

**Effizientere Nutzung bibliographischer
Metadaten im wissenschaftlichen Bereich
durch Verbesserung der Datenqualität**

**Dissertation
zur Erlangung des Grades eines
Doktors der Philosophie
der Philosophischen Fakultät III
der Universität des Saarlandes**

vorgelegt von

Robert Kolatzek

aus Heilsberg / Lidzbark Warmiński

Saarbrücken, 2012

Prof. Dr. J. Kubiniok

Berichterstatter

Prof. Dr. Harald H. Zimmermann

Prof. Dr. Hans Uszkoreit

Tag der Disputation: 6.2.2012

Inhaltsverzeichnis

1	Einführung.....	1
1.1	Wissenschaftliche Literatur und Informationsmanagement.....	1
1.2	Zielsetzung.....	5
1.3	Methoden.....	7
2	Grundbegriffe und Definitionen	9
2.1	Grundlagen der Formalerschließung.....	11
2.1.1	Dublin Core als Beispiel für ein einfaches Format für Metadaten	11
2.1.2	Regeln für die alphabetische Katalogisierung (RAK) und DIN 1505-1.....	15
2.1.3	Anglo-American Cataloguing Rules (AACR2).....	22
2.1.4	Transkription und Transliteration.....	24
2.2	Weitere Normen, Begriffe und Definitionen.....	26
2.2.1	DIN- und ISO-Normen im bibliographischen Bereich.....	27
2.2.2	Datenformate.....	30
2.2.2.1	Textbasierte Datenformate.....	32
2.2.2.2	Binäre Datenformate.....	52
2.2.3	Austauschprotokolle.....	57
2.2.3.1	Z39.50.....	58
2.2.3.2	MedLine (PubMed).....	61
3	Vorhandene bibliographische Softwaresysteme.....	67
3.1	Lokale bibliographische Softwaresysteme	70
3.1.1	Vorteile lokaler Speicherung und Verarbeitung.....	71
3.1.2	Schwachstellen lokaler Speicherung und Verarbeitung.....	73
3.1.3	Umsetzung.....	75
3.1.3.1	BibTeX.....	75
3.1.3.2	OpenOffice.....	79
3.2	Bibliographische Softwaresysteme mit Datenübertragung über das Internet (nach Server-Client-Modell)	82
3.2.1	Vorteile einer Übernahme bibliographischer Daten.....	83

3.2.1.1	Deutsche Einheitsaufnahme (CIP).....	84
3.2.1.2	Akzeptanz.....	85
3.2.2	Schwachstellen einer Übernahme bibliographischer Daten.....	87
3.2.2.1	Erschließung nach RAK bzw. AACR2 als Voraussetzung für die Datenqualität.....	89
3.2.2.2	Schwachstellen in der Verarbeitung von Quellenangaben in unterschiedlichen Zitiernormen.....	91
3.2.3	Umsetzung.....	94
3.2.3.1	Web-Gateway.....	95
3.2.3.2	Reference Manager als Beispiel für kommerzielle Umsetzung eines in ein persönliches Literaturverwaltungssystem integrierten Z39.50-Clients.....	100
3.2.3.3	Tellico als Beispiel für Umsetzung eines Z39.50-Clients im Rahmen einer Software zur Verwaltung von Sammlungen	102
3.2.4	MedLine als Beispiel für eine Artikeldatenbank.....	105
3.2.4.1	Gateway von MedLine.....	106
3.2.4.2	Web-Oberfläche.....	107
3.3	Vorläufiges Fazit.....	109
4	Neuere Entwicklungen im bibliographischen Bereich.....	111
4.1	Erwartungen.....	111
4.2	Umsetzung.....	115
4.2.1	ZING: SRU (SRW).....	116
4.2.1.1	SRU-Operationen.....	118
4.2.1.2	CQL als Abfrage-Sprache des SRU/W.....	119
4.2.1.3	XML-basierte Datenaustauschformate.....	120
4.2.2	Weitere Ansätze.....	132
1.	Vorläufiges Fazit.....	139
3.	Entwurf einer bibliographischen Auszeichnungssprache.....	144
1.	Anforderungen.....	145
1.	Einfachheit der Erfassung.....	147
2.	Internationalität.....	149
3.	Atomare Feldstruktur.....	150
4.	Maschinell verarbeitbar.....	151

5. Erweiterbar nach Bedarf.....	151
6. Verwendung in anderen Datenformaten.....	152
7. Identität, Integrität und Homogenität der Daten.....	153
2. Ansätze zur Realisierung.....	154
1. Drei-Ebenen-System.....	154
2. Unicode.....	158
3. XML.....	158
1. XSLT-Stile.....	161
2. Erweiterbar nach Bedarf	162
3. Einbindung in andere Dateien.....	163
4. Hierarchische und relationale Struktur.....	164
5. Vertikale Beziehungen (bibliographische Hierarchie).....	167
6. Horizontale Beziehungen.....	169
3. Umsetzung.....	170
1. Hauptklassen.....	171
1. Serie.....	172
2. Buch.....	174
3. Periodikum.....	176
4. Artikel.....	178
2. Hilfsklassen.....	180
1. Person.....	180
2. Organisation.....	185
3. Verlag.....	187
4. Abstract.....	188
3. Besondere Datentypen.....	189
1. Relation.....	189
2. Klassifikation, Schlagwort, Sprache.....	192
3. Autor.....	195
4. Herausgeber.....	196
5. Mitarbeiter.....	197
4. Maschinelle Verarbeitung.....	197
1. Auflösung von Abhängigkeiten.....	198
2. Normgerechte Einordnung.....	201

3. Normgerechte Formatierung.....	202
4.Abschluss.....	204
1. Vorteile des entwickelten Formats.....	205
2. Schwachstellen des neuen Formats.....	207
3. Fazit.....	209
5.Definition des Datenformats.....	I
1. Datenblätter.....	I
1. Hauptklassen	I
2. Hilfsklassen.....	VI
3. Datentypen.....	IX
2. XML-Schema für das Datenformat.....	XI
3. Verwendungsbeispiel.....	XXVII
6.Literaturverzeichnis.....	XXXIII
7.Glossar.....	XLIII

Abbildungsverzeichnis

Abbildung 1: Erzeugen von Hierarchien aus dem RIS-Format.....	40
Abbildung 2: Erzeugen von Hierarchien im BibTeX.....	44
Abbildung 3: Darstellung des Objektorientierten Modells am Beispiel eines Buches.....	51
Abbildung 4: Client-Server-Kommunikation mit Z39.50 (nach Wolfram Schneider).....	54
Abbildung 5: Gefilterte Ergebnisliste einer MedLine-Abfrage.....	57
Abbildung 6: Übersicht über die Funktionsweise von CiteProc.....	128
Abbildung 7: Der Zusammenhang der einzelnen CiteProc-Dateien und ihre Funktion.....	129
Abbildung 8: Die graphische Darstellung des Drei-Ebenen-Systems.....	147
Abbildung 9: Graphische Darstellung des Klassenmodells.....	158

Abkürzungsverzeichnis

AACR(2) – Anglo-American Cataloging Rules

CIP – Cataloging in Publication

COinS – Context Objects in Spans

CSS – Cascading Style Sheets

DDC – Dewey Decimal Classification

DC – Dublin Core

DIN – Deutsches Institut für Normung

DNB – Deutsche Nationalbibliothek

DTD – Document Type Definition

DOI – Digital Object Identifier

FRBR – Functional Requirements for Bibliographic Records

HTML – HyperText Markup Language

ILTIS – Integrierte Literatur-, Tonträger- und Musikalien-Informationssystem

ISBD – International Standard Bibliographic Description

ISBN – International Standard Book Number

ISI – Institute for Scientific Information (heute Teil von Thomson Corporation)

ISSN – International Standard Serial Number

ISSN-L – International Standard Serial Number Linking

ISO – International Standard Organisation

LCSH – Library of Congress Subject Headings

LoC – Library of Congress

RIS – Research Information System

MAB(2) – Maschinelles Austauschformat für Bibliotheken

MARC – Machine-Readable Cataloging

MeSH – Medical Subject Headings

MODS – Metadata Object Description Schema

NABD – Normenausschuss Bibliotheks- und Dokumentationswesen

OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting

OCLC – Online Computer Library Center

ONIX – Online Information eXchange

PDF – Portable Document Format

PubMed – kostenfreie Datenbank mit Metadaten medizinischer Publikationen
des US-amerikanischen National Center for Biotechnology
Information (NCBI)

PND – Personennamendatei

RDF – Ressource Description Framework

RAK-WB – Regeln für die alphabetische Katalogisierung in wissenschaft-
lichen Bibliotheken

Relax NG – eine XML-Schema-Grammatik

SQL – Structured Query Language

SRU – Search / Retrieve via URL

SRW – Search/Retrieve Web Service

SULB – Saarländische Universitäts- und Landesbibliothek

SWD – Schlagwortnormdatei der DNB (entspricht der LCSH der LoC)

TCP/IP – Transmission Control Protocol/Internet Protocol

UPC – Universal Product Code

URN – Uniform Resource Name

URI – Uniform Resource Identifier

VOYAGER – Datenbank mit Metadaten zu Publikationen der LoC

XHTML – Extensible HyperText Markup Language

XML – eXtensible Markup Language

XMLS – XML-Schema (des W3C)

XSLT – Extensible Stylesheet Language Transformations

XSL-FO – Extensible Stylesheet Language – Formatting Objects

Z39.50 – Protokoll für eine datenbankunabhängige Recherche im Internet

ZING – Z39.50 International New Generation (früherer Name für SRU und
SRW)

Kurzreferat (Abstract)

Diese Arbeit beschäftigt sich mit den bibliographischen Metadaten, wie sie in der wissenschaftlichen Praxis verwendet werden. Ausgehend von der Betrachtung bestehender Normen und Formate aus dem Bereich der wissenschaftlichen Bibliotheken und der Literaturverwaltungssoftware werden Verbesserungsvorschläge entwickelt, die in einen Entwurf eines neuen Datenschemas einfließen. Dieser Entwurf soll dem Wissenschaftler eine effizientere (multifunktionale) Nutzung dieser Daten ermöglichen, indem den folgenden Problemfeldern besondere Beachtung geschenkt wird: einfache und präzise Recherche im Datenbestand, persönliche Bewertung der Inhalte (Interpretation), verlustfreier Austausch der Daten (Portabilität, Persistenz), direkter Zugang zum Volltext der Werke, internationale Nutzung (Unabhängigkeit von Zeichensätzen und Sprachen), bedarfsgerechte Erweiterbarkeit sowie die Verarbeitung zu Quellenangaben. Abschließend wird der Entwurf definiert und auf seine Eignung hin untersucht. (Die Umsetzung in Form eines XML-Schemas wird mit einigen Verwendungsbeispielen im Anhang beigelegt.)

