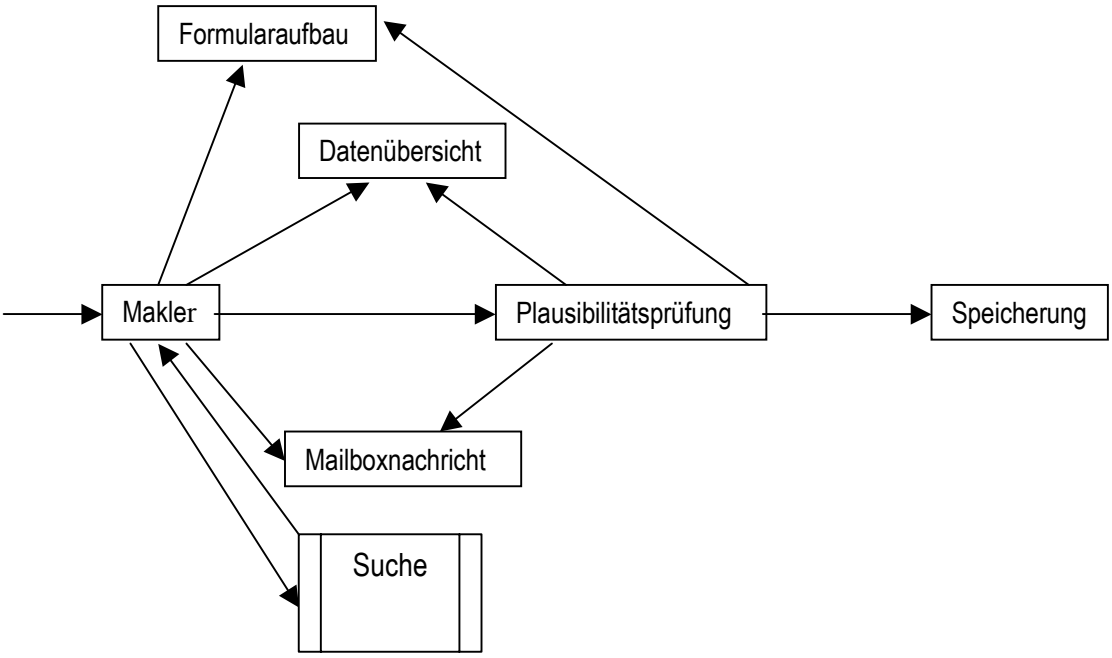


Abbildung 13: Modularisierung der Datenerfassung

Das Redaktionstool wurde teilweise objektorientiert implementiert, da die Geschwindigkeit der Anfragebearbeitung in diesem Bereich weniger entscheidend ist und eine Kapselung der Daten in eigenen Klassen vorteilhaft erschien. Dadurch ist eine übersichtliche und komplette Repräsentation der zu einem Satz gehörenden Daten möglich, unabhängig davon, ob diese im Online-Datenbestand gespeichert oder bei der Exportverarbeitung automatisch hinzugefügt werden. Ferner kann die Datenextraktion exakter von der Konvertierung und Ausgabe getrennt werden. Es wurde jeweils eine Klasse für SWD (einschließlich GKD)- und PND-Sätze geschaffen, von denen das Abfragemodul für jeden abzufragenden Datensatz jeweils eine neue Instanz anlegt. Diese wird bei Initialisierung mit den vorhandenen Defaultwerten belegt und bietet Methoden zum Laden der übrigen Werte aus der Datenbank und zur Rückgabe einer PICA/ILTIS-Repräsentation des Kandidatensatzes an. Letztere wird an das Ausgabemodul übergeben, wobei wahlweise Methoden für eine Browser- oder IBW-taugliche Ausgabe aufgerufen werden können. Insgesamt teilt sich das Redaktionstool in die in Abbildung 14 dargestellten Module auf.

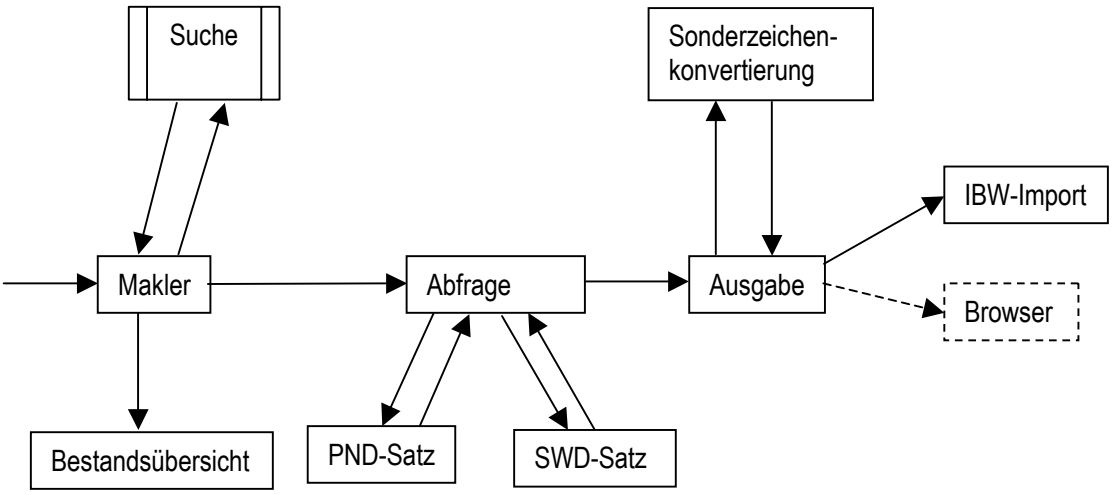


Abbildung 14: Modularisierung des Redaktionstools

Bei der Suche konnten – soweit sie sich auf den vorhandenen OSWD-Bestand bezieht – Komponenten wiederverwendet werden, die beim Aufbau der OSWD entstanden sind.

Zur Trennung der kontextspezifischen Ein- und Ausgaben von dem für die gesamte Anwendung einheitlichen BSZ-Layout wird durchweg mit HTML-Templates gearbeitet: Für die Ausgabe von statischen Teilen wie Seitenleiste, Kopf- und Fußzeile wird ein einheitlicher Code eingebunden; nur der HTML-Quelltext des Ein- und Ausgabebereiches wird in Abhängigkeit vom Kontext dynamisch generiert.

6.4.2 Dateneingabe

Durch die Beschränkung auf das Standardübertragungsprotokoll HTTP sind Benutzereingaben entweder durch Aufruf einer URL, an die ggf. zusätzliche Daten angehängt sind, oder durch Absendung eines HTML-Formulars möglich. Beide Möglichkeiten werden genutzt, wobei die Daten im ersten Fall über eine Umgebungsvariable des Webservers (*GET-Methode*) und im zweiten Fall über die Standardeingabe eingelesen werden (*POST-Methode*).

Auch die zur Dateneingabe angebotenen Formulare werden dynamisch generiert, um Eingaben über mehrere Arbeitsschritte hinweg zu erhalten und eine Anpassung der angezeigten Felder durch den Benutzer zu ermöglichen. Soweit dabei zusätzlich Informationen zur Zustandswahrung gespeichert werden müssen, werden versteckte Formularfelder verwendet.

Die Eingabe neuer Schlagwortkandidaten und Korrekturwünsche durch die Museen und Archive erfolgt mit Hilfe von Formularen, die über eine gemeinsame Startseite zu erreichen sind. Analog zu den bestehenden Normdateien stehen jeweils eigene Formulare für Personen, Körperschaften und den Schlagwort-Hauptteil der SWD (Geographika, Gegenstände und Sachverhalte) zur Verfügung. Ausschnitte aus den Formularen für den Schlagwort-Hauptteil und die Personensätze sind in Anlage 8 und 9 abgebildet.
Für die Datenerfassung obligatorische Felder sind besonders gekennzeichnet; sofern ein Feld wiederholbar ist, können durch Betätigung eines nebenstehenden Buttons die erforderlichen zusätzlichen

Eingabefelder aufgeklappt werden. Bei Verweisungen können die Verweisungsziele durch Wechsel in den Suchmodus und Auswahl des entsprechenden Schlagwortes eingefügt werden. Zusätzlich steht auf Knopfdruck eine Gesamtansicht der eingegebenen Daten zur Verfügung.

In Abhängigkeit davon, ob die Datenerfassung abgeschlossen ist, können die eingegebenen Daten nur zwischengespeichert oder aber zur Redaktion und weiteren Bearbeitung an die Online-Datenbank abgesandt werden. Im ersten Fall erfolgt keine Plausibilitäts- und Vollständigkeitsprüfung; im letzteren werden vorhandene Fehler angezeigt und müssen korrigiert werden, bevor eine endgültige Absendung möglich ist. Dabei wird das Ausfüllen von Pflichtfeldern, die inhaltliche Plausibilität einzelner Felder und die Konsistenz zwischen verschiedenen Feldern eines Datensatzes geprüft. Ferner erfolgt eine Dublettenprüfung auf bereits in der SWD oder dem Kandidatenbestand vorhandene Schlagwortsätze mit identischer Ansetzungsform.

Zwischengespeicherte Datensätze können jederzeit geladen, verändert und erneut gespeichert werden. Nach Abschluss der Erfassung ist dies ebenfalls noch möglich, falls noch keine redaktionelle Bearbeitung erfolgt ist.

6.4.3 Suche

Das Suchmodul erlaubt die Suche auf dem Datenbestand der OSWD und dem Bestand an Schlagwortkandidaten, die bereits von Museen und Archiven eingebracht wurden.

Im Rahmen der SWD-Erweiterung ist die Recherche in der OSWD insofern von Bedeutung, als vor Neuerfassung eines Kandidaten geprüft werden kann, ob die SWD bereits einen geeignete Entsprechung enthält. Ferner können mögliche Verweisungen ausgewählt und bestehende Hierarchie- und Verweisungsstrukturen eingesehen werden. Im Laufe des Projektes stellte sich außerdem heraus, dass der Vergleich mit qualitätssichernden Angaben in bereits vorhandenen SWD-Sätzen die entsprechenden Eintragungen bei der Erfassung von Kandidaten erleichtert. Die OSWD-Ausgabe für Museen und Archive wurde daher gegenüber der bei der Online-Verschlagwortung üblichen Sicht um die Felder Quelle, Benutzungshinweise und redaktionelle Bemerkungen erweitert. Ferner können zu einem angezeigten SWD-Satz Korrekturwünsche eingegeben werden, die analog zu den Kandidaten als neue Arbeitsaufgabe für die Redaktion abgespeichert werden.

Die Suche auf dem Kandidatenbestand ermöglicht die Prüfung, welche Kandidaten bereits eingebracht wurden und in welchem Bearbeitungsstand sie sich befinden. Außerdem können noch nicht redigierte Kandidaten geändert oder gelöscht werden.

Für die Suche in der OSWD und den Kandidaten stehen jeweils eine exakte Suche, eine Teilwortsuche mit Möglichkeit der Links- und Rechtstrunkierung und die Anzeige aller Datensätze in Form einer alphabetischen Liste, ausgehend vom Suchbegriff, zur Verfügung. Bei den Kandidaten ist darüber hinaus eine Suche anhand des Erstellers eines Datensatzes möglich. Falls mehrere Treffer vorhanden sind, werden aus Gründen der Übersichtlichkeit zunächst nur die Bezeichnungen der Treffer in Listenform ausgegeben. Durch Anklicken eines Buttons neben der betreffenden Bezeichnung kann auf die

Vollanzeige des Datensatzes gesprungen werden. Die Anlagen 10 und 11 zeigen die Eingabemaske für die Schlagwortsuche und die Vollanzeige eines Schlagwortkandidaten.

6.4.4 Mailboxsystem

Zur Gewährleistung der Kommunikation zwischen den Teilnehmern der SWD-Erweiterung bzw. den Teilnehmern und der Redaktion wurde ein Mailboxsystem aufgebaut, welches in Anlehnung an die Verbundsysteme und die Kommunikation in PICA/ILTIS datenbankgestützt ist. Die Nachrichten werden ebenfalls mit Hilfe von Webformularen eingegeben und in Tabellen der Datenbank gespeichert. Hierbei ist eine Nachricht jeweils mittels Fremdschlüsselbeziehung mit einem Datensatz verknüpft. Bei Anzeige eines Datensatzes werden auch die zu ihm verfassten Mailboxnachrichten angezeigt.

Diese Konzeption unterstützt forenartige Diskussionen, wie sie im Rahmen der Entwicklung eines Vokabulars häufig erforderlich sind: Jeder Teilnehmer sieht alle Beiträge, die zu einem Begriff abgegeben wurden und kann selbst Kommentare beisteuern, die sofort für alle anderen Nutzer sichtbar sind. Ferner ist die Datensatzorientierung insofern erwünscht, als in der Regel nur Kommunikation stattfinden soll, die sich auf die Entwicklung eines bestimmten Schlagwortsatzes oder mit einem Satz verbundene Probleme bezieht. Außerdem ist das Konzept und seine Funktionalitäten zahlreichen Teilnehmern aus vergleichbaren bibliothekarischen Systemen bekannt, die teilweise aus historischen Gründen in dieser Weise aufgebaut sind.

Obwohl die Nachrichten von allen Teilnehmern einsehbar sind, werden sie dennoch mit Absender- und Empfängerkennungen versehen, um auch eine zielgerichtete Kommunikation zu ermöglichen, wie sie beispielsweise bei der Korrektur von Erfassungsfehlern oder der Bearbeitung von Änderungswünschen hilfreich ist.

Mailboxnachrichten können im Rahmen der vorliegenden Anwendung entweder gemeinsam mit dem betreffenden Begriff neu eingebracht oder an einen vorhandenen Datensatz angehängt werden. Ferner kann ein Nutzer gezielt nach Datensätzen suchen, die an ihn gerichtete Nachrichten enthalten, und erledigte Nachrichten löschen.

6.4.5 Redaktionstool

Die Redaktion setzt zunächst eine Übersicht über den Zugang der Schlagwortkandidaten voraus. Hierzu wurde ein Redaktionstool geschaffen, welches neben den Such- und Anzeigemöglichkeiten der gewöhnlichen Schlagworterfassung einen tabellarischen Überblick über die neu eingegangenen Kandidaten, bereits redigierte Datensätze, Korrekturwünsche und Zwischenspeicherungen gibt. Ferner kann eine chronologische Sortierung der Kandidatenmenge ab einem vorzugebenden Eingangsdatum angezeigt werden. Neben der tabellarischen Anzeige eines Datensatzes ist außerdem die Konvertierung und Anzeige einzelner oder mehrerer ausgewählter Sätze im PICA/ILTIS-Format möglich.

Der Redaktionsvorgang selbst erfolgt auf der Arbeitsoberfläche der IBW. Diese sieht vor, dass der gesamte einzugebende Normdatensatz in einem editierbaren Textfeld angezeigt wird, wobei die einzelnen

Felder durch führende Feldkennungen unterschieden werden. Die Nutzung der IBW für Redaktionsarbeiten bringt den Vorteil, dass der Redakteur ein aus der Normdatenerfassung vertrautes Arbeitsinstrument einsetzen kann. Außerdem befinden sich die Daten bei Redaktion bereits im Zielformat und können mit Hilfe des Texteditors rasch und flexibel bearbeitet werden. Eine Redaktion auf dem Online-Datenbestand hätte im übrigen zu redundanten Arbeitsschritten geführt, da die IBW ohnehin die satzweise Eingabe und Absendung der Daten erfordert.

Durch Betätigung von in die IBW eingefügten Buttons können die noch nicht redigierten Datensätze in das Textfeld geladen werden (vgl. Anlage 12). Hierzu werden die Daten mittels des in VBScript referenzierten MSXML-Parsers abgerufen. Das ausgebende CGI-Skript konvertiert die Schlagwortsätze dabei on-the-fly in das PICA/ILTIS-Format. Ist ein Datensatz korrekt bzw. die Redaktion abgeschlossen, sendet der Redakteur ihn innerhalb von PICA/ILTIS ab und ruft den nächsten Kandidaten auf. Durch Auswahl des entsprechenden Abrufbuttons wird zugleich an den Kandidatenbestand zurückgemeldet, ob die Redaktion erfolgreich war. Das dabei erfolgende Setzen bzw. Nichtsetzen eines Redaktionsflags kann im Falle eines Fehlers oder nachträglichen Problems durch eine entsprechende Funktion des Redaktionstools rückgängig gemacht werden.

6.4.6 Berücksichtigung der Usability-Ziele

Aufgabenangemessenheit

Die Erfassungs- und Redaktionsaufgaben können mit Hilfe der angebotenen Funktionalitäten erledigt werden. Soweit möglich erfolgt dies effizient und komfortabel. Bei den Erfassungsformularen wurde der Kompetenzförderlichkeit teilweise Vorrang gegenüber der Effizienz gegeben. So ist es einem Experten sicherlich möglich, Daten mit Hilfe von Feldkennungen und Steuerzeichen schneller einzugeben als dies die fest strukturierten HTML-Formulare erlauben. Aus den in Kapitel 6.3.4 erwähnten Gründen wurde der leichten Erlernbarkeit und Bedienbarkeit der Eingabeschnittstelle jedoch Vorrang eingeräumt.

Medienbrüche und Sprünge zwischen Anwendungen wurden weitestgehend vermieden, indem die OSWD in die Erfassung integriert wurde und auch weitere Werkzeuge wie Ländercodes oder Systematikstellen entweder in Form von Menüs zur Verfügung stehen oder online zugänglich sind. Auch der automatisierte Import in die IBW folgt diesem Kriterium: Die Anwendung von „Copy&Paste“, die bei der sonstigen Übertragung von Normdaten noch häufig üblich ist, konnte dadurch vermieden werden.

Kompetenzförderlichkeit

Der Kompetenzförderlichkeit wurde v. a. durch folgende Maßnahmen Rechnung getragen:

- Ausschließliche Verwendung bekannter und standardisierter Gestaltungselemente für Webseiten
- Durch einheitliche Gestaltung, insbesondere bei den Eingabeformularen, wird der Benutzer schneller mit der Funktionalität vertraut.
- Betonung der Übersichtlichkeit durch Funktionen wie tabellarische Gesamtansicht und optionales Aufklappen von Zusatzfeldern
- Schaffung einer Online-Hilfe, die Inhalt und Eingabeformat der einzelnen Erfassungsfelder beschreibt.

- Prüfung auf Fehler und Inkonsistenzen und unmittelbare Rückmeldung vor Abschluss der Datenerfassung.
- Aktionen, die nicht direkt mit der Aufgabenerledigung zu tun haben, werden automatisch durchgeführt, wie z. B. Dublettenkontrolle, Umsetzung nach PICA/ILTIS, Ergänzung um Standardangaben.
- Verstärkter Einsatz von steuernden und erklärenden Dialogfenstern.

Handlungsflexibilität

Handlungsflexibilität konnte nur begrenzt hergestellt werden, da der Übersichtlichkeit und Fehlerrobustheit Priorität eingeräumt wurde. Sie spielte jedoch bei folgenden Gestaltungsmaßnahmen eine Rolle:

- Weitgehende Flexibilität in bezug auf die nachträgliche Änderung und Löschung von Kandidatensätzen bis zum Zeitpunkt der Redaktion
- Durch Zwischenspeicherung im Online-Bestand kann die Schlagworterfassung unterbrochen und zu einem späteren Zeitpunkt fortgesetzt werden.
- Bei den Feldern Ländercode und Empfänger ist alternativ zur Auswahl aus Menüs auch die direkte Eingabe bekannter Codes möglich.
- Verschiedene Möglichkeiten der Suche im Kandidatenbestand

7. Pilotversuch

7.1 Ablauf und Beteiligte

Das Vorhaben, Museen und Archive des Landes Baden-Württemberg aktiv an der SWD-Kooperation teilnehmen zu lassen, geht auf eine Anregung des BSZ zurück. In Absprache mit Der Deutschen Bibliothek wurde vereinbart, dass die Realisierbarkeit der Teilnahme zunächst im Rahmen eines sechsmonatigen Pilotprojektes getestet werden soll.

Nach einer ersten Spezifikation der Anforderungen und Produktdaten in Gesprächen mit den Projektbetreuern im BSZ wurde die Architektur der Anwendung konzipiert und ein Teil der Nutzerschnittstelle prototypisch realisiert. Dieser Prototyp wurde bei einem gemeinsamen Treffen mit den künftigen Anwendern und einer Vertreterin Der Deutschen Bibliothek vorgestellt. Als Ergebnis der dabei stattfindenden Diskussion konnten Korrekturen und Feinabstimmungen bei den zu erhebenden Daten und der Gestaltung der Nutzerschnittstelle vorgenommen werden.

Nachdem die Eingabe- und Speicherungskomponenten fertiggestellt waren, konnte der Pilotversuch am 01.01.02 wie geplant beginnen. An der Konvertierung in das PICA/ILTIS-Format und dem Import in die IBW wurde in der Folgezeit noch gearbeitet. Die Nutzer erhielten Ende Januar eine Schulung durch Die Deutsche Bibliothek zur Normdatenerfassung und der SWD.

Neben Der Deutschen Bibliothek und dem BSZ nehmen derzeit folgende Institutionen an dem Test teil:

- Badisches Landesmuseum Karlsruhe
- Landesmuseum für Technik und Arbeit Mannheim
- Reiss-Engelhorn-Museen Mannheim
- Landesarchivdirektion Stuttgart

7.2 Nutzung und Probleme

Bislang wurden 98 Kandidaten eingebracht, wobei es sich überwiegend um Schlagwörter handelt, die sich auf Geographika, Gegenstände und Sachverhalte beziehen. 78 davon konnten im Rahmen der im Wochenrhythmus stattfindenden Redaktion bearbeitet und nach PICA/ILTIS eingebracht werden. Alle Schlagwortsätze wurden in der Folge von Der Deutschen Bibliothek autorisiert und standen in der Regel binnen 2 Wochen nach Einbringung im Rahmen der wöchentlichen Datenlieferung als SWD-Satz zur Verfügung.

Bekannte Probleme der Nutzung bezogen sich auf kleinere Schwierigkeiten mit der Handhabung der Formulare und der Funktionalität des Mailboxsystems. Inhaltliche Schwierigkeiten ergaben sich zum einen dadurch, dass die in der SWD übliche Erfassung von Lebensdaten für Personen für die Zwecke der Museen unzureichend war und Missverständnisse zur Bedeutung des Datenfeldes Titel vorlagen. Diesen Problemen konnte durch Einfügung eines neuen Feldes bzw. Anpassung des Formulars begegnet werden.

Grundsätzlichere Schwierigkeiten bereiten nach Angaben der Anwender die oft unzureichend ausgeprägten Hierarchien in der SWD. So fehlen beispielsweise beim Schlagwort „Schach“, die für Museen wesentlichen übergeordneten Hierarchieebenen „Brett-/Gesellschaftsspiel“ und „Spiel“. Als untergeordnete Begriffe von „Schachfigur“ finden sich „Leichtfigur“ und „Schwerfigur“ inkonsistenterweise auf gleicher Hierarchieebene wie „Bauer <Schach>“ und König <Schach>“. Verbesserungen in der Hierarchiestruktur können nicht immer durch Einbringen eines neuen Begriffes oder einen einfachen Änderungswunsch erreicht werden, sondern erfordern teilweise größere Umstrukturierungen. Ob und wie diese erreicht werden können, wird zur Zeit zwischen den Projektteilnehmern und Der Deutschen Bibliothek diskutiert.

Ein weiteres häufiges Problem auf begrifflich-inhaltlicher Ebene stellt die Tatsache dar, dass manche Begriffe zwar vorhanden sind, ihre Bezeichnung für die neuen SWD-Anwender jedoch unzureichend ist. So bietet die SWD beispielsweise als Ansetzung „Iran“ und „Iran <Altertum>“ und die für Museen wichtige Bezeichnung „Persien“ stellt ein Synonym dar, welches auf diese Bezeichnungen verweist und somit nicht zur Ressourcenbeschreibung verwendet werden kann. Auch in dieser Hinsicht besteht voraussichtlich über die Einbringung von Korrekturwünschen hinausgehender Diskussionsbedarf auf redaktioneller, nichttechnischer Ebene.

Aus der Sicht der Redaktion bewegten sich die Probleme neben der geschilderten Schwierigkeiten mit der Hierarchie v. a. im Bereich der Einhaltung des Regelwerks, z. B. bei Angabe von Quellen, und der Beachtung von Formatvorgaben für einzelne Felder.

7.3 Bewertung

Trotz der relativ geringen Zahl an eingebrachten Kandidaten, wird das Pilotprojekt von allen Teilnehmerseiten positiv bewertet (vgl. auch [Ludewig 2002]). Es ist geplant, eine dauerhafte Teilnahme der Museen an der SWD-Kooperation zu etablieren. Hierbei soll die Redaktion – in Abweichung von dem Pilotversuch – von den Museen selbst organisiert werden. Die im Rahmen des Projektes realisierte Anwendung wird von den Benutzern als geeignetes Instrument für die SWD-Erweiterung angesehen. Soweit absehbar wird sich die dauerhafte Partizipation des entwickelten Erfassungs- und Übergabesystems bedienen.

Die Vorteile der Anwendung scheinen zum einen darin zu liegen, dass gängige Komponenten Sprachen und Protokolle des WWW für die Erweiterung der SWD-Kooperation eingesetzt werden. So sind die Nutzer mit dem Umgang mit Browsern, mit den gängigen Gestaltungselementen von HTML und dem Ablauf von Datenübertragungen im Rahmen des HTTP-Protokolls vertraut. Ferner haben sich die Entwicklungskomponenten in der Regel durch langjährigen Einsatz im WWW bewährt und stehen als offene Standards zur Verfügung. Eine Entwicklung einer proprietären Lösung hätte sicherlich den Zeit- und Kostenrahmen des Projektes gesprengt. Auch in bezug auf das eingesetzte relationale Datenbanksystem steht z. B. in Form von MySQL ein alternatives Open Source Produkt zur Verfügung, welches zumindest für die Speicherung und Suche auf den Kandidaten ausreichende Leistungen bieten dürfte. Ähnliche Realisierungen für die Vokabularentwicklung anderer Beschreibungscommunities des Semantic Web scheinen daher möglich.

Ein weiterer wesentlicher Aspekt liegt meines Erachtens darin, dass die Nutzerschnittstelle bewusst einfach gehalten und versucht wurde, den Anwender durch Hilfekomponenten und selbstbeschreibende Gestaltung weitestgehend zu unterstützen. Kontrollierte Vokabulare, die die Qualität der Beschreibung tatsächlich verbessern und eine effiziente Nutzung ermöglichen sollen, bringen wohl zwangsläufig komplexe Regeln und Formatvorgaben für die begriffliche Fortschreibung und Strukturpflege mit sich. Diese Komplexität kann jedoch nicht einfach an einen neuen Nutzer weitergegeben werden, der in der Regel nur begrenzt Zeit und Möglichkeiten hat, sich mit den Beschreibungskonventionen auseinander zu setzen. Die Einfachheit und die Nutzungsunterstützung einer Anwendung für die Vokabularentwicklung stellen daher ein wichtiges Moment für deren Akzeptanz dar.

Andererseits ergeben sich natürlich auch Einschränkungen durch eine einfache Nutzerschnittstelle und den Einsatz von Standard-Webtechnologien. So ist die Effektivität und insbesondere die Geschwindigkeit der Interaktion geringer als bei einem Expertensystem. Vorliegend ergab sich durch die Eingabe mit HTML-Formularen auch eine Limitierung des Zeichensatzes: Der im WWW übliche Zeichensatz ISO-8559-1 schränkt die Eingabe von Sonderzeichen und Diakritika deutlich gegenüber den Möglichkeiten ein, die PICA/ILTIS bzw. der für die SWD festgelegte Zeichensatz bieten würden. Dies wurde allerdings bereits im Rahmen der Anforderungsanalyse erkannt und zugunsten der geringeren Komplexität bewusst in Kauf genommen.

In bezug auf die Skalierbarkeit scheint die gewählte Realisierung ausreichend Spielraum zu bieten. Sowohl die WWW-Schnittstelle als auch die Datenbank würde ein Vielfaches der im Projekt erfolgten Nutzung erlauben. Hinsichtlich der Erweiterbarkeit wurde ausgenutzt, dass mit PICA/ILTIS auf einem sehr standardisierten und stabilen Zielformat aufgebaut werden kann. Bei häufigen Änderungen wäre die Speicherung und Ausgabe basierend auf XML zu erwägen, um flexible Änderungen der Datenstruktur zu ermöglichen. Auch bei der Modularität wäre in größeren und auf Wiederverwendung von Teilkomponenten ausgerichteten Systemen ein deutlicherer Akzent zu setzen.

Insgesamt scheint die Einbeziehung eines neuen Nutzerkreises in die Entwicklung und Pflege eines Vokabulars mit Hilfe einer Anwendung, wie sie vorliegend implementiert wurde, möglich zu sein. Das steigende Erfordernis einer einheitlichen, domänenübergreifenden Ressourcenbeschreibung, die mit der semantischen Nutzung des WWW einhergeht, lässt erwarten, dass der Einsatz und die Bedeutung entsprechender Instrumente in Zukunft zunehmen wird.