Abstract

The aim of the thesis is to present bibliographic metadata and the way they can be utilized by scientists. At first, it focuses on the existing norms and formats in academic libraries and in literature management software which, thoroughly analysed, influence the concept of the schema of such data. This concept should enable a more efficient (multifunctional) use of the data by concentrating on the following issues: easy and precise research in the database, individual assessment of the content (interpretation), data portability, direct access to full text, international use (independent of character encoding and language), extensibility of the model as required and citation processing. Finally, the concept is described and examined whether it can be applied in practice. (Definition of xml schema together with a set of examples is appended to the thesis.)

Danksagung und Widmung

Mein Dank gilt allen, die mich in der Zeit der Entstehung dieser Arbeit auf vielfältige Art und Weise unterstützt haben: Vor allem meiner Frau, meiner Familie und meinen Freunden.

Dankbar bin ich meinem Promotor - Prof. Harald H. Zimmermann - für das Aufzeigen des weiten Horizonts der Informationswissenschaft. Er hat diese Arbeit von der ersten Idee an begleitet und um viele Facetten bereichert.

Ich widme diese Arbeit allen Wissenschaftlern, die die Wahrheit in der Wissenschaft und im Glauben suchen, die ethischen Grenzen achten und ihr Wissen in den Dienst am Menschen stellen.

1. Einführung

„Wissenschaftler aller Fachrichtungen leiden unter der Last einer Informationsflut, zu der sie mit jeder neuen Veröffentlichung selbst beitragen. Denn auf der ganzen Welt ist ein Millionenheer von Forschern damit beschäftigt, unsere Erkenntnis über die Gesetzmäßigkeiten in Natur und Gesellschaft zu erweitern. Das Produkt dieser Tätigkeit ist Information, die über den Weg der Fachzeitschriften und Fachbücher wieder in den Wissenschaftsprozess einfließt.“¹

1.1. Wissenschaftliche Literatur und Informationsmanagement

Mit diesen Worten beschreiben Werner Marx und Gerhard Gramm in ihrem Online-Artikel „Literaturflut – Informationslawine – Wissensexpllosion. Wächst der Wissenschaft das Wissen über den Kopf?“ eine Entwicklung im wissenschaftlichen Bereich, die schon seit Jahrzehnten anhält: Die Vermehrung des Wissens ist mit der Vervielfachung der Literatur verbunden. Diese nutzbar zu machen, ist die Aufgabe der Bibliothekswissenschaft² im

1 Werner, Max ; Gramm, Gerhard: *Literaturflut – Informationslawine – Wissensexpllosion. Wächst der Wissenschaft das Wissen über den Kopf?* URL: <http://www.fkf.mpg.de/ivs/literaturflut.html>. Stand: 05.9.2006

2 Vgl. Umstätter, Walther: *Bibliothekswissenschaft als Teil der Wissenschaftswissenschaft - unter dem Aspekt der Interdisziplinarität*. In: Umstätter,

"modernen" Sinn. Sie entwickelte zahlreiche Methoden, um den Umgang – vor allem im Bezug auf das Auffinden und Beschaffen – für den Kunden zu erleichtern. Dazu gehören Ordnungsprinzipien, Datenbanken und verschiedene Formate und Normen, die von nationalen und internationalen Normierungsgremien verabschiedet werden.

In all diesen Lösungen bewahrheitet sich der Satz, den Paepke im Bezug auf digitale Bibliotheken³ 1996 formulierte: „Searching is not enough“⁴. Seine These lautet: Um die gesammelten Informationen effizient nutzen zu können, muss die digitale Bibliothek folgendes ermöglichen: die Suche und Auswahl relevanter Dokumente; die Bereitstellung der gesuchten Informationen; die Interpretation der Suchergebnisse; die lokale Verwaltung der empfangenen Metadaten sowie der Austausch dieser mit Anderen. Dabei handelt es sich um typische Merkmale eines Informationsmanagementsystems.

Die bibliographischen Metadaten enthalten Informationen, die in mehreren **Kontexten** verwendet werden (Darstellung, Übertragung und Speicherung). Als solche sind sie Gegenstand der Informationswissenschaft. Diese Disziplin entwickelte Kriterien zur Beurteilung der Informationsqualität sowie Methoden zur Optimierung der Informationsvermittlung⁵. Einen für diese Arbeit wichtigen Forschungsschwerpunkt stellt das Informationsmanagement dar. Die wissenschaftlichen Erkenntnisse aus diesem Bereich können dazu

Walther (Hrsg.); Wessel, Karl-Friedrich (Hrsg.): *Interdisziplinarität - Herausforderung an die Wissenschaftlerinnen und Wissenschaftler*. Kleine Verlag, Bielefeld 1999. S. 146 ff

3 Unter „digitaler Bibliothek“ ist eine reale Ansammlung von Dokumenten zu verstehen, die Informationen zu einem Werk – wenn auch in vielen Dokumenten – enthält. Dieser Begriff ist vom übertragenen Sinn zu unterscheiden, der die Internetpräsenz einer Bibliothek meint, welche ihren Bestand nicht elektronisch ausliefert.

4 Paepke, Andreas: *Digital Libraries: Searching is not enough : What We Learned On-Site*. In: *D-Lib Magazine*, 1996, Mai. ISSN: 1082-9873. URL: <http://www.dlib.org/~dlib/may96/stanford/05paepcke.html> Stand: 05.09.2006

5 Vgl. Universität des Saarlandes, Fachrichtung Informationswissenschaft (Hrsg.): *Studienführer Informationswissenschaft*. URL: <http://is.uni-sb.de/studium/~studienfuehrer/kap1.html> Stand: 22.6.2008

beitragen, einzelne Informationen miteinander so zu vernetzen, dass der Umgang mit bibliographischen Metadaten effizienter wird.

Durch den vermehrten Einsatz von Computern im Bereich der Literaturverwaltung und Bestandserfassung stellen sich den Bibliothekaren sowie seinen Kunden die Frage nach einer computergerechten Aufbereitung dieser Informationen. Dieses Problem ist Forschungsgegenstand der Computerlinguistik und der Informatik. Diese Disziplinen entwickeln Methoden, um die natürliche Sprache "dem Rechner verständlich zu machen" und die daraus gewonnenen Informationen zufriedenstellend zu verarbeiten⁶.

Alle hier genannten Wissenschaften stellen Wissen bereit, um einen effizienten Umgang mit der „Literaturflut“ zu ermöglichen: Die Bibliothekswissenschaft definiert die zu verarbeitenden Datentypen, die Informatik und die Computerlinguistik tragen zur erfolgreichen elektronischen Verarbeitung dieser bei, während der richtige (effiziente und störungsfreie) Umgang mit diesen auf Erkenntnisse der Informationswissenschaft aufbaut.

Dennoch scheint es an einem Fortschritt aus der Sicht des Endbenutzers (der Wissenschaftler) zu mangeln, weil veraltete Formate und Normen das Umfeld der Verarbeitung bibliographischer Metadaten prägen. Die für den Menschen gut lesbaren Informationen werden per Software nicht immer zufriedenstellend zu Quellenangaben verarbeitet. Viele erfasste Daten gehen während des Austausches verloren, andere können nicht erfasst werden.

Angesichts nahezu exponentiell steigender Zahl an wissenschaftlichen Veröffentlichungen ist die Suche nach einem Weg zur effizienteren Handhabung wichtiger denn je⁷. Dabei werden die modernen Identifikatoren wie URI und DOI, die Kontextdokumentation (persönliche Anmerkungen,

6 Vgl. Universität Duisburg-Essen, Fakultät für Ingenieurwissenschaften, Abteilung Informatik und Angewandte Kognitionswissenschaft (Hrsg.): *Was ist Computerlinguistik?* URL: http://www.uni-due.de/computerlinguistik/c_und_l.shtml Stand: 27.06.2009

7 Vgl. Lautenschlager, Michael; Sens, Iris : *Konzept zur Zitierfähigkeit wissenschaftlicher Primärdaten*. In: Information. Wissenschaft und Praxis, 2003. S. 463-466

Abstracts, Beschlagwortung, Kategorisierung) sowie ein auf semantische Analyse ausgerichtetes Austauschformat für Metadaten immer wichtiger. Sie dienen einer besseren Organisation und Vernetzung der erfassten Daten.

Mit www.bibsonomy.org, www.librarything.com oder www.connotea.org⁸ existieren bereits die ersten Social-Bookmarking-Dienste für die Verwaltung (Beschlagwortung, Erfassung und Katalogisierung) und den Austausch von Literaturquellen (in Interessengruppen wie öffentlich). Dies ist eine neue Nutzungsart der bibliographischen Daten, die noch vor einigen Jahren nicht denkbar war: Eine Wissenschaftsgemeinde kommuniziert ihre Sammlung relevanter Literatur nicht mehr mittels Veröffentlichung, sondern direkt und wirkt so dem Trend zur „papiernen Sintflut“ effektiv entgegen. Das Sammeln und Austauschen der Literaturnachweise in wissenschaftlichen Gruppen auf direktem Wege scheint – gemessen an der stets zunehmenden Zahl solcher "Austauschorte" im Internet – seit einigen Jahren immer beliebter und sogar zu einem weiteren "Informationskanal" zu werden. Die Frage "Wen und was zitiert er/sie?" wird in der Zukunft möglicherweise durch "Wen und was hat er/sie gebookmarkt?" abgelöst. Auch solche neuartigen Phänomene gilt es zu berücksichtigen, wenn man an die Bewältigung der Informations- und Publikationsflut denkt.

Derartige Entwicklungen wirken sich auf das Selbstverständnis der wissenschaftlichen Bibliotheken aus. Bemüht um größere Nähe zum „Kunden“ arbeiten auch Universitätsbibliotheken mit diesen Diensten zusammen oder versuchen diese sogar nachzuahmen⁹. Will der Benutzer

8 Ein Angaben zu Leistungen und Kosten von Literaturverwaltungsprogrammen und (bibliographischen) Social-Bookmarking-Diensten werden in dieser Arbeit grundsätzlich von der Diplomarbeit von Annett Kerschis bezogen:

Kerschis, Annett : *Literaturverwaltung und Wissensorganisation im Vergleich. Das Angebot von Literaturverwaltungsprogrammen und Social Bookmarking in Bezug auf die Benutzbarkeit in Bibliotheken*. Potsdam, Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Diplomarbeit, 2007

9 Vgl. Flimm, Oliver : *Die Open-Source-Software OpenBib an der USB Köln - Überblick und Entwicklungen in Richtung OPAC 2.0* . In: *Bibliothek. Forschung und Praxis*, Jg. 31 (2007) Nr. 2 . S. 9ff URL: <http://eprints.rclis.org/archive/00009891/01/openbib-opac20->

jedoch die eigens online gepflegte Bibliographie ohne Internetzugang nutzen, verliert er während der Datenübernahme einen nicht unerheblichen Teil dieser Daten, weil das Austauschformat veraltet ist oder nicht zu diesem Zweck entwickelt wurde.