8. Zusammenfassung

Der Metadatenerschließung im WWW kommt mit seinem weiteren Wachstum und seiner zunehmenden Nutzung als Medium zur Publikation und Suche von Informationen, für den Datenaustausch und für elektronische Dienstleistungen eine steigende Bedeutung zu. Es existiert eine Reihe von Ansätzen, die auf eine stärkere Berücksichtigung der Semantik von Webressourcen für die vorgenannten Zwecke abzielen. Ihnen ist gemeinsam, dass sie die vorherige Erschließung der Ressourcen mit Metadaten voraussetzen.

Die vorgeschlagenen Technologien decken dabei eine Reihe von semantischen Auszeichnungsstandards, deklarativen Sprachen und Architekturvorschlägen ab. Sie reichen von Dublin Core als gemeinsamer Kernmenge von inhaltsorientierten Beschreibungselementen für eine Vielzahl von Ressourcen in unterschiedlichen Domänen bis zu Ontologiesprachen, die darauf abzielen, die Semantik von Ressourcen formal und rechnerinterpretierbar darzustellen. XML liefert dabei vor allem die Syntax, um Daten flexibel mit anwendungsspezifischen Beschreibungselementen auszuzeichnen, zu speichern, zu transformieren und über Domänengrenzen hinweg auszutauschen. RDF und RDFS stellen eine Grundlage für eine syntaxunabhängige Darstellung der Semantik von domänenspezifischen Metadatenmodellen zur Verfügung und bieten eine Infrastruktur zur Nutzung der verteilten Metadatenschemen an.

Den umfassendsten Ansatz für eine semantische Nutzung stellt der „Semantic-Web“-Vorschlag von Tim Berners-Lee und dem W3C dar. Er vereint eine Vielzahl der vorgestellten Technologien mit weiteren Mechanismen, die Schlussfolgerungen, Beweisführungen und Authenzitätsnachweise auf den im WWW verfügbaren Daten zum Ziel haben. Insgesamt ergibt sich ein Schichtenmodell, welches eine deklarative Darstellung und rechnerbasierte Interpretation der Semantik von Informationen auf unterschiedlichen Niveaus ermöglichen soll. Dadurch sollen sich effektivere Möglichkeiten der Informationssuche, der Integration und der Kombination von Wissen und neuartige Webdienste eröffnen.

Während ein Teil des „Semantic-Web“-Konzeptes noch eine Zukunftsvision darstellt, sind Technologien von Dublin Core bis RDF bereits im praktischen Einsatz und erlauben eine inhaltsorientierte Beschreibung und Nutzung von Webressourcen, die über Volltextsuche und nutzergesteuerte Navigation hinausgehen.

Es wurde dargestellt, dass kontrollierten Vokabularen eine entscheidende Rolle bei der Beschränkung und Strukturierung der für die Beschreibung verwendeten Sprachelemente zukommt. Eine besondere Herausforderung stellt in diesem Zusammenhang die Entwicklung der Vokabulare in einer Weise dar, dass sie hohen Anforderungen an Konsistenz und Struktur genügen und zugleich zur Beschreibung von Ressourcen in verschiedensten Domänen geeignet sind. Dies kann nur gelingen, wenn die Fachgemeinschaften der zu integrierenden Wissens- und Anwendungsbereiche in der Lage sind, aktiv an der Entwicklung und Pflege des Vokabulars teilzunehmen.

Am Beispiel der OSWD wurde aufgezeigt, wie bei einem zur Metadatenerschließung im WWW genutzten Thesaurus die Partizipation des neuen Nutzerkreis der Museen und Archive ermöglicht werden kann. Im Rahmen eines Pilotprojektes habe ich WWW- und Datenbanktechnologien zur Implementation einer Anwendung eingesetzt, die das Einbringen neuer Begriffe und das Anmelden von Korrekturwünschen unter Beachtung der bestehenden qualitätssichernden Maßnahmen ermöglicht. Außerdem gibt die Anwendung die Möglichkeit zur Kommunikation innerhalb des Nutzerkreises, was für die für die Entwicklung eines Vokabulars essentiell ist.

Die im Pilotversuch gemachten Erfahrungen weisen darauf hin, dass eine derartige Erweiterung erfolgreich ist und die SWD sich durch die geplante Fortführung der aktiven Partizipation der Museen und Archive in Richtung eines universalen Vokabulars zur Ressourcenbeschreibung entwickeln kann. Diesem käme im Rahmen der Bemühungen zur semantischen Nutzung des WWW vermutlich eine enorme Bedeutung zu, da es zu den vorgestellten Technologien die adäquaten begrifflichen Repräsentationseinheiten für eine domänenneutrale und allgemein akzeptierte Auszeichnung liefert.

Weiterhin wurde die Erfahrung gemacht, dass sich auch für den Bereich der Vokabularentwicklung das allgemeine Prinzip der Ressourcenbeschreibung empfiehlt, auf im WWW gängige Sprachen und Anwendungen zurückzugreifen und einfach zu handhabende Werkzeuge zur Verfügung zu stellen. Dies steht in einem gewissen Spannungsverhältnis zur Tatsache, dass die Pflege eines hochentwickelten Vokabulars grundsätzliche Fachkenntnisse und nicht unerheblichen Arbeitsaufwand voraussetzt. Durch Integration von Hilfsmitteln, geeignete Oberflächengestaltung und Automatisierung kann dem jedoch teilweise begegnet werden, so dass rationelles Arbeiten möglich ist.

9. Anlagen

Anlage 1: Zusammenstellung relevanter Felder des PICA/ILTIS-Datenformats

1. SWD- Sätze (einschließlich Körperschaftssätze)

	Feld-kennung	Intern-bezeichnung			
	005	002@	$0	**Gattung und Status**	
				Pos. 1:	
				T	Thesaurus
				Pos. 2:	
				s	Schlagwortnormdatei (SWD)
				Pos. 3:	
				v	vorläufige Ansetzung; uncontrolled term
	011	008A		**Teilbestands- und Verwendungskennzeichen**	
*		008A	$a	"/..."	Teilbestandskennzeichen
				s	Sacherschließung
				t	vorläufige Ansetzung
*	**028**	007R	$0	**Angabe einer GKD-Nummer**	
	800	041A		**Hauptschlagwort**	
		041A	$p	"p|"	Personenschlagwort
		041A	$s	"s|"	Sachschlagwort
		041A	$g	"g|"	Geographikum
		041A	$c	"c|"	ortsgebundene Körperschaft
		041A	$k	"k|"	ortsungebundene Körperschaft
					alle (@{)
	801	041B		**1. Unterschlagwort**	
		041B	$p	"p|"	Personenschlagwort
		041B	$s	"s|"	Sachschlagwort
		041B	$g	"g|"	Geographikum
		041B	$c	"c|"	ortsgebundene Körperschaft
		041B	$k	"k|"	ortsungebundene Körperschaft
		041B	$x	"x|"	Schlagwort mit Indikator "Blank"
					alle (@{)
	802	041K		**2. Unterschlagwort**	
		041K	$x	"x|"	Schlagwort mit Indikator "Blank" (@{)
	803	041L		**3. Unterschlagwort**	
		041L	$x	"x|"	Schlagwort mit Indikator "Blank" (@{)
	804	041M		**4. Unterschlagwort**	
		041M	$x	"x|"	Schlagwort mit Indikator "Blank" (@{)
	805	041N		**5. Unterschlagwort**	
		041N	$x	"x|"	Schlagwort mit Indikator "Blank" (@{)
*	**808**	046A		**Erläuterungen zum Schlagwort**	
		046A	$S	"a|"	Indikator für Quelle
				"b|"	Indikator für Definition
				"c|"	Indikator für Benutzungshinweise
				"d|"	Indikator für redaktionelle Bemerkungen
				"z|"	Indikator für lokale Kennung
		046A	$a	- ohne -	Text gemäß Indikator
	810	042A		**Systematik-Nummer der DDB**	
*		042A	$a	-ohne-	weitere Nummern werden mit ";" angeschlossen
	811	042B		**Ländercode nach DIN EN 23166 (ISO 3166)**	
*		042B	$a	-ohne-	weitere Codes werden mit ";" angeschlossen
		042B	$b	"¬*"	Bemerkungen zum Ländercode
	812	042C		**Sprachencode nach ISO/TC46/SC4-N350**	
*		042C	$a	-ohne-	weitere Codes werden mit ";" angeschlossen
	813	042D		**Zeitcode nach UDK**	

*		042D	$a	-ohne-	Zeitcode; weitere Codes werden mit ";" angeschlossen
		042D	$b	"*"	Zeitcode für den Beginn eines Zeitintervalls
		042D	$c	"/"	Zeitcode für das Ende eines Zeitintervalls
*		042D	$d	":"	Zeitcode (Vergleich zum vorangehenden Zeitcode)
	814	042E		**Zeitcode**	
		042E	$a	-ohne-	Zeitcode
		042E	$b	"¬*"	Bemerkungen zum Zeitcode
	820	041D		**Alternativform zum Hauptschlagwort**	
		041D	$p	"p|"	Personenschlagwort
		041D	$s	"s|"	Sachschlagwort
		041D	$g	"g|"	Geographikum
		041D	$c	"c|"	ortsgebundene Körperschaft
		041D	$k	"k|"	ortsungebundene Körperschaft
					alle (@{)
	821	041E/01		**1. Unterschlagwort zur Alternativform**	
		041E/01	$p	"p|"	Personenschlagwort
		041E/01	$s	"s|"	Sachschlagwort
		041E/01	$g	"g|"	Geographikum
		041E/01	$c	"c|"	ortsgebundene Körperschaft
		041E/01	$k	"k|"	ortsungebundene Körperschaft
		041E/01	$x	"x|"	Schlagwort mit Indikator "Blank"
					alle (@{)
	822	041E/02		**2. Unterschlagwort zur Alternativform**	
		041E/02	$x	"x|"	Schlagwort mit Indikator "Blank" (@{)
	823	041E/03		**3. Unterschlagwort zur Alternativform**	
		041E/03	$x	"x|	Schlagwort mit Indikator "Blank" (@{)
	824	041E/04		**4. Unterschlagwort zur Alternativform**	
		041E/04	$x	"x|"	Schlagwort mit Indikator "Blank" (@{)
	825	041E/05		**5. Unterschlagwort zur Alternativform**	
		041E/05	$x	"x|"	Schlagwort mit Indikator "Blank" (@{)
*	**830**	041F		**Äquivalente Bezeichnung**	
		041F	$p	"p|"	Personenschlagwort
		041F	$s	"s|"	Sachschlagwort
		041F	$g	"g|"	Geographikum
		041F	$c	"c|"	ortsgebundene Körperschaft
		041F	$k	"k|"	ortsungebundene Körperschaft
		041F	$u	"¬/¬"	Weiterer Bestandteil einer Verweisung
					alle (@{)
		041F	$b	"¬*"	Ergänzende Angaben
*	**845**	041S		**Übergeordneter Begriff zu Individualbezeichnung**	
		041S	$p	"p|"	Personenschlagwort
		041S	$s	"s|"	Sachschlagwort
		041S	$g	"g|"	Geographikum
		041S	$c	"c|"	ortsgebundene Körperschaft
		041S	$k	"k|"	ortsungebundene Körperschaft
		041S	$u	"¬/¬"	Unterschlagwort oder weiteres Schlagwort in
					einer Schlagwortkette
					alle (@{)
		041S	$b	"¬*"	Ergänzende Angaben
*	**850**	039C		**Übergeordnetes Schlagwort**	
		039C	$p	"p|"	Personenschlagwort
		039C	$s	"s|"	Sachschlagwort
		039C	$g	"g|"	Geographikum
		039C	$c	"c|"	ortsgebundene Körperschaft
		039C	$k	"k|"	ortsungebundene Körperschaft
		039C	$u	"¬/¬"	Unterschlagwort oder weiteres Schlagwort in
					einer Schlagwortkette
					alle (@{)
		039C	$b	"¬*"	Ergänzende Angaben

*	**860**	039D		**Verwandtes Schlagwort**

		039D	$p	"p|"	Personenschlagwort
		039D	$s	"s|"	Sachschlagwort
		039D	$g	"g|"	Geographikum
		039D	$c	"c|"	ortsgebundene Körperschaft
		039D	$k	"k|"	ortsungebundene Körperschaft
		039D	$u	"¬/¬"	Unterschlagwort oder weiteres Schlagwort in einer Schlagwortkette alle (@{)
		039D	$b	"¬*"	Ergänzende Angaben
*	**870**	039E		**Schlagwort für eine frühere Benennung**	
		039E	$p	"p|"	Personenschlagwort
		039E	$s	"s|"	Sachschlagwort
		039E	$g	"g|"	Geographikum
		039E	$c	"c|"	ortsgebundene Körperschaft
		039E	$k	"k|"	ortsungebundene Körperschaft
		039E	$u	"¬/¬"	Unterschlagwort oder weiteres Schlagwort in einer Schlagwortkette alle (@{)
		039E	$b	"¬*"	Ergänzende Angaben
*	**880**	039F		**Schlagwort für eine spätere Benennung**	
		039F	$p	"p|"	Personenschlagwort
		039F	$s	"s|"	Sachschlagwort
		039F	$g	"g|"	Geographikum
		039F	$c	"c|"	ortsgebundene Körperschaft
		039F	$k	"k|"	ortsungebundene Körperschaft
		039F	$u	"¬/¬"	Unterschlagwort oder weiteres Schlagwort in einer Schlagwortkette alle (@{)
		039F	$b	"¬*"	Ergänzende Angaben
	901	047A/01		**Mail-Box-Nachrichten**	
		047A/01	$b	"...*"	Formatierter Text
		047A/01	$a	- ohne -	Freitext

Legende:

- Leerzeichen/Blank/Spatium wird durch das Zeichen "¬" dargestellt
- *. Feld/Unterfeld ist wiederholbar
- (@{) bzw. ({) Der Inhalt des betreffenden Unterfeldes ist sortierrelevant und enthält Steuerzeichen zur Kennzeichnung nicht-sortierender Bestandteile.

2. PND-Sätze

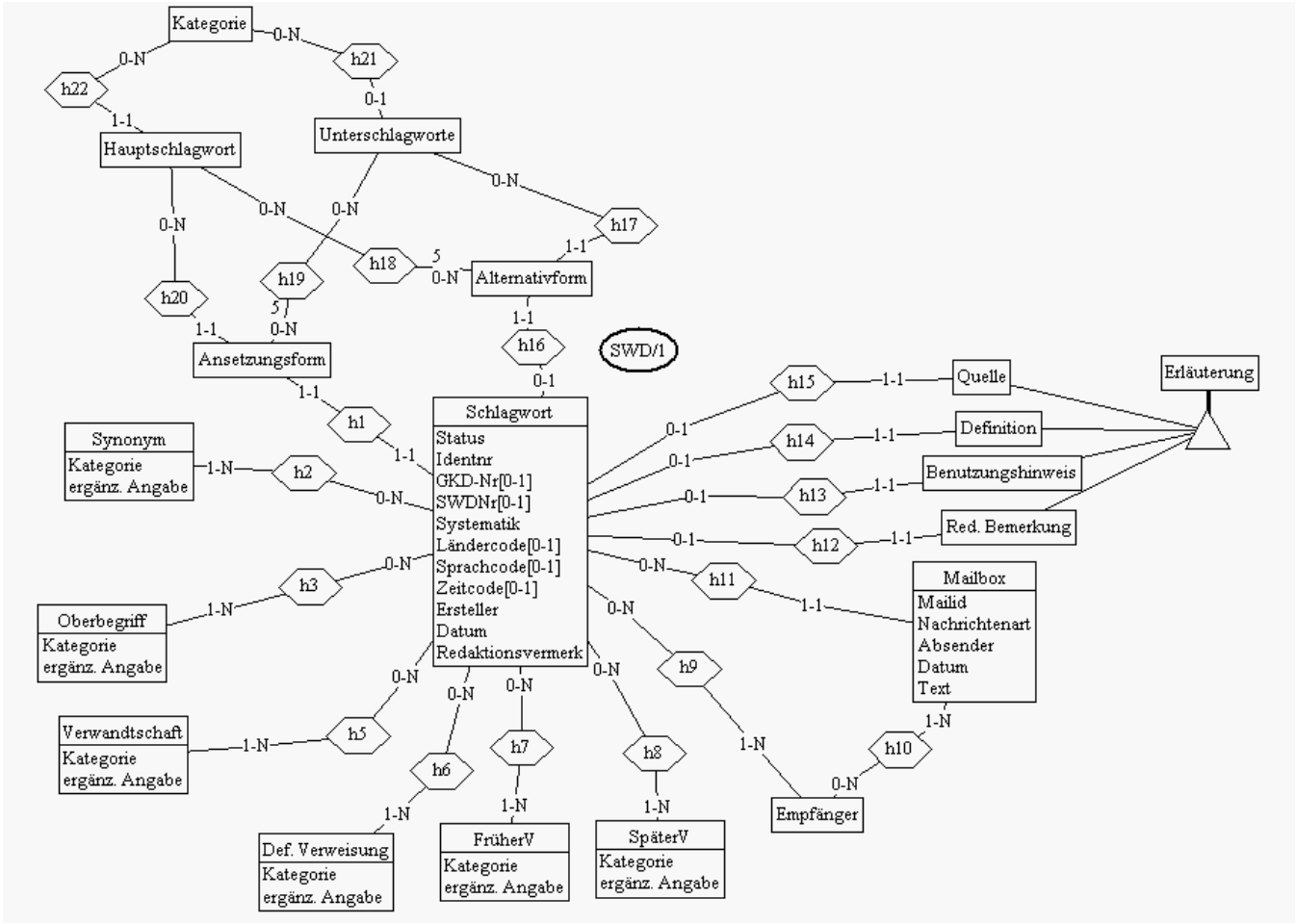
	Feld-kennung	Intern-bezeichnung		Kurzbeschreibung	
	005	002@	$0	**Satzart**	
				Pos. 1:	
				T	Thesaurus
				Pos. 2:	
				n	Name (Nichtindividualisierter Satz)
				p	Person (Individualisierter Satz)
	011	008A		**Teilbestands- und Verwendungskennzeichen**	
*		008A	$a	"/..."	Teilbestandskennzeichen
				a	Namen der Formalerschließung 1500 - 1850, Namen des Göttinger Altbestandserschließungsprojektes bis 1945
				e	Namen aus dem osteuropäischen und islamischen Kulturkreis, zu transliterierende Namen
				f	Formalerschließung (für Titelverknüpfung)
				m	Musikschaffende
				n	PMA, bereits in Druckausgabe erschienen
				o	PMA, noch nicht in Druckausgabe erschienen
				p	PAN
				s	Sacherschließung
				t	vorläufige Ansetzung
				z	ZKA
	100	028A		**Einheitliche Ansetzungsform oder Ansetzungsform nach RAK**	
				Die Dateneingabe und -präsentation erfolgt in invertierter Form:	
		028A	$5	"@..."	Persönlicher Name
		028A	$a	- ohne -	Familienname
		028A	$d	",¬"	Vorname(n)
		028A	$c	"¬/"	Präfix, dem Vornamen nachgestellt
		028A	$l	"¬<...>"	Ordnungshilfe
		028A	$f	"¬(...)"	RSWK-Homonymenzusatz, Ergänzung hinter dem Namen (nur in Satzart Tp)
*	**101**	046B		**Bemerkungen zur Ansetzungsform**	
		046B	$S	"|a|"	Indikator für Quelle
				"|c|"	Indikator für Benutzungshinweise
				"|d|"	Indikator für redaktionelle Bemerkungen
				"|e|"	Indikator für ergebnislos eingesehene Quelle
				"|z|"	Indikator für lokale Kennung
		046B	$a -	ohne - Text gemäß Indikator	
	120	028B		**Ansetzungsform nach RSWK**	
				Feldinterne Steuerzeichen wie 100	
	140	028C		**Ansetzungsform der LoC**	
				Feldinterne Steuerzeichen wie 100	
*	**145**	028D		**Alternative Ansetzungsform**	
		028D	$S	"|b|"	Alternative Ansetzungsform nach RAK-ÖB
				Weitere feldinterne Steuerzeichen wie 100, zusätzlich:	
		028D	$p	"¬%"	sonstige Angaben, weitere identifizierende Angaben, Bemerkungen
*	**200**	028@		**Verweisungsformen für alle Namensverweisungen**	
				Feldinterne Steuerzeichen wie 100, zusätzlich:	
		028@	$B	"¬[...]"	Erläuterung
		028@	$p	"¬%"	sonstige Angaben, weitere identifizierende Angaben, Bemerkungen
*	**300**	032A		**Lebens- und Wirkungsdaten der Person in normierter Form**	
		032A	$S	"|a|"	Lebensdaten
		032A	$a	- ohne -	Angabe von Jahreszahlen gemäß Indikator
		032A	$p	"¬%"	Bemerkungen
*	**305**	032E		**Lebens- und Wirkungsdaten der Person in nichtnormierter Form**	

		032E	$S	"|a|"	Lebensdaten
		032E	$a	- ohne -	Angabe von Jahreszahlen gemäß Indikator
		032E	$p	"¬%"	Bemerkungen
*	**310**	032B		**Geschlecht, Beruf und/oder Funktion der Person; sonstige identifizierende Angaben zur Person**	
				"|j|"	nichtnormierte Angabe Beruf und/oder Funktion
				"|u|"	nichtnormierte sonstige identifizierende Angaben
				"|w|"	Geschlechtsangabe 'm' bzw. 'f'
		032B	$a	- ohne -	Text gemäß Indikator
	810	042A		**Systematiknummer der DDB**	
*		042A	$a	- ohne -	weitere Codes werden mit ";" angeschlossen
	811	042B		**Ländercode nach DIN EN 23166 (ISO 3166)**	
*		042B	$a	- ohne -	weitere Codes werden mit ";" angeschlossen
		042B	$b	"¬*" Bemerkungen zum Ländercode	
*	**899**	046G		**Titelangaben** (nur aus Altdatenkonversion)	
		046G	$a	- ohne -	Titel werden nur deskriptiv erfaßt, es erfolgt keine Unterteilung in Unterfelder
	901	047A/01		**Mail-Box-Nachrichten**	
		047A/01	$b	"...*"	Formatierter Text
		047A/01	$a	- ohne -	Freitext

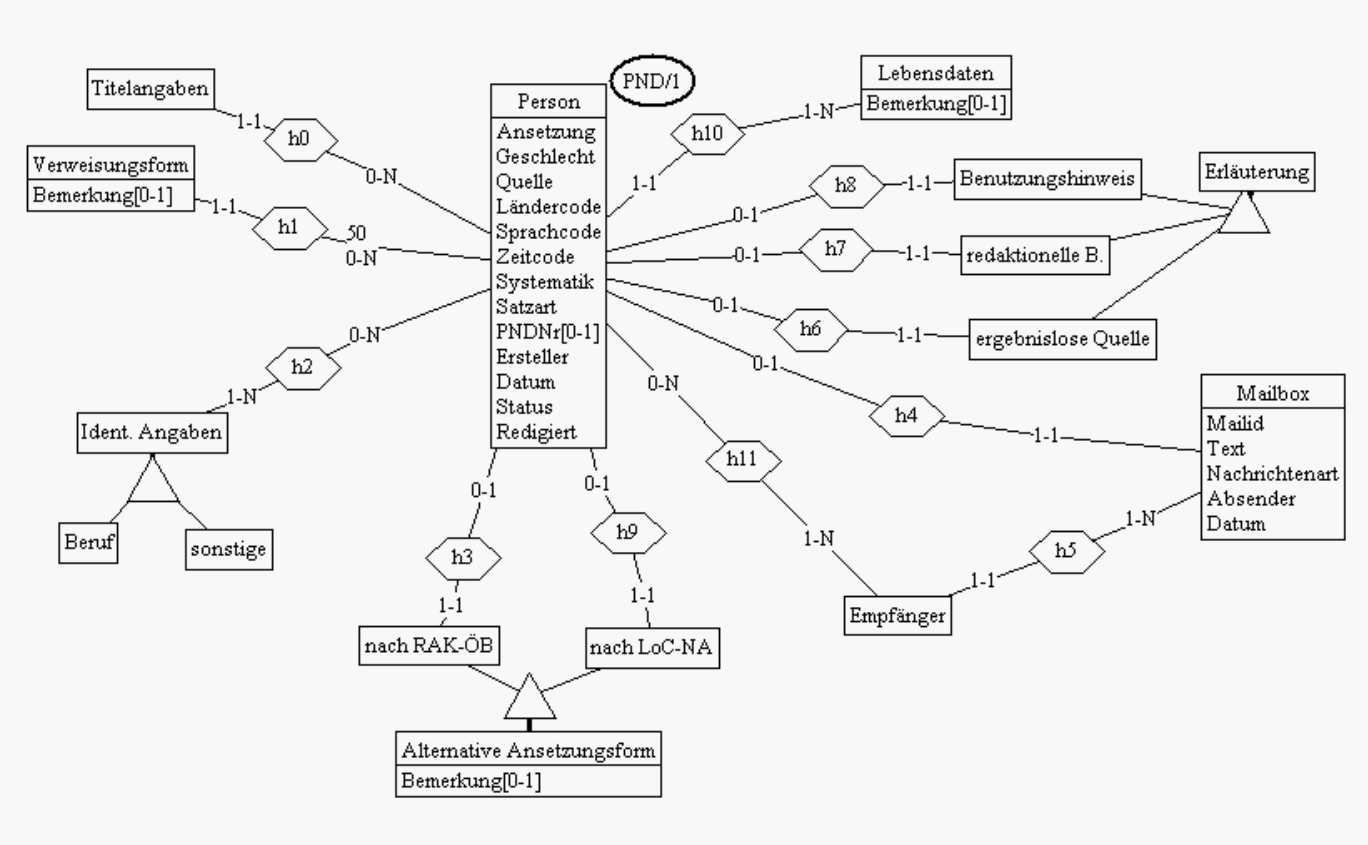
Legende:

- Leerzeichen/Blank/Spatium wird durch das Zeichen "¬" dargestellt
- *. Feld/Unterfeld ist wiederholbar
- (@{) bzw. ({) Der Inhalt des betreffenden Unterfeldes ist sortierrelevant und enthält Steuerzeichen zur Kennzeichnung nicht-sortierender Bestandteile.

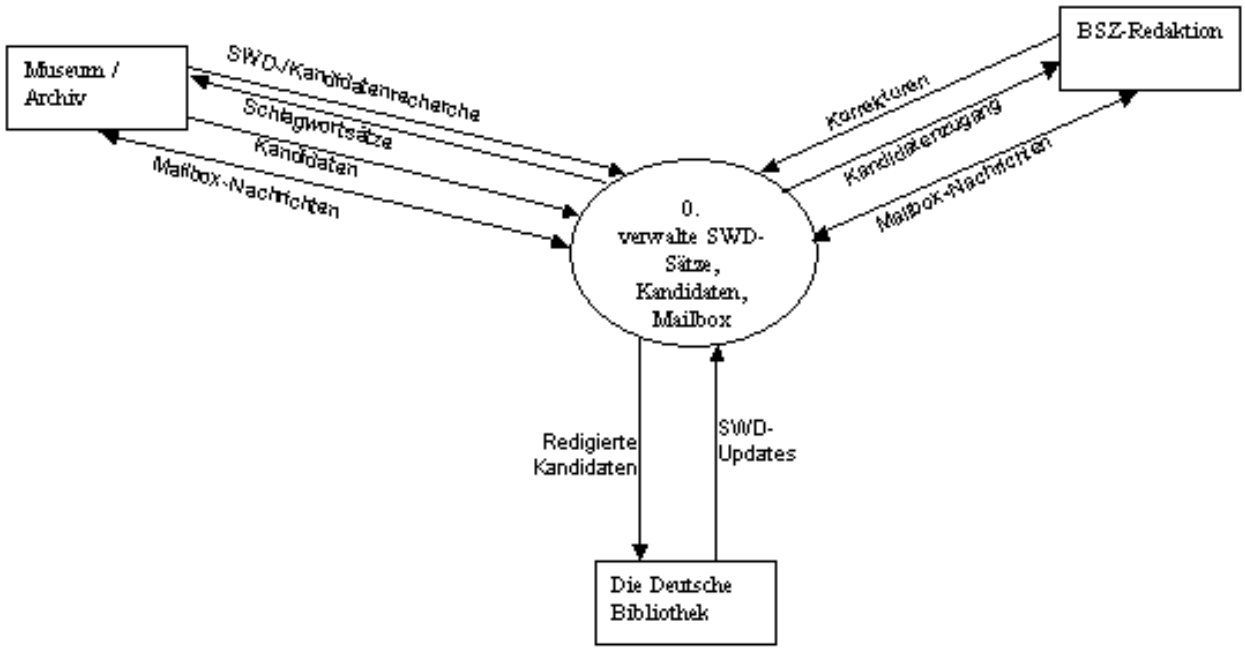
Anlage 2: Entity Relationship Diagramm eines SWD- bzw. GKD-Satzes