An Ideen für eine bessere Nutzbarkeit bibliographischer Metadaten scheint es nicht zu fehlen. Alle bisher entwickelten Datenformate und Normen bieten Lösungen für ein bestimmtes Einsatzgebiet wie Abgleich der Bestandsdaten (MARC und MAB), die Generierung von Quellenangaben (BibTeX) oder die Titelmeldung (ONIX for books beim VLB). Für eine effizientere Nutzung dieser Metadaten, die von der Erfassung über Kontextdokumentation, Austausch bis hin zur Generierung von Quellenangaben reicht, scheint jedoch eine tragende Säule zu fehlen: ein geeignetes und zukunftssicheres Datenformat.

1.2. Zielsetzung

In dieser Arbeit sollen die im wissenschaftlichen Umfeld gegenwärtig eingesetzten Lösungen hinsichtlich ihrer informationswissenschaftlichen Merkmale untersucht werden. Die gewonnenen Erkenntnisse werden für die **Verbesserung der Datenqualität** und die **Optimierung des Umgangs** mit eben diesen Metadaten verwendet. Als konkrete Ausgestaltung wird ein System von Datenfeldern angesehen, welches eine **effizientere Nutzung dieser Daten** in den **Kontexten** "Erfassung und Interpretation", "direkter Austausch der Rohdaten" sowie "Zitieren und Präsentieren" ermöglicht.

Um dem Benutzer eines solchen Systems unnötige Hürden bei der Erfassung zu ersparen, wird in dieser Arbeit auf Vollständigkeit der Veröffentlichungsarten sowie einige selten verwendete Datentypen verzichtet. Es gilt, eine möglichst hohe Qualität und Homogenität der bibliographischen Metadaten und ein zufriedenstellendes Recall-Precision-Verhältnis auf Basis dieser Daten zu erreichen sowie Redundanzen zu

vermeiden. Die nach diesem System erfassten Daten sollen zugleich für die automatische Generierung von Quellenangaben geeignet sein.

Das zu entwickelnde System wird in Form eines XML-RelaxNG-Schemas¹⁰ beschrieben. Dies soll eine Verwendung innerhalb anderer XML-basierter Formate ermöglichen und zugleich auf ein weltweit akzeptiertes Basis-Datenformat zurückgreifen. Die Erweiterbarkeit ist neben der internationalen Ausrichtung¹¹ eine der wesentlichen Eigenschaften des XML, über die auch das System verfügen soll.

Es sollen Technologien und wissenschaftliche Erkenntnisse der Informatik, der Computerlinguistik, der Bibliotheks- und der Informationswissenschaft genutzt werden, die bei der Erstellung eines solchen Systems als hilfreich erscheinen. Aus Gründen der Weitläufigkeit und des Detailreichtums solcher Lösungen werden sie lediglich ansatzweise mit ihren jeweiligen Stärken und Schwächen dargestellt. Sie sind kein Gegenstand dieser Arbeit, sondern als Hilfsmittel und zugleich als Modelle zu verstehen, die in dieser oder jener Form zum Einsatz kommen können wie z. B.: Thesauri und Ontology Web Language, XML und RDF oder Dewey Decimal Classification.

Das zu erstellende System von Datenfeldern ist kein Selbstzweck. Durch die Verbesserung der Datenqualität soll der Umgang mit diesen Daten in Literaturverwaltungssystemen verbessert werden, indem folgende **Operationen** eine besondere Beachtung bekommen:

- **Retrieval** = einfache und zugleich präzise Suche nach (Vor-, Nach-, Original- oder historischen) Namen, Titeln, Serien (und deren Teilen),

10 Die XML-Schema-Sprache RelaxNG wurde als internationaler Standard (ISO/IEC 19757-2) verabschiedet. Nicht zuletzt wegen der Verwendung für die Definition des OpenDocument-Formats (ISO/IEC 26300:2006) und ihrer Einfachheit erfreut sie sich einer großen Beliebtheit. Diese ist für die formelle Beschreibung der Syntax des zu entwickelnden Systems zweckmäßig.

11 Die „Internationalität“ von XML wird unter 5.3.3.2 auf Seite 192 genauer beschrieben. Hier soll nur darauf hingewiesen werden, dass es möglich sei, mehrere Sprachen in einem Dokument gleichzeitig zu verwenden und sie auszuzeichnen, so dass die Inhalte Sprachen-spezifisch verarbeitet werden können.

Schlagwörtern oder Kategorien (unter Beachtung möglicher Homonymien bei Schlagworten, Personennamen, Titeln etc)

- **Interpretation** = möglichst automatische Auswertung von Zusammenhängen zwischen Publikationen (Übersetzungen, Fortsetzungen, Serien); kurze Wiedergabe des Inhalts und Interpretation von vielen Benutzern
- **Austausch** = verlustfreier Export und Import von Daten (für den Austausch zwischen Arbeitsplätzen oder Benutzern)
- **Zugang zum Volltext** = die Möglichkeit der Verlinkung der bibliographischen Daten mit dem Volltext der Publikation
- **Verarbeitung zu Quellenangaben** = möglichst genaue Abbildung der Daten auf verschiedene Zitiernormen
- **Internationalisierung** bibliographischer Daten = die Daten sollen in beliebigen Zeichensätzen erfasst werden können (z. B. Kyrillisch, Hebräisch), wie auch deren Entsprechung in lateinischer Schrift oder einer phonetischen Umschrift (unter Wahrung der Äquivalenz solcher Sprachversionen)

1.3. Methoden

Ausgangspunkt dieser Arbeit ist die Betrachtung der im wissenschaftlichen Umfeld verwendeten bibliographischen Datenformate und Normen. Auf eine tiefer gehende Untersuchung dieser wird aufgrund ihrer hohen Zahl verzichtet. Neuartige und besonders erwähnenswerte Merkmale dieser Datenformate und Normen werden genauer geschildert. Die relevanten Grundbegriffe und Definitionen werden im Kapitel 2 zusammengetragen und erklärt. Sie sind nicht chronologisch geordnet und beziehen sich zum Teil auf Kapitel 3 und 4 – damit nehmen sie einige Entwicklungen, Normen und Standards vorweg.

Ein Blick auf die Ansätze und die Art und Weise der Verarbeitung der bibliographischen Daten in den vorhandenen Systemen soll im Kapitel 3

geschehen. Dargestellt werden darin Systeme und Software, die im XX. Jh. entstanden und nicht primär für den Einsatz im WWW entwickelt worden sind.

Die neueren Entwicklungen im Bereich der Recherche und der Übermittlung von bibliographischen Daten, die neben der Erneuerung des Z39.50-Protokolls und des MARC-Datenformats betreffen und auf XML als Datenformat setzen, sollen im Kapitel 4 hinsichtlich ihrer Eigenart skizziert werden.

Die in Kapiteln 2, 3 und 4 gewonnenen Erkenntnisse über die Beschaffung der Datenformate und deren Nutzung (Einsatzgebiete) sollen mit der beabsichtigten Nutzung der bibliographischen Metadaten und somit deren Anforderungen verglichen werden. Von diesem Vergleich ausgehend, werden im Kapitel 5 die notwendigen technischen Mittel gesucht, die eine Umsetzung ermöglichen. Anschließend wird anhand benötigter Datentypen ein System von Klassen sowie deren Relationen beschrieben.

Im Abschluss dieser Arbeit wird das entwickelte System einer Prüfung unterzogen, um seine Schwachstellen und seine Neuerungen kurz zusammenzufassen.

Dieses System ist auf die Nutzung in Computern bzw. in Computernetzwerken ausgerichtet, stellt selbst jedoch keine Software dar, so dass es nicht (oder nur hinsichtlich einiger weniger Merkmale) empirisch untersucht werden kann. Eine Beschreibung der Datenelemente und ein XML-Schema (begleitet von einigen Beispielen für seine Verwendung) wird im Anhang eingefügt.

2. Grundbegriffe und Definitionen

Der wichtigste Begriff dieser Arbeit ist „die Formalerschließung“ (auch als „formale Erfassung“ verwendet). Diese ist zusammen mit der Sacherschließung dem Gegenstand dieser Arbeit vorgelagert und bildet den Ausgangspunkt für die Untersuchung der Nutzungskontexte und der Datenformate. Zugleich wird diese Arbeit versuchen, Lösungen für mögliche Problemfelder aufzuzeigen, um eine höhere Effizienz bei der Nutzung der so entstandenen Daten zu erreichen. Die Formalerschließung wird wie folgt definiert:

„Unter Formalerschließung (auch formale Analyse, Formalkatalogisierung, alphabetische Katalogisierung oder Titelaufnahme) versteht man die konsistent nach bestimmten Regeln erfolgende Beschreibung und das Wiederauffindbar-Machen von Dokumenten nach äußerlichen, formalen Kriterien. ... Sie beinhaltet nicht nur eine bibliographische Beschreibung des Dokuments, sondern auch die Festlegung der Elemente, unter denen das Dokument suchbar sein soll (Eintragung). Man nennt diese Daten auch Auswertungselemente: Beispiele sind Titelangaben, Verfasser, Ausgabe- und

*Auflagenbezeichnung, ISBN-Nummer usw. Dafür ist jeweils die Ansetzungsform zu bestimmen.*¹²

Die **Formalerschließung** gibt „formale“ Regeln vor, nach denen jeder Mensch eine gültige Wiedergabe erreichen dürfte. Die Kenntnis des Inhalts eines Werkes ist dafür nicht erforderlich. Alle Daten (Auswertungselemente) können aus dem Titelblatt übernommen werden.

Die **Sacherschließung** bedient sich ebenfalls dieser Regeln, benötigt jedoch die Sachkenntnis des Erfassenden. Diese wird für die richtige Wiedergabe des Inhalts in Form von Klassifikation, Schlagworten oder Abstracts vorausgesetzt (bzw. als gegeben angenommen). Diese Angaben reichern die bibliographische Beschreibung an.

Beide Formen der Erschließung beschreiben ein Werk bzw. dessen Inhalt und generieren sog. Metadaten. Ein möglichst effizienter Umgang mit diesen ist der Gegenstand dieser Arbeit. Die Metadaten definiert KSS-Glossar wie folgt:

*„Mit Metadaten werden die Daten bezeichnet, die semantische, strukturelle, administrative und technische Daten über andere Daten bereitstellen. ... Die Interpretierbarkeit der einzelnen Datenelemente und ihrer Inhalte muss durch ausreichend präzise semantische Regeln sichergestellt sein.“*¹³

Daten, die bibliographische Werke beschreiben, nennt man „**bibliographische Metadaten**“. Die Untersuchung der bestehenden Regeln für bibliographische Metadaten wird in dieser Arbeit parallel zur Betrachtung der Erfassungsregeln vorgenommen, da beide sich gegenseitig beeinflussen und bedingen.