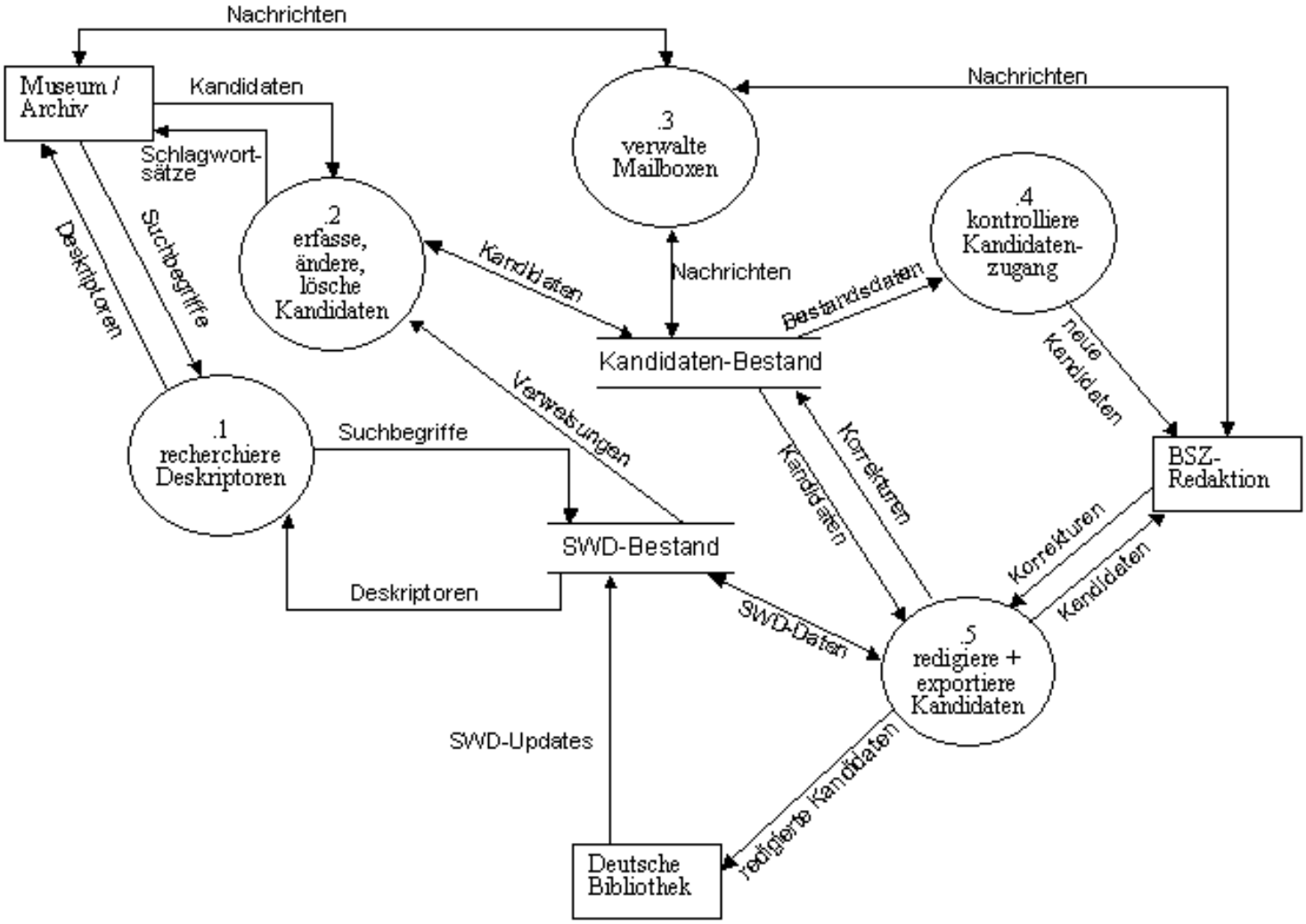
Anlage 3: Entity Relationship Diagramm eines PND-Satzes



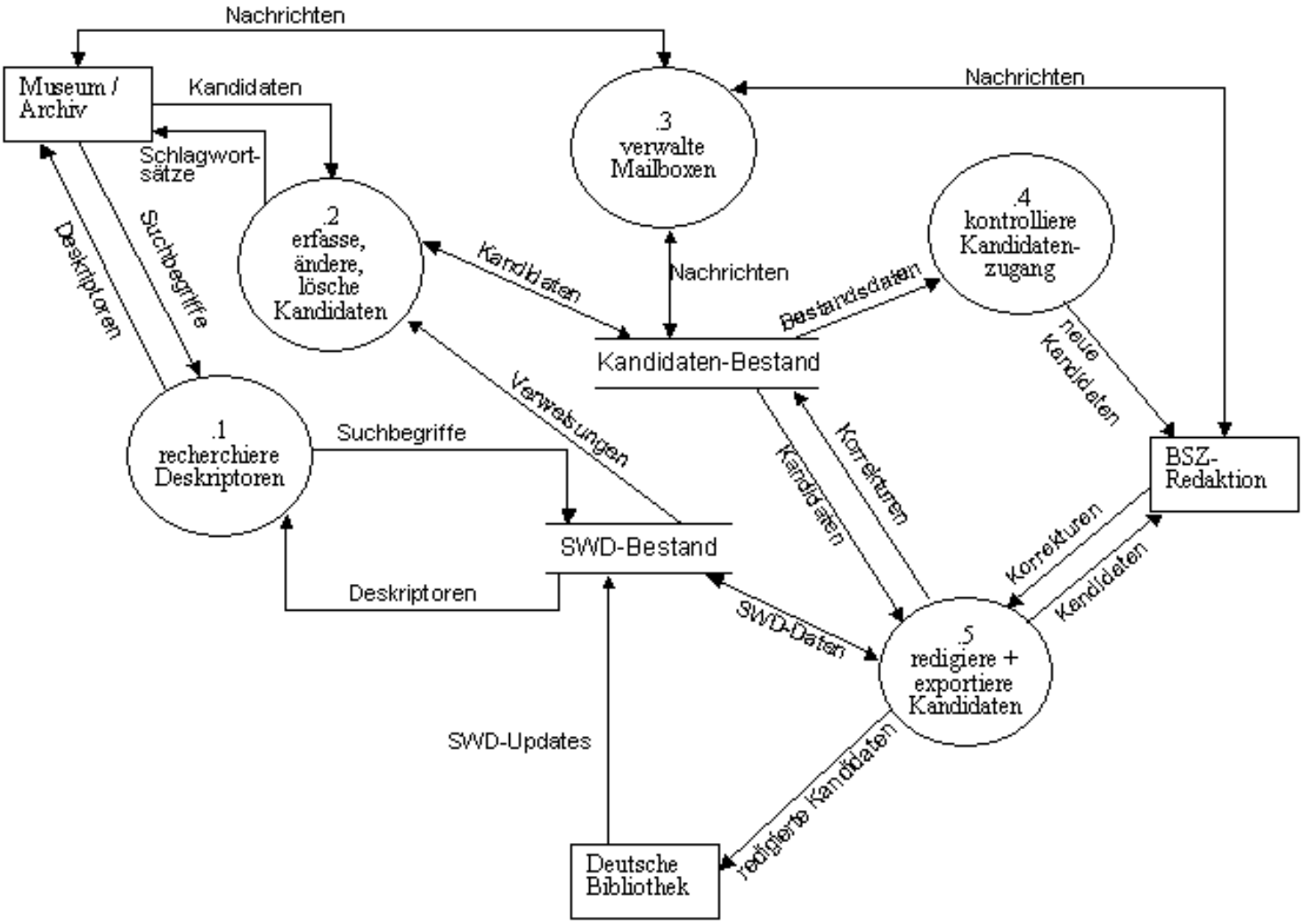
Anlage 4: Kontextdiagramm der Erweiterungsanwendung



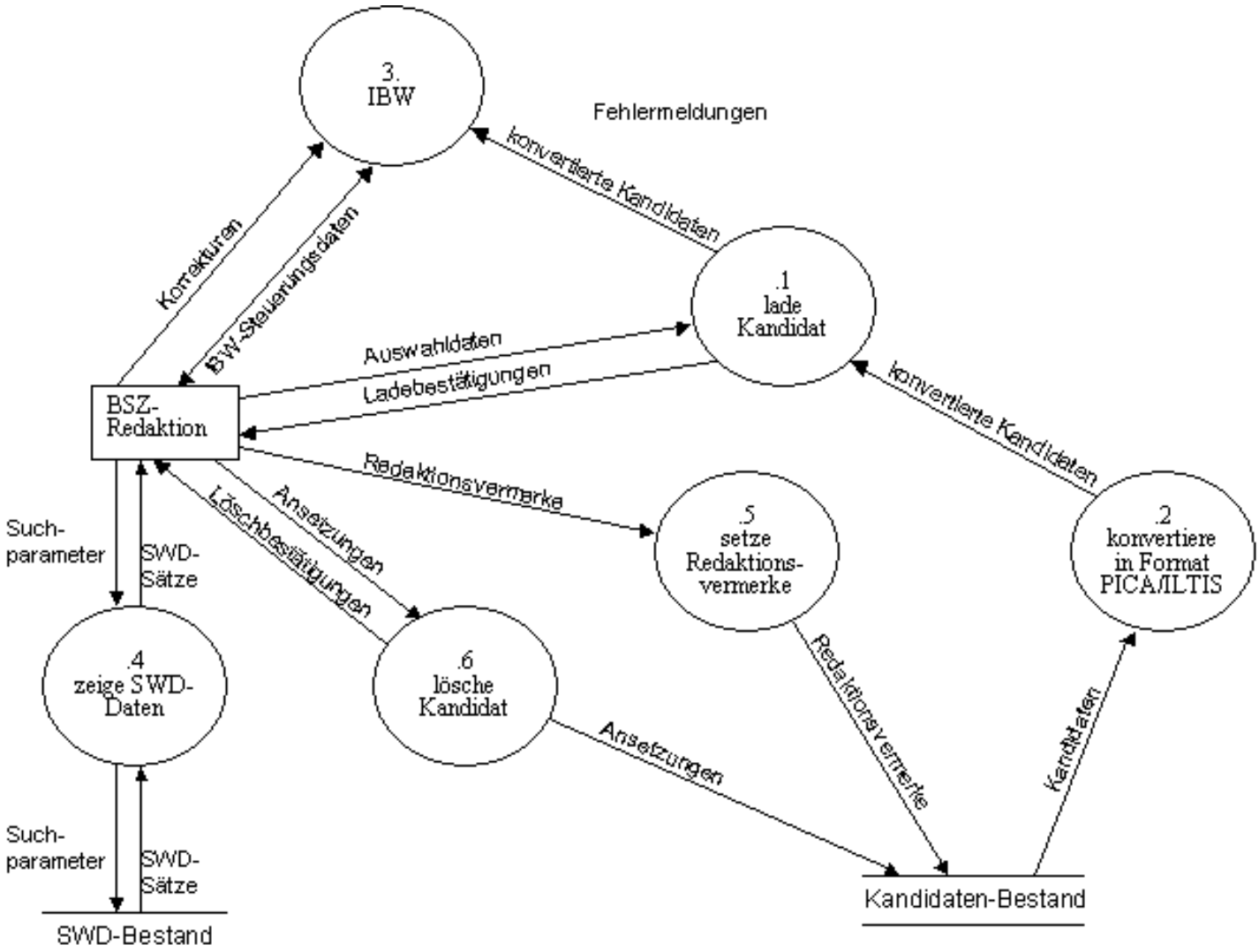
Anlage 5: Datenflussdiagramm 0 der Erweiterungsanwendung (DFD 0)



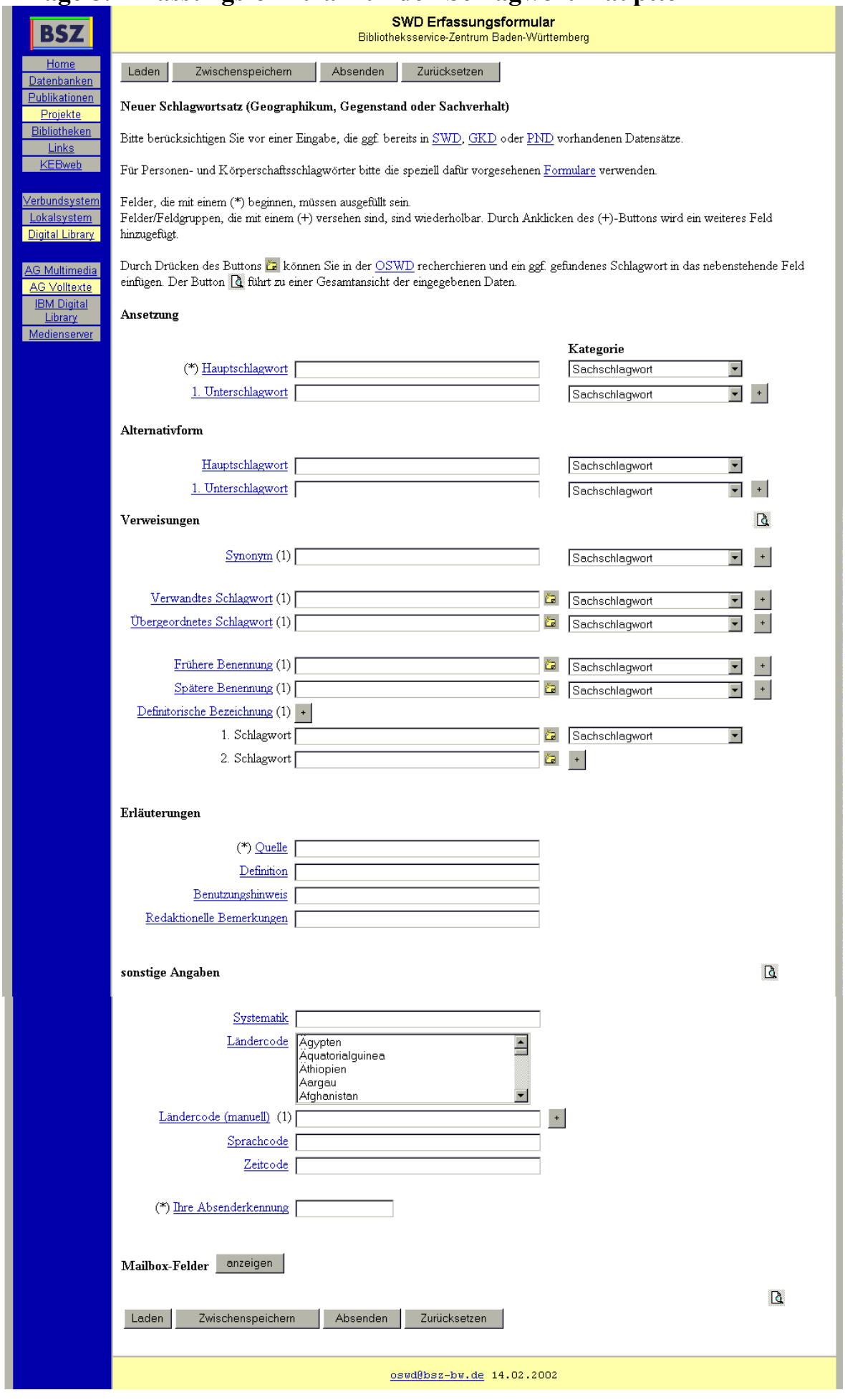
Anlage 6: Verfeinertes Datenflussdiagramm der Kandidatenerfassung (DFD 2)



Anlage 7: Verfeinertes Datenflussdiagramm der Redaktion (DFD 5)



Anlage 8: Erfassungsformular für den Schlagwort-Hauptteil



BSZ

Home
Datenbanken
Publikationen
Projekte
Bibliotheken
Links
KEBweb

Verbundsystem
Lokalsystem
Digital Library

AG Multimedia
AG Volltexte
IBM Digital Library
Medienserver

SWD Erfassungsformular
Bibliotheksservice-Zentrum Baden-Württemberg

Laden | Zwischenspeichern | Absenden | Zurücksetzen

Neuer Schlagwortsatz (Geographikum, Gegenstand oder Sachverhalt)

Bitte berücksichtigen Sie vor einer Eingabe, die ggf. bereits in SWD, GKD oder PND vorhandenen Datensätze.

Für Personen- und Körperschaftsschlagwörter bitte die speziell dafür vorgesehenen Formulare verwenden.

Felder, die mit einem (*) beginnen, müssen ausgefüllt sein.
Felder/Feldgruppen, die mit einem (+) versehen sind, sind wiederholbar. Durch Anklicken des (+)-Buttons wird ein weiteres Feld hinzugefügt.

Durch Drücken des Buttons können Sie in der OSWD recherchieren und ein ggf. gefundenes Schlagwort in das nebenstehende Feld einfügen. Der Button führt zu einer Gesamtansicht der eingegebenen Daten.

Ansetzung

Kategorie

(*) Hauptschlagwort — Sachschlagwort
1. Unterschlagwort — Sachschlagwort +

Alternativform

Hauptschlagwort — Sachschlagwort
1. Unterschlagwort — Sachschlagwort +

Verweisungen

Synonym (1) — Sachschlagwort +
Verwandtes Schlagwort (1) — Sachschlagwort +
Übergeordnetes Schlagwort (1) — Sachschlagwort +
Frühere Benennung (1) — Sachschlagwort +
Spätere Benennung (1) — Sachschlagwort +
Definitorische Bezeichnung (1) +
1. Schlagwort — Sachschlagwort
2. Schlagwort +

Erläuterungen

(*) Quelle
Definition
Benutzungshinweis
Redaktionelle Bemerkungen

sonstige Angaben

Systematik
Ländercode: Ägypten, Äquatorialguinea, Äthiopien, Aargau, Afghanistan
Ländercode (manuell) (1) +
Sprachcode
Zeitcode

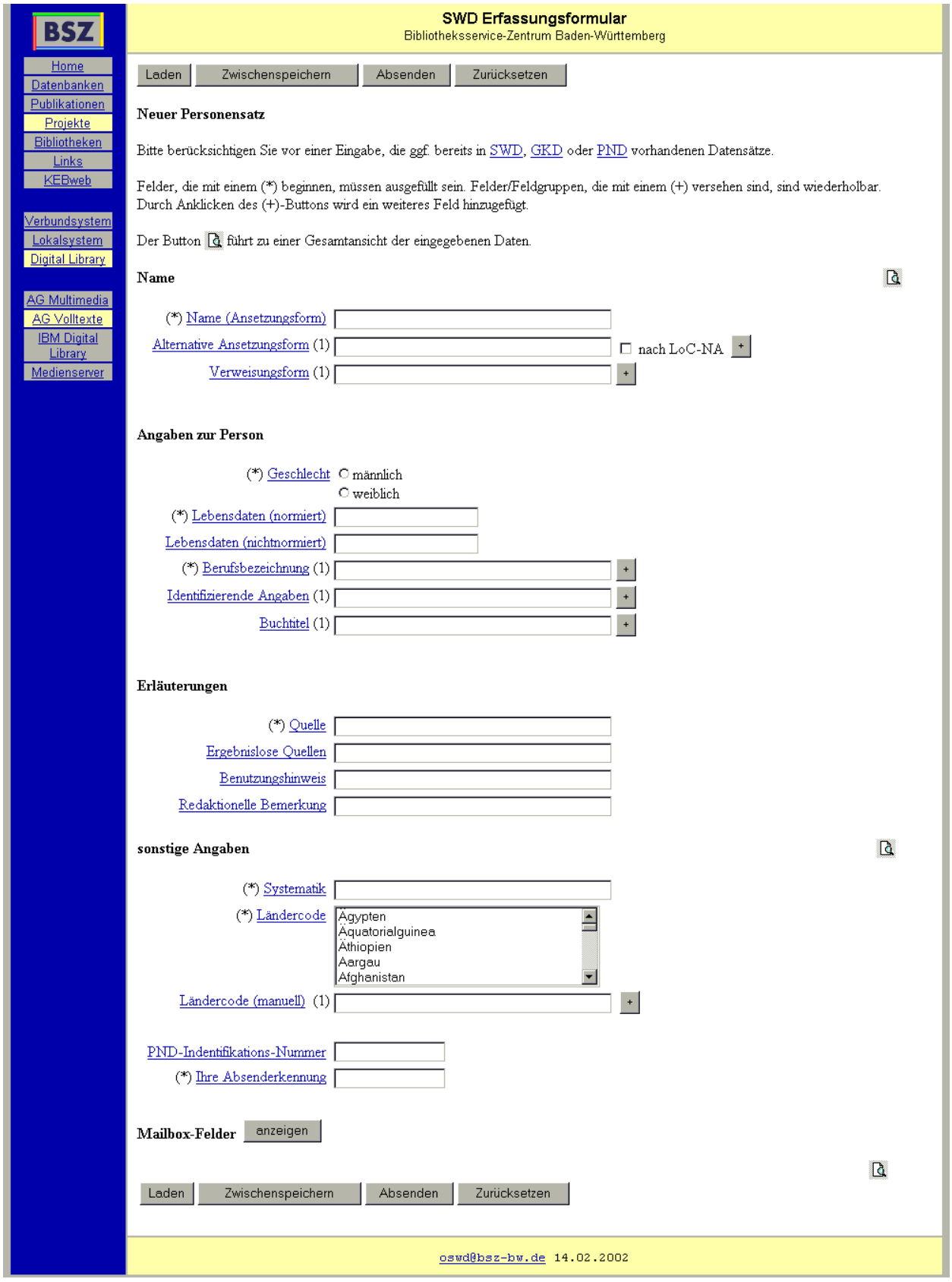
(*) Ihre Absenderkennung

Mailbox-Felder anzeigen

Laden | Zwischenspeichern | Absenden | Zurücksetzen

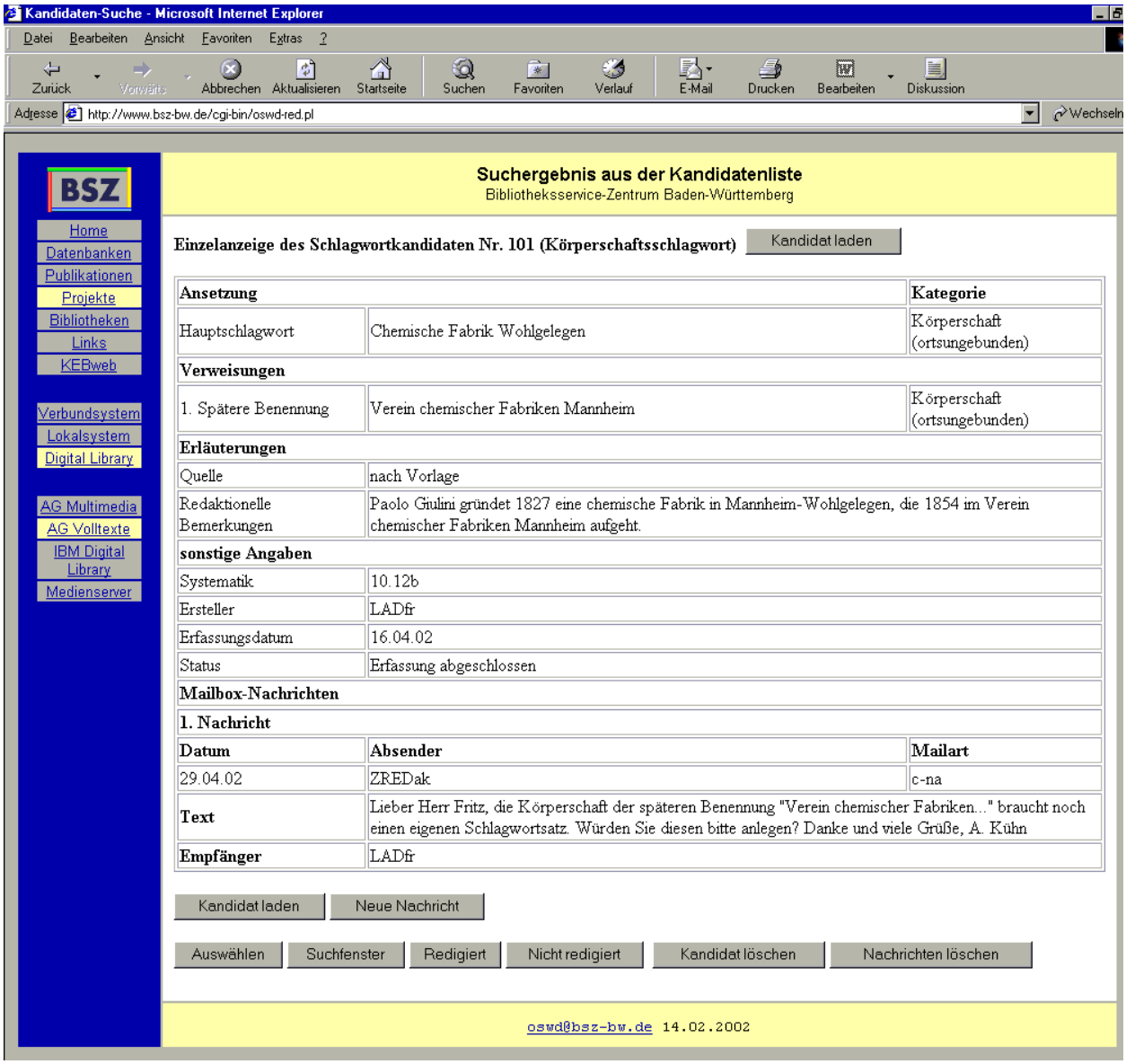
oswd@bsz-bw.de 14.02.2002

Anlage 9: Erfassungsformular für Personensätze



Anlage 10: Eingabemaske für die Suche auf den Schlagwortsätzen

Anlage 11: Einzelanzeige eines Schlagwortkandidaten



Kandidaten-Suche - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion

Adresse http://www.bsz-bw.de/cgi-bin/oswd-red.pl Wechseln

BSZ

Home
Datenbanken
Publikationen
Projekte
Bibliotheken
Links
KEBweb

Verbundsystem
Lokalsystem
Digital Library

AG Multimedia
AG Volltexte
IBM Digital Library
Medienserver

Suchergebnis aus der Kandidatenliste
Bibliotheksservice-Zentrum Baden-Württemberg

Einzelanzeige des Schlagwortkandidaten Nr. 101 (Körperschaftsschlagwort) Kandidat laden

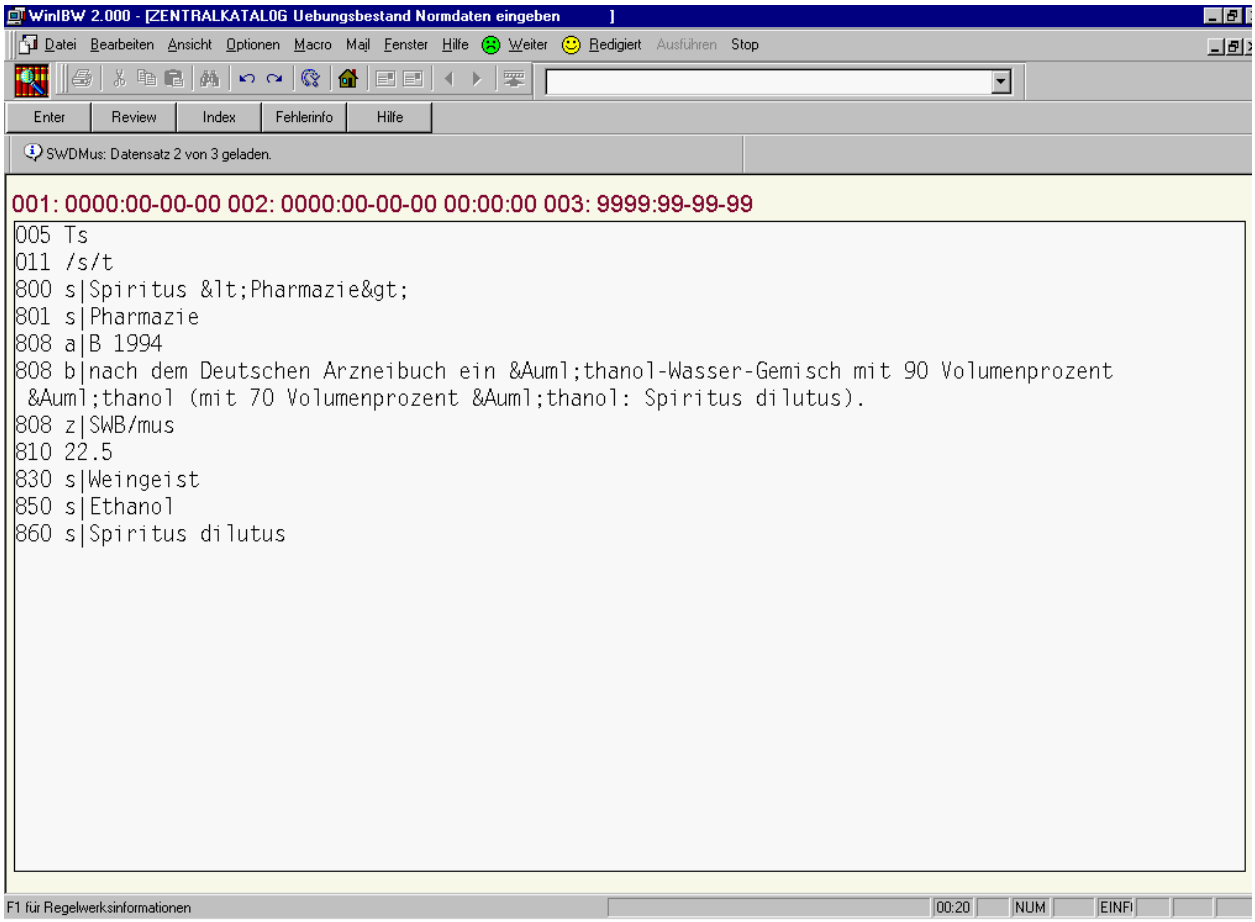
Ansetzung		**Kategorie**
Hauptschlagwort	Chemische Fabrik Wohlgelegen	Körperschaft (ortsungebunden)
Verweisungen		
1. Spätere Benennung	Verein chemischer Fabriken Mannheim	Körperschaft (ortsungebunden)
Erläuterungen		
Quelle	nach Vorlage	
Redaktionelle Bemerkungen	Paolo Giulini gründet 1827 eine chemische Fabrik in Mannheim-Wohlgelegen, die 1854 im Verein chemischer Fabriken Mannheim aufgeht.	
sonstige Angaben		
Systematik	10.12b	
Ersteller	LADfr	
Erfassungsdatum	16.04.02	
Status	Erfassung abgeschlossen	
Mailbox-Nachrichten		
1. Nachricht		
Datum	**Absender**	**Mailart**
29.04.02	ZREDak	c-na
Text	Lieber Herr Fritz, die Körperschaft der späteren Benennung "Verein chemischer Fabriken..." braucht noch einen eigenen Schlagwortsatz. Würden Sie diesen bitte anlegen? Danke und viele Grüße, A. Kühn	
Empfänger	LADfr	