12 *Formalerschließung*. In: Kuhlen, Reiner (Hrsg.) ; Seeger, Thomas (Hrsg.) ; Strauch, Dietmar (Hrsg.) : *Grundlagen der praktischen Information und Dokumentation*. Bd. 2 : *Glossar*. 5. Auflage. München: K G Saur, 2004. - ISBN 3-598-11674-8 (dieses Standardwerk wird als „**KSS-Glossar**“ abgekürzt verwendet)

13 *Metadaten*. In: *KSS-Glossar*, S. 84

2.1. Grundlagen der Formalerschließung

An dieser Stelle sollte man die Notwendigkeit solcher „Formalien“ hinterfragen. Warum werden sie benötigt? Inwieweit muss reguliert werden? Welche Arten gibt es? Das folgende Anwendungsbeispiel wird bei der Klärung nützlich sein.

Um den Aufwand der Erfassung zu minimieren, wird im bibliothekarischen Alltag¹⁴ auf extern erfasste Daten zurückgegriffen. Eine unkontrollierte Übernahme fremder Daten kann zur Minderung der Qualität und Homogenität des Datenbestandes führen oder die daraus generierten Quellenangaben¹⁵ negativ beeinflussen. Um die bibliographischen Daten unterschiedlicher Quellen homogen (bzw. vergleichbar) zu gestalten, wurden Normen und Regeln entworfen, an welche die erschließenden Bibliothekare gebunden sind. Grundsätzlich kann man diese in drei Gruppen aufteilen: ein **Format für Metadaten** (das sog. „Datenformat“ - hier am Beispiel des Dublin Core), **Regeln für die formale Erfassung von Metadaten** (RAK und AACR2) und **Normen für Umschrift von Zeichen zwischen Sprachen mit unterschiedlichen Zeichensätzen** (Transkription und Transliteration).

2.1.1. Dublin Core als Beispiel für ein einfaches Format für Metadaten

Dublin Core geht auf die Initiative einer Gruppe von Teilnehmern einer Konferenz zum Thema WWW im Oktober 1994 zurück. Im Jahr darauf (März

14 vgl. Hilberer, Thomas : *Aufwand vs. Nutzen : Wie sollen deutsche wissenschaftliche Bibliotheken künftig katalogisieren?* In: Bibliotheksdienst 2003, Nr. 6, S. 758. Dieser Aufsatz bietet eine kritische Betrachtung der Formalerschließung in den deutschen wissenschaftlichen Bibliotheken. Nach fast zehn Jahren sind einige Forderungen umgesetzt worden. Einige Kritikpunkte werden in einem länger andauernden Reformprozess berücksichtigt.

15 Das Wort „**Quellenangabe**“ oder „**Quellennachweis**“ wird als Synonym zum Wort „**Zitat**“ verwendet. Dabei bedeutet das **Zitat** die wörtliche Übernahme einer fremden Aussage, während die **Quellenangabe** den Ort belegt, aus dem ein Zitat stammt. In dieser Arbeit wird der Begriff "Zitieren" im weiteren Sinne verwendet: als Synonym für das Erstellen von Quellenangaben (ungeachtet dessen, ob der Inhalt der Quelle wörtlich wiedergegeben wird oder nicht).

1995) in Dublin/Ohio wurden die Überlegungen im Rahmen des OCLC/NCSA Metatadata Workshops ausformuliert und in Form eines Formats für Metadaten verabschiedet. Dieser enthält 15 Datenelemente, die der Beschreibung von Web-Ressourcen dienen sollen. Diese sind¹⁶:

- Title (de. Titel)
- Subject (de. Thema – in Form von freien Schlagwörtern)
- Description (de. Beschreibung bzw. Abstract)
- Creator (de. Urheber – Autor oder Produzent)
- Contributor (de. andere beteiligte Personen)
- Date (de. Datum; Empfehlung: in einer der zulässigen Schreibweisen von <http://www.w3.org/TR/NOTE-datetime>)
- Type (de. Art der Ressource; Empfehlung: DCMITYPE <http://dublincore.org/documents/dcmi-type-vocabulary/>)
- Format (de. Format; Empfehlung: MIME <http://www.iana.org/assignments/media-types/>)
- Identifier (de. Identifikator; nicht genormt)
- Source (de. Quelle; nicht genormt)
- Language (de. Sprache – Empfehlung: abgekürzt nach RFC4646)
- Relation (de. Beziehung [zu anderen Dokumenten]; nicht genormt)
- Publisher (de. Verlag – die für die Verfügbarkeit zuständige Einheit)
- Coverage (de. Erfassung = zusätzliche Orts- oder Zeitangaben)

16 Angaben nach Schütz, Thomas : *Dokumentenmanagement*. In: Kuhlen, Reiner (Hrsg.) ; Seeger, Thomas (Hrsg.) ; Strauch, Dietmar (Hrsg.) : *Grundlagen der praktischen Information und Dokumentation. Bd. 1 : Handbuch zur Einführung in die Informationswissenschaft und -praxis*. 5. Auflage. München: K G Saur, 2004. - ISBN 3-598-11674-8 (dieses Standardwerk wird als „KSS“ abgekürzt verwendet), S. 341. Vgl. auch Dublin Core Metadata Initiative (Hrsg.) : *Dublin Core Metadata Element Set, Version 1.1*. URL: <http://dublincore.org/documents/2010/10/11/dces/> Stand: 11.10.2010

- Rights (de. Rechte = Nutzungsbedingungen bzw. Vervielfältigungsrechte; nicht genormt)

Mit Hilfe dieser Datenfelder soll der Benutzer bzw. ein Computer mit Informationen versorgt werden, die für den Umgang mit einem Dokument im WWW nützlich sind. Diese Angaben können zum Beispiel für die Erfassung in Datenbanken oder für die Generierung von Quellenangaben verwendet werden. Mit „coverage“, „description“ oder „keywords“ gehen sie über das absolute Minimum (der Formalerschließung) hinaus und sollen einer besseren (Sachgebiet-) Zuordnung dienen. Diese Felder wurden vor allem von Suchmaschinen benutzt, um die Suchergebnisse zu verbessern. Seitdem ein Missbrauch dieses Formates durch zahlreiche Webseitenbetreiber eingesetzt hat, ist sein Ansehen stark gesunken. Dennoch werden diese Daten nach wie vor nach einer Kontrolle durch einen Menschen übernommen und vielseitig weiterverwendet.

```
<head profile="http://dublincore.org/documents/dcq-html/">
  <title>Dublin Core</title>
  <link rel="schema.DC"
    href="http://purl.org/dc/elements/1.1/" />
  <link rel="schema.DCTERMS"
    href="http://purl.org/dc/terms/" />
  <meta name="DC.format" scheme="DCTERMS.IMT"
    content="text/html" />
  <meta name="DC.type" scheme="DCTERMS.DCMIType"
    content="Text" />
  <meta name="DC.publisher" content="Jimmy Whales" />
  <meta name="DC.subject" content="Dublin Core Metadaten-
    Elemente, Anwendungen" />
  <meta name="DC.creator" content="Björn G. Kulms" />
  <meta name="DCTERMS.license" scheme="DCTERMS.URI"
    content="http://www.gnu.org/copyleft/fdl.html" />
  <meta name="DCTERMS.rightsHolder" content="Wikimedia
    Foundation Inc." />
  <meta name="DCTERMS.modified" scheme="DCTERMS.W3CDTF"
    content="2006-03-08" />
</head>
```

Text 1: Beispiel für die Verwendung von Dublin Core (mit Terms) in einer HTML-Datei nach http://de.wikipedia.org/wiki/Dublin_Core (Stand: 07.12.2007)

Die grobe und überschaubare Struktur des Dublin Core, die im Vergleich zu ähnlichen Formaten sicherlich zu seinen Stärken zählt, stellt sich im Bereich

der wissenschaftlichen Bibliotheken als ein Nachteil heraus. Zum Beispiel erlauben die Felder „creator“ und „contributor“ keine weitere Typisierung der Inhalte. Dabei ist es aus der bibliographischen Sicht nicht unerheblich ob Autor oder Herausgeber, eine Person oder eine Körperschaft im Feld „creator“ genannt wird (aktuelle DC Fassung¹⁷) oder ob der Herausgeber als einer unter vielen Beteiligten im Feld „contributor“ (Fassung von 1998) in seiner Funktion zu erkennen ist. Gleiches trifft auf „relation“ zu, wo ein Text, ein Link oder eine ID eine andere Ressource nennt, zu welcher das zu beschreibende Objekt in Beziehung steht. Diese Beziehung ist nicht typisiert und kann erst durch die Sacherschließung beider Objekte erkannt werden.

Viel exakter ist die nachträglich entwickelte Erweiterung namens „DC Terms“¹⁸, die zum Beispiel die Beziehungen zwischen den Teilen eines Werkes und dem Gesamten („isPartOf“, „hasPart“) ausdrücken oder der Versionsverwaltung dienen („replaces“, „isReplacedBy“, „created“). Da Terms nur an wenigen Stellen wie "accessRights" in "rights" oder "alternative" in "title" die genannten Schwächen ausgleichen, wird das Dublin Core als Standard für Erfassung von Bibliotheksbeständen für nicht geeignet angesehen¹⁹, obwohl immer mehr Bibliotheken (wie z. B. die Library of Congress) auch in diesem Format bibliographische Daten bereitstellen.

Dieser relativ junge Ansatz, der nur auf einem sehr begrenzten Satz an Erfassungsregeln beruht, brachte einige Neuerungen, die sicherlich im elektronischen Zeitalter von Bedeutung sind. Dazu zählen: die Möglichkeit der "unbegrenzten" Wiederholung einzelner Felder (z. B. Autor, beteiligte

17 Vgl. Dublin Core Metadata Initiative (Hrsg.) : *Dublin Core Metadata Element Set, Version 1.1*. URL: <http://dublincore.org/documents/2010/10/11/dces/> Stand: 11.10.2010

18 Im Gegensatz zu den „elements“ werden „terms“ nicht direkt als Elemente des DC angesehen und bilden einen eigenen Satz an Elementen, sog. „Verfeinerungs-Elementen“, Kodierung-Schematas und Ressource-Typen („vocabulary-term“). Dies wird auch dadurch deutlich, dass sie nicht unter der URI <http://purl.org/dc/elements/1.1/> sondern unter <http://purl.org/dc/terms/> zu finden sind.

19 Vgl. Wiesenmüller, Heidrun : *Informationsaufbereitung I: Formale Erfassung*. In: KSS S. 176

Personen) oder die Vernetzung von Ressourcen ohne weitere Typisierung von Beziehung (z. B. „relation“) oder mit Typisierung (z. B. „isPartOf“).

Neben Dublin Core gibt es andere Formate für die Übertragung von Metadaten. Im Kontext einer wissenschaftlichen Bibliothek sind hier die älteren Formate wie MARC oder MAB aber auch die jüngeren – XML-basierten – wie ONIX oder MODS zu erwähnen. Die starke Ausrichtung an der Praxis hat den Vorteil, dass die entsprechenden Daten genauer und sachgerechter erfasst bzw. ausgetauscht werden können. Solche speziellen Formate sind exakter, erfordern jedoch komplexere Erfassungsregeln und eine darauf ausgerichtete Ausbildung. Zu solchen Regeln gehören die RAK und AACR2.

2.1.2. Regeln für die alphabetische Katalogisierung (RAK) und DIN 1505-1

Mit Regeln zur alphabetischen Katalogisierung (RAK) – vor allem den RAK-WB – soll eine deutsche Norm für die Erfassung bibliographischer Daten dargestellt werden. Da sie jedoch nicht unabhängig von anderen deutschen Normen im selben Bereich ist und zu diesen teilweise im Widerspruch steht, wird am Beispiel von DIN 1505²⁰ sichtbar.

Die RAK erschienen zum ersten Mal 1976 / 1977 und lösten die „Preußischen Instruktionen“ ab. In den darauf folgenden Jahren erschienen mehrere Ausgaben dieses Standardwerkes für verschiedene Bereiche: Musik, Karten, Nichtbuchmaterialien, für Parlaments- und Behördenbibliotheken, für alte Drucke, zur Erschließung von Nachlässen und für die bereits erwähnten Wissenschaftlichen Bibliotheken mit ihrem

²⁰ DIN 1505 besteht aus drei Teilen, wobei jeder Teil einem anderen Problem gewidmet ist: Teil I regelt die Titelaufnahme des Schrifttums, Teil II stellt die Zitierregeln auf. Teil III beschäftigt sich mit der Erstellung von Literaturverzeichnissen. Mit der Bezeichnung „DIN 1505“ wird in dieser Arbeit vor allem Teil II (DIN 1505-2) gemeint. In dieser Arbeit wird die folgende Ausgabe benutzt: Deutsches Institut für Normung (Hrsg.) : *Titelangaben von Dokumenten : Zitierregeln. DIN 1505 Teil 2.* (Verabschiedet im Januar 1984) In: Deutsches Institut für Normung e. V. (Hrsg.) : *Publikation und Dokumentation* 2. 3. Auflage. Berlin : Beuth, 1989 (DIN-Taschenbuch 154). - ISBN 3-410-12343-1 (Wird abgekürzt mit „**DIN-Taschenbuch-154**“)

Vorlage: Johann Wolfgang von Goethe
 Ansetzungsform: Goethe, Johann Wolfgang von

Vorlage: Gallus Anonymus (*Pseudonym*)
 Ansetzungsform: Anonymus <Gallus>

Vorlage: Johann Strauss Sohn
 Ansetzungsform: Strauss, Johann <Sohn>

Vorlage: Caroline von der Heydt (*Britin*)
 Ansetzungsform: VonDerHeydt, Caroline
 Verweisung von: Heydt, Caroline von der
 Verweisung von: DerHeydt, Caroline von

Vorlage: Karoline von der Heydt (*Deutsche*)
 Ansetzungsform: Heydt, Karoline von der

Vorlage: Theodor Meyer zum Gottesberge
 Ansetzungsform: Meyer ZumGottesberge, Theodor

Vorlage: Nikolaj Andreevič Rimskij-Korsakov (*Andreevič =
 Sohn des Andreas*)
 Ansetzungsform: Rimskij-Korsakov, Nikolaj Andreevič

Vorlage: Karl Graf zu Inn- und Knypenhausen
 Ansetzungsform: Inn- und Knypenhausen, Karl zu

Vorlage: Der Apostel Matthäus
 Ansetzungsform: Matthaesus <Apostolus>

Vorlage: Gregor von Nazianz
 Ansetzungsform: Gregor <Nazianzenus>

Vorlage: Kaiser Wilhelm I.
 Ansetzungsform: Wilhelm < Deutsches Reich, Kaiser, I.>

Vorlage: Heinrich XXIV. von Reuß (*alle männlichen Nachkommen
 von Reuß bekommen den Vornamen „Heinrich“*)
 Ansetzungsform: Reuß, Heinrich <XXIV., 1984->

Vorlage: Papst Benedikt XVI. (*Joseph Ratzinger*)
 Ansetzungsform: Benedict <Papa, XVI.>
 Verweisung von: Benedikt <Papst, XVI.>

Text 2: Ansetzungsformen unterschiedlicher Namen und Pseudonyme nach RAK-WB. (Kommentare kursiv dargestellt)

Weitere Anmerkungen: Die russischen Namen der Väter werden im Deutschland meist fehlerhaft als zweiter Vorname verwendet. Aufgrund einer starken Verbreitung dieses „Irrtums“ kann diese Form als richtig angesehen werden. Die Zählung der von Reuß wird jedes Jahrhundert neu begonnen, so dass diese Regel lediglich zusammen mit Lebensdaten eindeutige Identifikation ermöglicht.

Gegenstück für österreichische Bibliotheken²¹. Sie geben die Normen vor, nach welchen der Bestand der Bibliotheken und Archiven einheitlich katalogisiert werden sollen. Dies geschah bis in die 90er Jahre in Form von Karteikarten. An dieser Praxis wurden die Regeln der alphabetischen Katalogisierung (RAK) ausgerichtet. Im digitalen Zeitalter verlieren die strengen Normen für Haupt- und Nebeneintragen oder die Schreibweise von Personennamen an Bedeutung.

In mehreren hundert Paragraphen der RAK-WB²² in der Fassung von 1993 (mit einigen Ergänzungen von 1995, 1996, 1998 und 2002) befinden sich viele Regeln, die dringend einer Überarbeitung bedürfen. So schreibt Thomas Hilberer in einem Aufsatz in der Zeitschrift „Bibliotheksdienst“:

„Die RAK-WB wurden für Zettelkataloge entwickelt und entsprechen nicht den Bedingungen von Datenbanken und Online-Katalogen. Sie beinhalten z. B. ausführliche Vorschriften für die Frage, unter welchem Element des Titels die 'Haupteintragung' zu leisten sei, also wo die 'Haupttitelkarte' in den Zettelkasten einzuordnen und von welchen anderen Elementen wie darauf zu verweisen ist. Das spielt im elektronischen Katalog überhaupt keine Rolle mehr, wo jedes Element der bibliographischen Beschreibung als Sucheinstieg gewählt werden kann. Sind, um noch ein Beispiel aufzuführen, mehr als drei Verfasser an einem Werk beteiligt, so wird nur der erste berücksichtigt, um die Zahl der Zettel klein zu halten. Der Online-Katalog hingegen sollte alle beteiligten Personen aufführen, um möglichst viele Informationen zu bieten.“²³

21 Vgl. Wiesenmüller, Heidrun : *Informationsaufbereitung I : Formale Erfassung* In: KSS. S. 168 f

22 Deutsches Bibliotheksinstitut (Hrsg.) : *Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken*. 2., überarbeitete Ausgabe. Berlin , 1993. - ISBN 3-87068-436-4

23 Hilberer, Thomas : *Aufwand vs. Nutzen : Wie sollen deutsche wissenschaftliche Bibliotheken künftig katalogisieren?* In: *Bibliotheksdienst* 2003, Nr. 6, S. 755

Der Umfang und die Genauigkeit der RAK-WB ist nicht unumstritten. An der Ausbildung und die Tätigkeit der erschießenden Bibliothekare wird ersichtlich, dass auch eine großen Menge an Vorschriften der RAK-WB nicht immer zur Homogenität der erfassten Daten führt, da jeder Bibliotheksverbund eigene – mehrere Ordner umfassende – Auslegungs- und Anwendungsregeln befolgen muss²⁴. Es sind nicht nur die kleinen aber bedeutenden Unterschiede in der Deutung, sondern auch die die Erschließungskosten von ca. 20 € pro Buch²⁵, die die Notwendigkeit eines solchen Regelwerkes in Frage stellen.

Seit Mitte der 90er Jahre erwartete man eine Neuauflage von RAK. Die Hauptanliegen wie: Vereinfachung, Wirtschaftlichkeit, die Fähigkeit, Online-Quellen zu erfassen sowie die Vereinheitlichung (hin zu den international anerkannten Normen) blieben unerfüllt. Im Jahre 2001 sprach sich der Standardisierungsausschuss im sog. Nikolaus-Beschluss für einen Umstieg auf AACR2 aus. Im Rahmen des Projektes „Umstieg auf internationale Formate und Regelwerke (MARC21, AACR2)“ sollte untersucht werden, welche Vorteile oder Nachteile ein Wechsel zu AACR2 mit sich bringt und welche finanziellen Folgen sich daraus ergeben. Im Abschlussbericht des Projektes ist die Projektleitung (DDB) vom Nutzen des Wechsels überzeugt. Sie räumt jedoch ein, dass ein konsequenter Wechsel nur durch Druck von Seiten der Entscheidungsträger gelänge:

„Die Untersuchung zeigt erhebliche Vorteile eines Umstiegs auf. Diese Vorteile lassen sich jedoch in der bibliothekarischen Öffentlichkeit emotional nicht 'verkaufen'. Eine Trendwende könnte nur durch deutliche Meinungsäußerungen seitens der maßgeblichen Förderorganisationen wie DFG und BMBF oder seitens der politischen

24 ebd.

25 ebd. (Angaben aus dem Jahr 2003! Es ist zu vermuten, dass trotz der immer öfter eingesetzter Übernahme der fremder Metadaten diese Kosten vergleichbar hoch geblieben sind, da eine Nachbearbeitung durch das Personal erforderlich ist.)

*Entscheidungsträger, wie sie in der KMK repräsentiert sind, eingeleitet werden.*²⁶

Aus der Sicht vieler Bibliothekare²⁷ ist ein Umstieg jedoch eine Fehlentscheidung:

*„Ein Umstieg auf die AACR2 würde bedeuten, dass man mit hohem Aufwand und ohne wirklichen Nutzen ein überkompliziertes und veraltetes Regelwerk durch ein anderes ersetzen würde, das ebenso schwerfällig und nicht mehr zeitgemäß ist.“*²⁸

Diese Kontroverse zeigt, wie wichtig es ist, im Alltag einfache, eindeutige und aktuelle Regeln anwenden zu können. Dass die RAK kompliziert und zugleich mehrdeutig und verwirrend sein kann, soll im Folgenden anhand von einigen Regeln aufgezeigt werden.