Kandidat laden | Neue Nachricht

Auswählen | Suchfenster | Redigiert | Nicht redigiert | Kandidat löschen | Nachrichten löschen

oswd@bsz-bw.de 14.02.2002

Anlage 12: Eingabefeld der IBW mit geladenem Kandidatensatz



10. Literatur

[Abiteboul et al 2000] Abiteboul, Serge; Buneman, Peter, Suci, Dan: Data on the Web: From Relations to Semistructured Data and XML. San Francisco (Morgan Kaufmann Publishers) 2000

[Ankolekar et al 2001] Ankolekar, Anupriya; Burstein, Mark; Hobbs, Jerry; Lassila, Ora; Martin, David; McIlraith, Sheila; Narayanan, Srini; Paolucci, Massimo; Payne, Terry; Sycara, Katia; Zeng, Honglei: DAML-S: Semantic Markup For Web Services. In: Proceedings of the International Semantic Web Working Symposium (SWWS), July 30-August 1, 2001
Online verfügbar unter:
http://www.daml.org/services/SWWS.pdf
(Stand: 06.04.2002)

[Balzert 1996] Balzert, Helmut: Lehrbuch der Software-Technik. Software-Entwicklung. Heidelberg, Berlin (Spektrum) 1996.

[Bearman, Sochats 1995] Bearman, David; Sochats, Ken: Metadata Requirements for Evidence. Automating 21st Century Science - The Legal, Regulatory, Technical and Social Aspects of Electronic Laboratory Notebooks and Collaborative Computing. In: R&D. TeamScience Publishing, 1996
Online verfügbar unter:
http://www.archimuse.com/papers/nhprc/BACartic.html
(Stand: 08.02.2002)

[Becker et al 2000] Becker, Hans; Lepschy, Petra; Tappenbeck, Inka: 3. META-LIB Workshop an der SUB Göttingen. Metadata: New developments – new frontiers. In: Bibliotheksdienst, Heft 10, 2000. 2000
Online verfügbar unter:
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_2000/00_10_07.htm
(Stand: 06.04.2002)

[Beckett 2002] Beckett, Dave: Dave Beckett's Resource Description Framework (RDF) Resource Guide. 2002
Online verfügbar unter:
http://www.ilrt.bris.ac.uk/discovery/rdf/resources/#sec-apps
(Stand: 25.03.2002)

[Behme 2000] Behme, Henning: XML in der Praxis : Professionelles Web-Publishing mit der Extensible Markup Language. 2., erweiterte Auflage. München (Addison Wesley) 2000

[Berners-Lee 1998] Berners-Lee, Tim: Why RDF model is different from the XML model. 1998
Online verfügbar unter:
http://www.w3.org/DesignIssues/RDF-XML.html
(Stand: 25.03.2002)

[Berners-Lee 1998a] Berners-Lee, Tim: Semantic Web Road map. 1998
Online verfügbar unter:
http://www.w3.org/DesignIssues/Semantic.html
(Stand: 03.04.2002)

[Berners-Lee 1998b] Berners-Lee, Tim: What the Semantic Web can represent. 1998
Online verfügbar unter:
http://www.w3.org/DesignIssues/RDFnot.html
(Stand: 10.04.2002)

[Berners-Lee 1999] Berners-Lee, Tim: Web Architecture from 50,000 feet. 1999
Online verfügbar unter:
http://www.w3.org/DesignIssues/Architecture.html
(Stand: 10.04.2002)

[Berners-Lee et al 1999] Berners-Lee, Tim; Connolly, Dan; Swick, Ralph: Web Architecture: Describing and Exchanging Data. W3C Note 7 June 1999.
Online verfügbar unter:
http://www.w3.org/1999/06/07
(Stand: 10.04.2002)

[Berners-Lee 2000] Berners-Lee, Tim: Semantic Web on XML. Vortrag bei XML-2000 Konferenz, 3.-8.12.2000, Washington DC. 2000
Vortragsfolien online verfügbar unter:
http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html
(Stand: 10.04.2002)

[Berners-Lee, Connolly 1995] Berners-Lee, Tim; Connolly, Dan: Hypertext Markup Language - 2.0. 1995
Online verfügbar unter: http://www.w3.org/MarkUp/
(Stand: 04.03.2002)

[Berners-Lee, Fischetti 1999] Berners-Lee, Tim; Fischetti, Mark: Der Web-Report. München (Econ) 1999.

[Berners-Lee, Hendler 2001] Berners-Lee, Tim; Hendler, James: Nature Debates: Scientific publishing on the ‘semantic web’. In: Nature, Volume 410, 26.04.2001, S. 1023f. Macmillan Publishers (London) 2001.
Online verfügbar unter:
http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm
(Stand: 10.04.2002)

[BSZ 2002] Bibliotheksservice-Zentrum Baden-Württemberg: Nachweis Bestands- und Titelmeldungen der Jahre 1986 – 2001. Stand 31.12.2001.
Online verfügbar unter:
http://www.bsz-bw.de/statistik/best-tit.html
(Stand: 19.04.2002)

[Broekstra et al 2000] Broekstra, Jeen; Klein, Michel; Decker, Stefan; Fensel, Dieter; Horrocks, Ian: Adding formal semantics to the Web building on top of RDF Schema. In: Electronic Proceedings ECDL 2000 Workshop on the Semantic Web, 21. September 2000, Lissabon (Portugal), 2000.
Online verfügbar unter:
http://www.ics.forth.gr/proj/isst/SemWeb/proceedings/session2-2/paper.pdf
(Stand: 03.04.2002)

[Burke 2002] Burke, Murray: Darpa Agent Markup Language (DAML).
Online verfügbar unter:
http://dtsn.darpa.mil/ixo/daml%2Easp
(Stand: 03.04.2002)

[Burkart 1997] Burkart, Margarete: Thesaurus. In: Buder, Marianne; Rehfeld, Werner; Seeger, Thomas; Strauch, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation. 4. völlig neu gefasste Aufl. München (KG Saur) 1997.

[Capelleveen 1997] Capelleveen, Remco van: Zur Bedeutung von Metadaten für die Erschließung von Internetquellen (Dublin Core, Warwick Framework etc.).
Online verfügbar unter:
http://www.ub.fu-berlin.de/~rvc/web/metadaten.html
(Stand: 02.02.2002)

[Clark 1999] Clark, Scott: Back To The Basics: META Tags, Part 2.
Online verfügbar unter:
http://www.webdeveloper.com/html/html_metatags_part2.html
(Stand: 05.02.2002)

[CEN 2000] EUROPEAN COMMITTEE FOR STANDARDIZATION (CEN): CEN Workshop Agreement 13988: Agreement on the use of Dublin Core in Europe. 2000.
Online verfügbar unter:
http://www.cenorm.be/isss/cwa_download_area/cwa13988.pdf
(Stand: 08.03.2002)

[Cover 2002] Cover, Robin: The XML Cover Pages, Extensible Markup Language (XML)
Online verfügbar unter:
http://xml.coverpages.org/xml.html#applications
(Stand 18.03.2002)

[DAML 2002] DARPA Agent Markup Language Initiative: Language Feature Comparison. 2002
Online verfügbar unter:
http://www.daml.org/language/features.html
(Stand: 05.04.2002)

[DAML 2002a] DARPA Agent Markup Language Initiative: DAML Applications. 2002
Online verfügbar unter:
http://www.daml.org/applications/
(Stand: 05.04.2002)

[Das et al 2001] Das, Aseem; Wu, Wei; McGuinness, Deborah: Industrial Strength Ontology Management. Stanford Knowledge Systems Laboratory Technical Report KSL-01-09 2001. In: Proceedings of the International Semantic Web Working Symposium, Stanford, CA, Juli 2001.
Online verfügbar unter:
http://www.ksl.stanford.edu/people/dlm/papers/verticalNetOntologyBuilder/Industrial%20Strength%20Ontology%20Management2.pdf
(Stand: 05.04.2002)

[Decker et al 2000] Decker, Stefan; Melnik, Sergey; van Harmelen, Frank; Fensel, Dieter; Klein, Michel; Broekstra, Jeen; Erdmann, Michael; Horrocks, Ian: THE SEMANTIC WEB. The Roles of XML and RDF. In: IEEE Internet Computing, Vol. 4, No. 5, S. 63-74, Sept/Oct 2000. 2000
Online verfügbar unter:
http://www.ontoknowledge.org/oil/downl/IEEE00.pdf
(Stand: 25.03.2002)

[Decker et al 2002] Decker, Stefan; Fensel, Dieter; van Harmelen, Frank; Horrocks, Ian; McGuinness, Deborah; Patel-Schneider, Peter; Staab, Steffen; Stein, Lynn: OIL Case Studies.
Online verfügbar unter:
http://www.ontoknowledge.org/oil/case-studies/
(Stand: 05.04.2002)

[Dekkers, Weibel 1999] Dekkers, Makx; Weibel, Stuart: Dublin Core Metadata Initiative Progress Report and Workplan for 2002. In: Corporation for National Research Initiatives, Reston, Virginia (Hrsg.): D-Lib Magazine, Volume 8, Nr. 2, Februar 2002.
Online verfügbar unter:
http://www.dlib.org/dlib/february02/weibel/02weibel.html
(Stand: 08.03.2002)

[Dempsey et al 1997] Dempsey, Lorcan: Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats.
Online verfügbar unter:
http://www.ukoln.ac.uk/metadata/DESIRE/overview/
(Stand 04.03.2002)

[DCMI 1999] Dublin Core Metadata Initiative: Dublin Core Metadata Element Set, Version 1.1: Reference Description. 1999
Online verfügbar unter:
http://dublincore.org/documents/dces/
(Stand 06.03.2002)

[DCMI 2000] Dublin Core Metadata Initiative: Dublin Core Qualifiers. 2000.
Online verfügbar unter:
http://dublincore.org/documents/dcmes-qualifiers/
(Stand 06.03.2002)

[DCMI 2002a] Dublin Core Metadata Initiative: History of the Dublin Core Metadata Initiative.
Online verfügbar unter: http://dublincore.org/about/history/
(Stand: 05.03.2002)

[DCMI 2002b] Dublin Core Metadata Initiative: Dublin Core Projects
Online verfügbar unter: http://dublincore.org/projects/index.shtml
(Stand 08.03.2002)

[DCMI 2002c] Dublin Core Metadata Initiative: DCMI Frequently Asked Questions (FAQ)
Online verfügbar unter:
http://dublincore.org/resources/faq/
(Stand 08.03.2002)

[DBI 1990] Deutsches Bibliotheksinstitut: Die Schlagwortnormdatei, Entwicklungsstand und Nutzungsmöglichkeiten: Vorträge eines Kolloquiums zur Schlagwortnormdatei (SWD) in Frankfurt a. M. am 5. und 6. Oktober 1989. Berlin (DBI) 1990.

[DDB 2000] Die Deutsche Bibliothek: Schlagwortnormdatei (SWD). Ausgabe Oktober 2000. Frankfurt am Main (Die Deutsche Bibliothek) 2000.

[DIN 1988] Deutsches Institut für Normung e. V.: DIN 1463, Teil 1: Erstellung und Weiterentwicklung von Thesauri. Einsprachige Thesauri. Berlin 1988.

[DLMI 2000] Digital Library Metadata Initiative: Digital Library Metadata Initiative (DLMI).
Online verfügbar unter:
http://enite.zdv.uni-tuebingen.de/jdlmeta/dlmi/index.jsp
(Stand: 18.03.2002)

[Eggers 2001] Eggers, Björn: Die Schlagwortnormdatei als Erschließungsinstrument für Online-Publikationen. 2001.
Online verfügbar unter:
http://www.bsz-bw.de/opus/volltexte/2001/48/pdf/DA.pdf
(Stand 22.04.2002)

[Eversberg 1999] Eversberg, Bernhard: Was sind und was sollen Bibliothekarische Datenformate. Überarbeitete und erweiterte Neuausgabe. In: Brandes, Dietmar (Hrsg.): Veröffentlichungen der Universitätsbibliothek Braunschweig, Heft 9. Braunschweig 1999
Online verfügbar unter:
http://www.biblio.tu-bs.de/allegro/formate/formate.htm
(Stand 04.03.2002)

[Fensel 1999] Fensel, Dieter: Relating Ontology Languages and Web Standards. In: Ebert, Jürgen (Hrsg.): Modelle und Modellierungssprachen in Informatik und Wirtschaftsinformatik : Beiträge des Workshops "Modellierung 2000", St. Goar, 5. - 7. April 2000, S. 111-128, Koblenz (Fölbach) 2000.
Online verfügbar unter:
http://www.cs.vu.nl/~dieter/ftp/paper/mod2000.pdf
(Stand: 03.04.2002)

[Fensel et al 2001] Fensel, Dieter; van Harmelen, Frank; Horrocks, Ian; McGuinness, Deborah; Patel-Schneider, Peter: OIL: An Ontology Infrastructure for the Semantic Web.
In: IEEE Intelligent Systems, Vol. 16, No. 2, March/April 2001
Online verfügbar unter:
http://www.ksl.stanford.edu/people/dlm/papers/ieee-oil.pdf
(Stand: 05.04.02)

[Fikes, McGuinness 2001] Fikes, Richard; McGuinness, Deborah: An Axiomatic Semantics for RDF, RDF Schema and DAML+OIL. KSL Technical Report KSL-01-01, 2001. Updated October, 2001.
Online verfügbar unter:
http://www.ksl.stanford.edu/people/dlm/daml-semantics/daml+oil-axioms-october2001.htm
(Stand: 04.04.2002)

[Gennusa 1999] Gennusa, Pamela: Evolution and use of generic markup languages. In: Möhr, Wiebke; Schmidt, Ingrid (Hrsg.): SGML und XML. Anwendungen und Perspektiven. Berlin, Heidelberg (Springer) 1999.