Unter Paragraphen 301 ff. wird in RAK-WB eine Konvention für die Erfassung von Personennamen eingeführt. In allen Werken einer Person soll immer ein und derselbe Name verwendet werden, in der Form, die die Person selber benutzt hat oder die als gebräuchlich gilt, wobei andere Namen bzw. Schreibweisen derer auf diese Form zu verweisen haben. Mit Verweisungen, die z. B. in der Personennamendatei der Deutschen Nationalbibliothek²⁹ festgehalten werden, kann man für die Identität des

26 Die Deutsche Bibliothek (Hrsg.) : *Projekt Umstieg auf internationale Formate und Regelwerke (MARC21, AACR2) – Abschlussbericht*. URL: http://www.ddb.de/standardisierung/pdf/dfg_abschlussbericht.pdf Stand: 21.09.2006

27 Vgl. Siebert, Irmgard : *Positionen zu RAK/AACR. Vortrag auf der Sitzung des DBV, Sektion IV*, Heidelberg, 19.05.2005

28 Hilberer, Thomas : *Aufwand vs. Nutzen : Wie sollen deutsche wissenschaftliche Bibliotheken künftig katalogisieren?* In: *Bibliotheksdienst* 2003, Nr. 6, S. 757

29 „In der PND werden alle für Formal- und Sacherschließung sowie nationale Katalogisierungsunternehmungen wesentlichen Namen zusammengeführt. Sie stellt vor allem eine Dienstleistung zur Vermeidung von Mehrfacharbeit bei der Ansetzung von Personennamen sowie zur Vereinheitlichung der Ansetzungsformen dar. Darüber hinaus ist sie die Bezugsdatei bei der internationalen Kooperation im Bereich der

Autors in allen seinen Werken sorgen. Die Ausnahme, dass mehrere Namen einer Person gebräuchlich sind – wie es z. B. beim Benedikt XVI. und Joseph Ratzinger der Fall ist – wird jedoch nicht gelöst, da hier keine Verknüpfung vorgesehen ist. Oft ist auch die grundsätzliche Forderung nach einheitlicher Verwendung des Namens mit dem Verlust einer wichtigen Information verbunden: Die Worte Joseph Ratzingers als Professor haben nicht denselben Stellenwert, wie die Äußerungen des Benedikts XVI. Die Ansetzungsform des Namens ist für einen Nichtbibliothekar manchmal verwirrend: Der Nachname kommt an erster Stelle vor, der dazugehörige Präfix jedoch in einigen Fällen vor dem Nachnamen und in anderen hinter den Vornamen. Einige Namensbestandteile werden mit weiteren Informationen in spitzen Klammern erfasst. Die richtige Verwendung Solcher Namen kann nur mit Hilfe aufwendiger computerlinguistischer Verfahren oder Kenntnis entsprechender Normen sichergestellt werden³⁰.

Der Grund für einige dieser Regeln liegt im Karteikarten-Katalog, der in der Entstehungszeit des RAK üblich war. Mit dem Einsatz von Datenbanken und damit verbundenen Volltextsuchen gewinnt die stärkere Strukturierung (Unterteilung) der Daten und die Nähe zum orts- und zeitbezogenem Sprachgebrauch gegenüber der Notwendigkeit einer „Identifikation über die Schreibweise“ an Bedeutung. Eine positive Entwicklung, die auf diesem Gebiet verzeichnet werden kann, ist die Zusammenführung von

Ansetzung der Personennamen.“ Aus: Deutsche Nationalbibliothek (Hrsg.): *DNB, Personennamendatei (PND)*. URL: <http://www.d-nb.de/standardisierung/✓normdateien/pnd.htm> Stand: 16.11.2008

30 Zum Problem der Verarbeitung von Personennamen (besonders bei maschinellen Übersetzungen und Cross-Language Information Retrieval) vgl. Womser-Hacker, Christa : *Zur Rolle von Eigennamen im Cross-Language Information Retrieval*. In: Harms, Ilse (Hrsg.) ; Luckhardt, Heinz-Dirk (Hrsg.) ; Giessen Hans W. (Hrsg.) : *Information und Sprache : Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern : Festschrift für Harald H. Zimmermann*. München: K G Saur, 2006. – ISBN-13 978-3-598-11754-1 (dieses Werk wird als "**Festschrift Zimmermann**" abgekürzt verwendet), S. 121 ff.

Personennamendatei (PND) und des Name Authority File (LCNAF) der Library of Congress im Rahmen des VIAF-Projekts³¹.

An die Konvention zur Schreibweise der Personennamen grenzt ein weiteres Problemfeld an – die Angaben zu den Autoren eines Werkes. RAK unterscheidet zwischen Verfasserwerken bzw. gemeinschaftlichen Werken und anonymen Werken. Verfasserwerke und gemeinschaftliche Werke sind Werke mit bis zu drei Verfassern. Alle anderen Fälle werden als anonyme Werke behandelt. Die Urheber solcher Drucke werden meist mit sonstigen beteiligten Personen gleichgesetzt und an dieser Stelle (z. B. USMARC-Felder Nr. 7XX) angegeben. Diese Regel verstößt gegen die Vorgaben der DIN 1505-2 (§ 3.1 und § 3.2)³², die der Vollständigkeit wegen die Nennung aller Autoren und Herausgeber fordert: „Gibt es keine Verfasser, so werden wichtige beteiligte Personen (z. B. Herausgeber) und körperschaftliche Urheber angegeben.“³³ An dieser Stelle es zu erwähnen, dass einige der bibliographischen Formate wie z. B. das von Der Deutschen Bibliothek entwickelte MAB (welches der DIN 1506³⁴ genügen soll) nur drei Felder für Autoren oder Urheber vorsieht. Der MAB und die RAK – ähnlich wie DIN 1505-1³⁵ – verfolgen ein anderes Ziel als DIN 1505-2. Der Verwendungskontext von RAK ist die formale Erfassung – der Bestandsnachweis in (Zettel-)Katalogen. DIN 1505-2 soll hingegen die Identifikation eines Werkes (vor allem im Bezug auf seine Wiederbeschaffung) ermöglichen. Als RAK-WB entstand, war eine automatisierte Verarbeitung

31 Gömpel, Renate ; Frodl, Christine ; Hengel, Christel ; Kutz, Martin ; Münnich, Monika ; Werner, Claudia : *Aus den Veranstaltungen der Division IV Bibliographic Control des Weltkongresses Bibliothek und Information, 69. IFLA-Generalkonferenz in Berlin*. In: *Bibliotheksdienste* 2003, Nr. 10. S. 1239

32 Vgl. *DIN-taschenbuch-154*, S. 93

33 ebd.

34 Deutsches Institut für Normung (Hrsg.) : *Format für den Austausch von bibliographischen Daten. DIN 1506*. (Verabschiedet im März 1978) In: *DIN-Taschenbuch-154*

35 Deutsches Institut für Normung (Hrsg.) : *Titelangaben von Dokumenten : Titelaufnahme von Schrifttum. DIN 1505 Teil 1*. (Verabschiedet im Mai 1984) In: *DIN-Taschenbuch-154*

der auf diese Art und Weise erfassten Daten zu Quellenangaben (DIN 1505-2 und -3) nicht einmal „eine Zukunftsvision“. Es ist zu hoffen, dass die Revision des DIN 1505-2 und -3, welche nach der Verabschiedung einer „Online-fähigen“ Revision des ISO 690³⁶ (Zusammenführung von Teil 1 und 2³⁷) geschehen soll, einen neuen Anstoß für die Erarbeitung eines Nachfolgers von RAK und DIN 1505-1 gibt.

Die formale Erfassung ist im deutschsprachigen Bereich mehr oder weniger einheitlich geregelt. Sie gilt jedoch als nicht eindeutig und veraltet. Ihr Einsatzgebiet – der Karteikartenkatalog – wird kaum noch verwendet und stellt somit die RAK an vielen Stellen in Frage. Eine Diskrepanz zwischen der angenommenen (alphabetischer Katalog) und der tatsächlichen Nutzung (elektronische Kataloge, Quellenangaben) ist die Folge. Ein bibliographisches Informationssystem auf der Grundlage der RAK wäre in elektronischer Form nicht leistungsfähig genug: Mangelnde Homogenität der Daten, ihre schwache Untergliederung und die mangelhafte Vernetzung der bibliographischen Einheiten würde lediglich eine Suche nach bekannten Merkmalen in einer Liste erfasster Werke zulassen.

2.1.3. Anglo-American Cataloguing Rules (AACR2)

Die Anglo-American Cataloguing Rules³⁸ ist die Entsprechung zur RAK im englischsprachigen Bereich. Sie entstand unter internationaler Beteiligung englischsprachiger Bibliotheksverbände wie American Library Association, Canadian Library Association und the Chartered Institute of Library and Information Professionals. Wie der Name schon sagt, orientiert und richtet

36 International Standard Organisation (Hrsg.) : *Information and documentation – Bibliographic references - Content, form and structure. ISO 690*. (Die letzte Revision stammt aus dem Jahre 1987)

37 ISO 690-2 versucht, elektronische Medien „erfassbar“ zu machen. Nach dreizehn Jahren seit der Erstellung werden jedoch immer mehr Schwachstellen (sowie die mangelhafte Verbreitung) dieses Standards sichtbar.

38 AACR besitzt im Gegensatz zu RAK zwei Ausgaben. Die zweite (Second Edition) liegt in der aktuellen Fassung aus dem Jahre 2002 vor. (In dieser Arbeit wird grundsätzlich auf die Arbeits-Teilübersetzung von Margarete Payer <http://www.payer.de/aacr/> ✓ [aacr00.htm](http://www.payer.de/aacr00.htm) ff., Stand 24.2.1996 / 13.08.2011, Bezug genommen.)

sich dieses Regelwerk am Bedarf und den Gepflogenheiten dieser Länder. Dennoch wird er auch in anderen Ländern wie z. B. in der Schweiz (die Schweizerische Landesbibliothek in Bern und Informationsverbund Deutschschweiz)³⁹ und dreißig weiteren Ländern – adaptiert oder unverändert – angewendet.

AACR2 orientiert sich stärker an ISBD⁴⁰ und lässt dem Katalogisierer mehr Interpretationsfreiheit als die RAK. Diese Tatsache macht es möglich, den Katalog stärker auf die Benutzer hin auszurichten. So werden Personen nicht unter dem von ihnen verwendeten Namen, sondern unter dem Namen erfasst, welcher den Katalogbenutzern geläufig ist. Die Benutzerfreundlichkeit⁴¹ ist jedoch mit dem Nachteil verbunden, dass der Katalog auf einen bestimmten Kreis von Personen ausgerichtet ist. Die sehr offen gehaltenen Regeln machen es nötig, lokale Interpretationsvorschriften zu erarbeiten. Dies bedeutet auch einen stärkeren Verlust der Homogenität der erfassten Daten unter einzelnen Bibliotheken bzw. Bibliotheksverbänden wie es bei RAK-WB der Fall ist.

Obwohl die Regeln des AACR2 nach Ansicht einiger Bibliothekare nicht weniger komplex und umfangreich sind als die der RAK⁴², bringen sie dennoch für den ungeübten Katalogisierer (in unserem Fall: ein Wissenschaftler oder ein Studierender) einige Vorteile. So ist es in AACR2 erlaubt „ein Buch nach ISBD [zu] beschreiben, ohne auf Eintragungs- und Ansetzungsregeln zu achten“⁴³. Die Personennamen werden in der

39 Vgl. Wiesenmüller, Heidrun : *Informationsaufbereitung I : Formale Erfassung* In: KSS. S. 168

40 International Standard Bibliographic Description (ISBD) ist ein internationaler Standard für die Beschreibung von Werken wie Monographien (ISBD-M) oder kartographischem Material (ISBD-CM). Mit Hilfe von wenigen Deskriptionszeichen werden darin die grundlegenden bibliographischen Merkmale eines Werkes beschrieben.

41 Gemeint ist im engeren Sinne die „Lokalisierung“ (Orientierung am ortsbezogenen Sprachgebrauch), die ein Teil der Usability darstellt.

42 Vgl. Hilberer, Thomas : *Aufwand vs. Nutzen : Wie sollen deutsche wissenschaftliche Bibliotheken künftig katalogisieren?* In: Bibliotheksdienst 2003, Nr. 6, S. 756

43 Payer, Margarete : *AACR2 - RAK : grundsätzliche Unterschiede.* URL:

Landessprache (als Haupteintrag) angegeben z. B.: „Benedikt XVI., Papst“ statt „Benedict <Papa, XVI.>“ (RAK). Die Namen aus nichtlateinischen Sprachen werden dabei nach ALA/LC⁴⁴ transliteriert. Für den deutschen Sprachbereich müssten die entsprechenden DIN-Normen angewendet werden. Bei Körperschaften, die oft als Urheber eines Werkes gelten, sind nur spezifische Namen erlaubt, der Begriff der Körperschaft jedoch weiter definiert als bei RAK.

Die Interpretationsfreiheit bei der Ansetzungsform sowie die Umschrift auf der Ebene der Benutzersprache verringert die Anzahl der Regeln, hat aber den Nachteil, dass die erfassten Daten weniger homogen sind und das „automatisierte Zitieren“ oder das Retrieval deutlich erschweren. Auf einer solchen Grundlage ist somit ein leistungsstarkes bibliographisches Informationssystem undenkbar.

Die AACR2 und RAK-WB haben trotz vieler Unterschiede vieles gemeinsam: Beide wurden für Karteikatenkataloge entwickelt; beiden fehlt die Möglichkeit, Personennamen oder Titel von Werken in der Originalschreibweise zu führen; die Ansetzungsformen können nur unzureichend von Computern verstanden und verarbeitet werden. An dieser Stelle wird die Kritik vieler Betroffenen an einem Umstieg von RAK-WB auf AACR2 z. B. als „Verschwendung von Steuergeldern“⁴⁵ nachvollziehbar.

2.1.4. Transkription und Transliteration

Mit Transkription und Transliteration gibt es zwei Vorgehensweisen einer Umschrift zwischen zwei Sprachen bzw. Zeichensätzen. Meyers Lexikon Online definiert **Transkription** wie folgt:

<http://www.payer.de/einzel/aacr2rak.htm> Stand: 09.07.2006

44 Die ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts wurden von der Library of Congress und der American Library Association erarbeitet. Mit Hilfe dieser Tabellen kann man nichtlateinische Sprachen ins Englische transliterieren. Die deutsche Entsprechung für die ALA/LC- Romanization Tables sind die DIN-Normen.

45 Vgl. *Meinungsäußerungen zur AACR/MARC-Frage* URL: <http://www.allegro-c.de/formate/umfr-3.htm> Stand: 27.09.2006

„Umsetzung einer Schrift (z. B. kyrillisch) in eine andere (z. B. lateinisch) für die Sprecher einer bestimmten Sprache, also in für diese lautlich in etwa entsprechende Buchstaben(kombinationen). Transkription ist also im Gegensatz zur Transliteration stets an eine bestimmte Sprache gebunden.“⁴⁶

Die Transliteration wird von derselben Quelle mit folgenden Worten erklärt:

*„**Transliteration** [lateinisch] die, buchstabentreue Umsetzung eines in einer Buchstabenschrift (z. B. hebräische, griechische Schrift) geschriebenen Textes in eine andere Buchstabenschrift (z. B. lateinische Schrift), eventuell unter Verwendung diakritischer Zeichen; der Text kann mithilfe einer Transliterationstabelle korrekt in die Originalschrift zurückübertragen werden.“⁴⁷*

Der Unterschied zwischen Transkription und Transliteration besteht darin, dass die Transkription um die Beibehaltung der originalen Aussprache der Vorlage bemüht ist, während Transliteration die Umsetzung von Zeichen eines in der Zielsprache nicht abzubildenden Zeichensatzes realisiert. Je nach Verwendungskontext kann die Transkription oder die Transliteration von Vorteil sein. Im Falle von ANSEL, dem USMARC Zeichensatz der Library of Congress, wird die Transliteration dazu verwendet, um z. B. „polnische Buchstaben“ wie *ą*, *ę*, oder *ł* mit Hilfe von ASCII-Zeichensatz abzubilden. Bei Personennamen hat sich die Transkription als geeignet erwiesen, da sie die Möglichkeit bietet, einen ursprünglich schwer lesbaren Namen mehr oder weniger richtig auszusprechen. Sie wird jedoch nicht auf alle Namen angewandt. So werden z. B. französische Namen niemals ins Deutsche

46 Bibliographisches Institut & F. A. Brockhaus AG (Hrsg.) : *Transkription*. In: *Meyers Lexikon online*. Stand: 25.08.2006. URL: <http://lexikon.meyers.de/index.php?title=✓Transkription&oldid=54765> (zuletzt online: 23.03.2009)

47 Bibliographisches Institut & F. A. Brockhaus AG (Hrsg.) : *Transliteration*. In: *Meyers Lexikon online*. Stand: 25.08.2006. URL: <http://lexikon.meyers.de/index.php?title=✓Transliteration&oldid=54768> (zuletzt online: 23.03.2009)

transkribiert, obwohl sie von vielen deutschen Bürgern nicht richtig ausgesprochen werden. Dies gründet wahrscheinlich in der Geschichte: Französisch galt als Diplomatensprache und somit als international. Eine weitere Folge davon ist, dass Russen oder Ukrainer in ihrem Reisepass die französische Transkription ihres Namens haben, während eine Übersetzung eines Werkes dieser Personen ins Englische unter einem anders geschriebenen (transkribierten) Namen veröffentlicht wurden. Die Bezeichnungen aus dem Kyrillischen können in Deutschland entweder **nach DIN 1460 transliteriert** werden (z. B. „Galaza“ = de. Augen) oder **nach Duden transkribiert** werden („Galasa“ = de. Augen). Stamm sie jedoch aus der Ukraine, wäre die Transkription „Halasa“ angebracht, da im Ukrainischen ein „G“ wie ein hartes „H“ ausgesprochen wird.

Die Transkription weist gegenüber der internationalen Transliteration einige Nachteile mit sich. Unterschiedliche Umschriften desselben Namens führen zu Problemen beim Retrieval. Eine Verlinkung mehrerer Schreibweisen desselben Namens mit der originalen Schreibweise könnte sich unter Verwendung von Unicode und weiterer Merkmale als der Königsweg erweisen, zumal es für Philologen von Vorteil wäre, auch nichtlateinische Zeichen in der Suche oder in den Quellenangaben einsetzen zu können.

2.2. Weitere Normen, Begriffe und Definitionen

Im Unterschied zu Zettelkatalogen, die immer in derselben physikalischen Form erstellt werden, können elektronische Kataloge sehr unterschiedlich realisiert werden. Die bibliographischen Daten können als binäre Datenbanken oder als Textdateien abgespeichert werden. Sie können entweder nur lokal (auf einem PC) eingegeben, gesucht und verarbeitet werden oder mit anderen Computern ausgetauscht werden (Client-Server-Modell), wobei die Suche in Form eines bestimmten Suchbefehls auf dem Client generiert, zum Server geschickt und dort ausgeführt wird, um die gefundenen Datensätze vom Server zu übernehmen. (Ein P2P-Modell kann bisher nur durch den Umweg des Exports und Import bestimmter

Datensätze, der von den Benutzern ausgeführt werden muss, realisiert werden.) Dabei spielen Datenformate als Form der lokalen Speicherung und des Austausches eine sehr wichtige Rolle. Es wird sichtbar, wie wichtig fundierte Kenntnisse der EDV für die Bibliothek (bzw. für deren elektronischen Katalog) sein können. Deshalb sollen an dieser Stelle neben den bibliographischen Begriffen auch einige relevante Begriffe aus der Informationstechnik, die die Erfassung und Speicherung der Metadaten betreffen, kurz dargestellt werden.

Bezeichnungen, die z. B. der Informatik entliehen wurden, aber für das Verständnis der folgenden Abschnitte sehr wichtig sind, können im Glossar im Anhang gefunden werden.

2.2.1. DIN- und ISO-Normen im bibliographischen Bereich

Die Normung spiegelt die Idee wieder, verbindliche Regeln zu schaffen und die Vereinheitlichung der Maßeinheiten und Vorgehensweisen zu fördern.

Im bibliographischen Bereich spielen nationale und internationale Normen eine große Rolle. Mit DIN 1505 und 1506 wurden bereits zwei nationale Normen erwähnt. Neben den Normen, die einheitliche Regelungen für die Tätigkeit der Bibliotheken darstellen, gibt es weitere Normen aus dem linguistischen Bereich, die z. B. während der Katalogisierung zu beachten sind. Zu den spezifischen Normen der Bibliotheken zählen⁴⁸:

- DIN 1426 – Inhaltsangaben von Dokumenten; Kurzreferate, Literaturberichte (*entspricht ISO 214*)
- DIN 1461 – Lochung in Katalogkarten und zugehörige Schließstangen (*wird kaum noch verwendet*)
- DIN 1463-2 bzw. ISO 6964 – Erstellung und Weiterentwicklung von Thesauri; Mehrsprachige Thesauri
- DIN 1464 – Loseblattausgaben (-werke); Ergänzungslieferungen, Form und Einordnung

48 Titelangaben zu ISO-Normen nach: Herzog Gottfried ; Wiesner Hans-Jörg : *Normung*. In: KSS. S. 629-633 (teilweise erweitert)

-
- DIN 1505-1 – Titelangaben von Dokumenten; Titelaufnahme von Schrifttum
 - DIN 1505-4 – Titelangaben von Dokumenten; Titelaufnahme von audio-visuellen Medien
 - DIN 1506 – Format für den Austausch von bibliographischen Daten
 - DIN 31638 – Bibliographische Ordnungsregeln
 - DIN 32705 – Information und Dokumentation – Anforderungen an die Aufbewahrung von Archiv- und Bibliotheksgut
 - DIN EN⁴⁹ ISO 2789 – Information und Dokumentation – Internationale Bibliotheksstatistik
 - ISO 18 – Dokumentation; Inhaltsverzeichnis von Zeitschriften und anderen Veröffentlichungen
 - DIN ISO 2108 – Information und Dokumentation – Internationale Standard-Buchnummer (ISBN)
 - DIN ISO 11620 – Information und Dokumentation – Leistungsindikatoren für Bibliotheken
 - DIN ISO 15489 – Information und Dokumentation – Schriftgutverwaltung
 - ISO 832 – Information und Dokumentation – Bibliographische Beschreibung und bibliographische Nachweise – Regeln für die Abkürzung von bibliographischen Bezeichnungen
 - ISO 3297 – Information und Dokumentation – Internationale Standardnummer für fortlaufende Sammelwerke (ISSN)
 - ISO 6630 – Dokumentation; Steuerzeichenvorrat für bibliographische Daten
 - ISO 7275 – Dokumentation; Präsentation von Titeln von Serien