[Goldfarb 1999] Goldfarb, Charles:Future Directions in SGML/XML. In: Möhr, Wiebke; Schmidt, Ingrid (Hrsg.): SGML und XML. Anwendungen und Perspektiven. Berlin, Heidelberg (Springer) 1999.

[Gruber 1993] Gruber, Tom: A Translation Approach to Portable Ontology Specifications. In: Knowledge Acquisition, Vol. 5, S. 199-220, 1993.
Online verfügbar unter:
http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html
(Stand 29.03.2002)

[Guarino, Giaretta 1995] Guarino, Nicola; and Giaretta, Pierdaniele: Ontologies and Knowledge Bases. Towards a Terminological Clarification. In: N. Mars (Hrsg.): Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing. S. 25-32. Amsterdam (IOS Press) 1995.

[Guarino 1998] Guarino, Nicola: Formal Ontology and Information Systems. In: Guarino, Nicola (Hrsg.): Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6. –8. Juni 1998, S. 3-15, Amsterdam (IOS Press) 1998.

Online verfügbar unter:
http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf
(Stand: 16.04.2002)

[Hacker 2000] Hacker, Rupert: Bibliothekarisches Grundwissen. 7., neu bearbeitete Aufl. München (K G Saur) 2000

[Heflin, Hendler 2000] Heflin, Jeff; Hendler, James: Dynamic Ontologies on the Web
In: Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000), S. 443-499. Menlo Park, CA (AAAI/MIT Press) 2000.
Online verfügbar unter:
http://www.cs.umd.edu/projects/plus/SHOE/pubs/aaai2000.pdf
(Stand: 03.04.2002)

[Hendler 2001] Hendler, James: Agents and the Semantic Web. In: IEEE Intelligent Systems Journal, Special Issue on the Semantic Web, Volume 16, Nr. 2, März/April 2001, S. 30-37. 2001.
Online verfügbar unter:
http://www.cs.umd.edu/users/hendler/AgentWeb.html
(Stand: 10.04.2002)

[Huemer 2001] Huemer, Christian: Electronic Business XML, Grundlagen und Nutzung. In: Turowski, Klaus; Fellner, Klement (Hrsg.): XML in der betrieblichen Praxis, Standards Möglichkeiten Praxisbeispiele. Heidelberg (dpunkt-Verlag) 2001.

[IFLA 2002] International Federation of Library Associations and Institutions, Den Haag: DIGITAL LIBRARIES: Metadata Resources
Online verfügbar unter:
http://ifla.inist.fr/II/metadata.htm
(Stand: 04.03.2002)

[INT Media 2002] INT Media Group: Search Engine Features For Webmasters.
Online verfügbar unter: http://searchenginewatch.com/webmasters/features.html
(Stand 11.02.2002)

[Kokkelink, Schwänzl 2001] Kokkelink, Stefan; Schwänzl, Roland: Expressing Qualified Dublin Core in RDF/XML. 30.11.2001
Online verfügbar unter:
http://dublincore.org/documents/2001/11/30/dcq-rdf-xml/
(Stand: 08.03.2002)

[Kunze 1999] Kunze, John: Encoding Dublin Core Metadata in HTML. 1999
Online verfügbar unter:
http://www.ietf.org/rfc/rfc2731.txt
(Stand: 07.03.2002)

[Lagoze 2000] Lagoze, Carl: Accomodating Simplicity and Complexity in Metadata: Lessons from the Dublin Core Experience. 2000
Online verfügbar unter:
http://www.archiefschool.nl/docs/lagopdf.pdf
(Stand 07.03.2002)

[Lancaster 1986] Lancaster, Frederick: Vocabulary Control for Information Retrieval. 2. Auflage, Arlington (Information Resources Press) 1986.

[LoC 2001] Library of Congress: Metadata Encoding & Transmission Standard (METS), METS: An overview and tutorial.
Online verfügbar unter:
http://www.loc.gov/standards/mets/METSOverview.html
(Stand 18.03.2002)

[Loeser 1998] Loeser, Henrik: Techniken für Web-basierte Datenbankanwendungen: Anforderungen, Ansätze, Architekturen. In: Informatik, Forschung und Entwicklung. Volume 13, Issue 4 (1998), S. 196-216. Berlin, Heidelberg (Springer) 1998.

[Lowe, Wilde 2001] Lowe, David; Wilde, Erik: Improving Web linking using XLink. In: Proceedings of Open Publish 2001. First Open Publish Conference. July 30, 2001 - August 02, 2001. Sydney Hilton Hotel, Sydney, Australia 2001.
Online verfügbar unter:
http://www.planetpdf.com/mainpage.asp?webpageid=1550
(Stand: 17.03.2002)

[Ludewig 2002] Ludewig, Karin: Nutzung der SWD bei der Objektdokumentation in Museen. Vortrag beim 92. Deutschen Bibliothekartag. 09.04-12.04.2002, Augsburg. 2002

[Lynch et al 1995] Lynch, Clifford; Michelson, Avra; Summerhill, Craig; Preston, Cecilia: White Paper on Networked Information Discovery and Retrieval NIDR. Washington, DC (Coalition for Networked Information) 1995.
Online verfügbar unter:
http://www.cni.org/projects/nidr
(Stand: 11.02.2002)

[Lynch 1998] Lynch, Clifford: The Dublin Core Descriptive Metadata Program: Strategic Implications for Libraries and Networked Information Access. In: A Bimonthly Report in Research Library Issues and Actions from ARL, CNI and SPARC, Issue 196. Februar 1998.
Online verfügbar unter:
http://www.arl.org/newsltr/196/dublin.html
(Stand: 11.02.2002)

[Mädche et al 2001] Mädche, Alexander; Staab, Steffen; Studer, Rudi: Ontologien. In: Wirtschaftsinformatik, WI-Schlagwort, 4/2001. Wiesbaden (Friedrich Vieweg & Sohn) 2001.
Online verfügbar unter:
http://www.aifb.uni-karlsruhe.de/~sst/Research/Publications/stichwort.pdf
(Stand: 05.04.2002)

[Mädche, Staab 2001] Mädche, Alexander; Staab, Steffen: Applying Semantic Web Technologies for Tourism Information Systems. In Karl Wöber, Andrew Frew, Martin Hitz (Hrsg.): Proceedings of the 9th International Conference for Information and Communication Technologies in Tourism, ENTER 2002, 23 – 25. Januar 2002. Innsbruck (Springer) 2002.
Online verfügbar unter:
http://www.aifb.uni-karlsruhe.de/~sst/Research/Publications/ENTER-2002-Paper_maedche-staab.pdf
(Stand: 10.04.02)

[Mainberger 2001] Mainberger, Christof: BAM-Projekt in Baden-Württemberg (Bibliotheken-Archive-Museen). 11.04.2001
Online verfügbar unter:
http://www.bsz-bw.de/diglib/bam/projekt.html
(Stand: 23.04.2002)

[McIlraith et al 2001] McIlraith, Sheila; Son, Tran; Zeng, Honglei: Semantic Web Services. In: IEEE Intelligent Systems, Special Issue on the Semantic Web, Volume 16, No. 2, S. 46-53, März/April 2001.
Online verfügbar unter:
http://www.daml.org/services/ieee01-KSL.pdf
(Stand: 05.04.2002)

[Miller 1996] Miller, Paul: Metadata for the masses. In: Ariadne, Version 5.
Online verfügbar unter:
http://www.ariadne.ac.uk/issue5/metadata-masses/
(Stand: 10.02.2002)

[Miller 1998] Miller, Eric: An Introduction to the Resource Description Framework. In: Corporation for National Research Initiatives, Reston, Virginia (Hrsg.): D-Lib Magazine, Mai 1998.
Online verfügbar unter:
http://www.dlib.org/dlib/may98/miller/05miller.html
(Stand: 23.03.2002)

[Milstead, Feldman 1999]: Milstead, Jessica; Feldman, Susan:
Metadata: Cataloging by Any Other Name ... .Online Januar 1999.
Online verfügbar unter: http://www.onlinemag.net/OL1999/milstead1.html
(Stand: 05.02.2002)

[Münz 2001] Münz, Stefan: SELFHTML – Die Energie des Verstehens: HTML-Dateien selbst erstellen. Version 8.0 vom 27.10.2001
Online verfügbar unter:
http://selfhtml.teamone.de/index.htm
(Stand 05.03.2002)

[OAI 2001] Open Archives Initiative: The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 1.1 of 2001-07-02.
Online verfügbar unter:
http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html
(Stand 18.03.2002)

[Palmer 2001] Palmer, Sean: The Semantic Web: An Introduction.
Online verfügbar unter:
http://infomesh.net/2001/swintro/
(Stand: 10.04.02)

[Pott, Wielage 1999] Pott, Oliver; Wielage, Günter: XML: Praxis und Referenz. München (Markt und Technik) 1999.

[Rusch-Feja 1997] Rusch-Feja, Diann: Mehr Qualität im Internet: Entwicklung und Implementierung von Metadaten. In: Ockenfeld, Marlies/Schmidt, Ralph (Hrsg.): 19. Online-Tagung der DGD. Die Zukunft der Recherche – Rechte, Ressourcen und Referenzen. Frankfurt am Main 14 bis 16. Mai 1997, S. 113-130. Frankfurt/Main (DGD) 1997

[Sonnenreich, Macinta 1998] Sonnenreich, Wes; Macinta, Tim: Web Developer.com Guide to Search Engines. New York, NY (John Wiley & Sons) 1998.

[Stephan, Scholze 1999] Stephan, Werner; Scholze, Frank: Online Publikationsverbund: Erfassung und Organisation elektronischer Hochschulschriften. In: Bibliotheksdienst, Heft 1, S. 92-102, 1999.

Online verfügbar unter:
http://elib.uni-stuttgart.de/opus/volltexte/1999/424/pdf/424_1.pdf
(Stand: 23.04.2002)

[Tennant 1998]: Tennant, Roy: 21st-Century Cataloging. In: Digital Library Journal, April 1998.
Online verfügbar unter:
http://www.libraryjournal.com/articles/infotech/digitallibraries/19980415_2627.asp
(Stand: 07.02.2002)

[van Harmelen, Fensel 1999] van Harmelen, Frank; Fensel, Dieter: Practical Knowledge Representation for the Web. In: Fensel, Dieter (Hrsg.): Proceedings of the IJCAI'99 Workshop on Intelligent Information Integration. 1999
Online verfügbar unter:
http://www.cs.vu.nl/~frankh/postscript/IJCAI99-III.html
(Stand: 03.04.2002)

[van Harmelen et al 2001] van Harmelen, Frank; Patel-Schneider, Peter; Horrocks, Ian: Reference description of the DAML+OIL (March 2001) ontology markup language. 2001
Online verfügbar unter:
http://www.daml.org/2001/03/reference.html
(Stand: 04.04.02)

[van Harmelen et al 2001b] van Harmelen, Frank; Patel-Schneider, Peter; Horrocks, Ian: A Model-Theoretic Semantics for DAML+OIL (March 2001). 2001
Online verfügbar unter:
http://www.daml.org/2001/03/reference.html
(Stand: 04.04.02)

[W3C 1998] World Wide Web Consortium: HTML 4.0 Specification W3C Recommendation revised on 24-Apr-1998. 1998
Online verfügbar unter:
http://www.w3.org/TR/1998/REC-html40-19980424/
(Stand 07.03.2002)

[W3C 1998a] World Wide Web Consortium: Extensible Markup Language (XML) 1.0 W3C Recommendation 10-February-1998. 1998
Online verfügbar unter:
http://www.w3.org/TR/1998/REC-xml-19980210
(Stand 14.03.2002)

[W3C 1999] World Wide Web Consortium: Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999. 1999
Online verfügbar unter:
http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/
(Stand: 23.03.02)

[W3C 2000] World Wide Web Consortium: Resource Description Framework (RDF) Schema Specification 1.0. W3C Candidate Recommendation 27 March 2000. 2000
Online verfügbar unter:
http://www.w3.org/TR/2000/CR-rdf-schema-20000327/
(Stand: 23.03.02)

[W3C 2001] World Wide Web Consortium: RDF/XML Syntax Specification (Revised). W3C Working Draft 18 December 2001. 2001
Online verfügbar unter:
http://www.w3.org/TR/rdf-syntax-grammar/
(Stand: 23.03.02)

[W3C 2002] World Wide Web Consortium: W3C XML Pointer, XML Base and XML Linking.
Online verfügbar unter:
http://www.w3.org/XML/Linking
(Stand: 17.03.2002)

[W3C 2002a] World Wide Web Consortium: XML Schema.
Online verfügbar unter:
http://www.w3.org/XML/Schema
(Stand: 18.03.2002)

[W3C 2002b] World Wide Web Consortium: Requirements for a Web Ontology Language. W3C Working Draft 07 March 2002. 2002
Online verfügbar unter:
http://www.w3.org/TR/2002/WD-webont-req-20020307/
(Stand: 25.03.2002)

[W3C 2002c] World Wide Web Consortium: Hypertext Markup Language Homepage. 2002
Online verfügbar unter:
http://www.w3.org/MarkUp/
(Stand: 06.04.2002)

[Waldt, Drummond 2002] Waldt, Dale; Drummond, Rik: EBXML: The Global Standard for Electronic Business.
Online verfügbar unter:
http://www.ebxml.org/presentations/global_standard.htm
(Stand 19.03.2002)

[Weibel 1995] Weibel, Stuart: Metadata: The Foundations of Resource Description. In: Corporation for National Research Initiatives, Reston, Virginia (Hrsg.): D-Lib-Magazine, Juli 1995.
Online verfügbar unter:
http://www.dlib.org/dlib/July95/07weibel.html
(Stand 06.02.2002)

[Weibel 1997] Weibel, Stuart: The Evolving Metadata Architecture for the World Wide Web: Bringing Together the Semantics, Structure and Syntax of Resource Description.
In: Proceedings on the International Symposium on Research, Development and Practice in Digital Libraries : ISDL'97. Tsukuba (Japan), 18. – 21.11.1997.
Online verfügbar unter:
http://www.dl.ulis.ac.jp/ISDL97/proceedings/weibe.html
(Stand: 28.02.2002)

[Weibel 1999] Weibel, Stuart: The State of the Dublin Core Metadata Initiative April 1999. In: Corporation for National Research Initiatives, Reston, Virginia (Hrsg.): D-Lib Magazine, Volume 5, Nr. 4, April 1999.
Online verfügbar unter:
http://www.dlib.org/dlib/april99/04weibel.html
(Stand: 08.03.2002)

[Wersig 1985] Wersig, Gernot: Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. 2. ergänzte Auflage, München (KG Saur) 1985.