49 EN steht für European Norm (de. Europäische Norm)

- ISO 23950 - Information und Dokumentation – Wiederauffinden von Informationen (Z39.50) – Festlegung des Anwendungsdienstes und Beschreibung des Protokolls (*hervorgegangen aus dem ANSI/NISO Z39.50*)
- ISO 10324 – Information und Dokumentation – Zusammenfassende Bestandsangaben in Katalogen

Je nach Ausrichtung der Bibliothek können weitere DIN oder ISO Normen greifen wie z. B. DIN ISO 10444 (Internationale Nummer für Forschungsberichte), DIN ISO 10957 (Internationale Nummer für Musikalien) u. Ä.. Für deutsche Bibliotheken gelten natürlich die DIN- und – wenn diese fehlen – die ISO-Normen. Für die Umsetzung der internationalen in nationale Normen im Bereich des Informations- und Dokumentationswesens ist der Normenausschuss Bibliotheks- und Dokumentationswesen (NABD) zuständig.

Weitere Normen, die zwar nicht direkt Bibliotheken und ihre Tätigkeit zum Gegenstand haben aber dort zur Geltung kommen, stammen aus dem allgemeinen Bereich der Informationsverarbeitung, besonders der Verarbeitung von Sprache. So gelten bei der Erfassung und Verarbeitung von bibliographischen Daten folgende Normen (nach Themenbereichen gegliedert):

- Transkription und Transliteration: ISO 9 bzw. DIN 1460 für kyrillische Buchstaben; ISO 843 bzw. DIN 31634 für griechische Zeichen; ISO 233 bzw. DIN 31635 für arabische Buchstaben; ISO 259 bzw. DIN 31636 für die Umschrift der hebräischen Buchstaben; ISO 3602 – Transkription des Japanischen; ISO 7098 – Transkription des Chinesischen; ISO 11941 für Transliteration der koreanischen Schrift [u. v. a. m.]
- Abkürzungen und Codes: DIN EN ISO 3166 – Codes für die Namen von Ländern und deren Untereinheiten; ISO 4 – Information und Dokumentation – Regeln für das Kürzen von Wörtern in Titeln und für

das Kürzen der Titel von Veröffentlichungen; ISO 639 – Codes für Sprachnamen.

- Indexierung und Informationsaustausch: DIN 5007 – Ordnen von Schriftzeichenfolgen; DIN 31630-1 – Registererstellung; Begriffe, Formale Gestaltung von gedruckten Registern; DIN 32705 – Klassifikationssysteme; Erstellung und Weiterentwicklung von Klassifikationssystemen; DIN 31623 – Indexierung zur inhaltlichen Erschließung von Dokumenten; DIN 31631 – Kategorienkatalog für Dokumente; ISO 5963 – Dokumentation; Methoden zur Analyse von Dokumenten, Bestimmung ihres Inhaltes und Selektion von Indexier-Benennungen.

Es ist zu beachten, dass ISO-Normen durch gleichwertige Normen des DIN ersetzt werden können. Bei einigen Normen wie DIN 1506 handelt es sich um noch verwendete, aber nicht mehr zeitgemäße Regelungen. Der Fortschritt der Technik und die damit einhergehende Entwicklung neuer Standards wie des **XML** (eine Anwendung der SGML⁵⁰ von hoher Akzeptanz im Software-Bereich) lässt vermuten, dass einige ältere Normen langfristig den neuen weichen werden.

2.2.2. Datenformate

Wenn man von elektronischer Datenverarbeitung spricht, meint man eine Verarbeitung von Eingaben (in Form von Daten) durch einen Computer⁵¹. Eingaben, die nur aus Text bestehen, konnten mit nur wenigen unsichtbaren

50 Standard Generalized Markup Language wurde als ISO 8879 verabschiedet und dient als Grundlage für viele verschiedene Dateiformate wie z. B. TeX oder XML.

51 Diese Eingaben werden in binärer Form kodiert und in achtstelligen Gruppen (sog. Bytes) verarbeitet. In einem Byte gibt es also 256 mögliche Variationen. Jede von ihnen kann dazu genutzt werden, eine Zahl oder einen Buchstaben zu repräsentieren. Außer den für Menschen sichtbaren Zeichen, gibt es einige sog. Steuerzeichen. Einige von ihnen stellen den Zeilenumbruch (meist als NL [new line] und CR [carriage return] kodiert) oder Einrückung (Tabulator) dar. Einige haben nur für Rechner eine Bedeutung. Vgl. Bekavac, Bernard: *Information und Kommunikationstechnologien*. In: *KSS*, S. 326 ff

Zeichen wie Zeilenumbruch, Spatium und Tabulator formatiert und abgespeichert werden. Auf diese Art und Weise abgespeicherten Daten nennt man deshalb **Text-basierte Datenformate**. Andere Eingaben brauchen auch weitere nichtdruckbare Bit-Folgen, um zusätzliche Daten abspeichern zu können⁵². Die Bedeutung von Zeilenumbrüchen oder Tabulatoren kann je nach Zusammenhang variieren. Dadurch sind sie auf die Interpretation durch geeignete Editoren angewiesen. Diese Art von Daten nennt man **binäre Datenformate**.

Die binären Datenformate haben gegenüber den reinen Textdateien den Vorteil, dass sie die Daten mit einigen Steuerzeichen interpretieren. Vor allem Datenbanken machen sich dieses Merkmal zunutze. Mit Einzug der **Auszeichnungssprachen**, die meistens auf der Basis des SGML entwickelt wurden, konnten auch reine Textdateien mit Steueranweisungen angereichert werden. Ein sehr altes Beispiel dafür ist das TeX-Format. Darin wurde das sog. „Backslash“ („\“) sowie die eckigen und geschweiften Klammern dazu verwendet, eine Verarbeitungsanweisung für eine bestimmte Textstellen zu setzen. Ein anderes Beispiel für die Popularität einer solchen Auszeichnungssprache sind das HTML, das XHTML und das XML. Mit XML fand ein relativ einfacher Satz von Syntax-Regeln ein sehr positives Echo unter Programmierern. Auf seiner Basis entstanden viele sehr populäre Dateiformate wie RDF, XHTML oder das OASIS Open Document Format for Office Applications⁵³ (Kurzform: OpenDocument) von Open Office und Star Office.

Die Auszeichnungssprachen erweitern die Textdateien um die Typisierung der Daten, so dass sie für Menschen wie Computer verständlich sind und zugleich von Rechnern interpretiert werden können. Für die Bearbeitung dieser Dateien reicht ein üblicher Texteditor aus – was die Gemeinsamkeit

52 Wikimedia Foundation Inc. (Hrsg.) : *Binärformat*. URL: <http://de.wikipedia.org/w/index.php?title=Bin%C3%A4rformat&oldid=6801306>

53 Das von OASIS spezifizierte OpenDocument wurde als ISO-Norm ISO/IEC DIS 26300 verabschiedet. Diese Entwicklung bewegte einige Staaten dazu, es als einzig zulässiges Austauschformat in Ämtern einzusetzen.

mit allen textbasierten Datenformaten darstellt. Die zu vergebenden Auszeichnungen werden in der Document Type Definition⁵⁴ oder dem sog. Schema festgelegt. Jede Abweichung von diesen Anwendungsregeln kann zu einer ungültigen (für Computer unverständlichen) Datei führen. Dieses Problem haben die Auszeichnungssprachen mit den meisten binären und einigen textbasierten Datenformaten gemeinsam.

Die binären Datenformate werden oft dort bevorzugt, wo es um die Verarbeitung großer Datenmengen geht, da die Steueranweisungen kürzer und dadurch schneller zu verarbeiten sind als lange und oft verschachtelte Auszeichnungen in Textform. Diese Tatsache bedeutet eine Ersparnis von Speicherplatz und eine Verkürzung der Verarbeitungszeit – die sog. Performance. Beides spielt angesichts der Leistung und der Größe der Speichermedien heutiger Rechner bei kleinen und mittelgroßen bibliographischen Datenbeständen jedoch kaum eine Rolle.

Im Folgenden sollen einige gängige Datenformate vorgestellt werden, die der Speicherung und Übermittlung von bibliographischen Daten dienen. Einen Sonderfall stellt hier das XML dar, da es nicht zweckgebunden ist, aber für Metadaten wie Dublin Core verwendet werden kann.

2.2.2.1. Textbasierte Datenformate

2.2.2.1.1. MARC und MAB

Mit **Machine-Readable Cataloging (MARC)** und dem **Maschinellen Austauschformat für Bibliotheken (MAB)** existieren zwei Text-basierte Datenformate für den bibliographischen Bereich. MARC wurde unter der Federführung der Library of Congress entwickelt und implementiert den ISO 2709-Standard „Format für Informationsaustausch“ sowie sein amerikanisches Gegenstück ANSI/NISO Z39.2. MARC wurde in einigen

54 „Die Dokumenttyp-Definition (DTD) ist die Definition der erlaubten Elemente und Attribute für eine Klasse von Dokumenten in SGML oder XML.“ Aus: DTD. In: KSS-Glossar, S. 31

Ländern erweitert und an die lokalen Gegebenheiten angepasst (USMARC, UK-MARC etc.)⁵⁵.

In Deutschland wurde von Der Deutschen Bibliothek ein ähnliches, wenn auch nicht MARC-kompatibles, Format entwickelt. Das Maschinellen Austauschformat für Bibliotheken (MAB), welches auf DIN 1506 (entspricht ISO 2709) basiert und denselben Zweck verfolgt. Die neuste Version des MAB-Formats stammt von 1995 und wird als MAB2 bezeichnet. MAB2 besteht aus mehreren Teilformaten wie MAB-Titel (Format für bibliographische Daten), MAB-PND (Format für Personennamen) etc.

Beide Formate nutzen außer druckbaren Zeichen nur Zeilenumbruch und Tabulator (sowie einige wenige andere nicht druckbare Zeichen für die Darstellung fremder Zeichensätze) – damit zählen sie zu den textbasierten Datenformaten. Einigen Zeichen kommt eine besondere Bedeutung zu: In MARC wird das Dollarzeichen „\$“ zum Maskieren von Teilfeldern verwendet. Da sowohl MARC als auch MAB für den Austausch von Daten zwischen Bibliotheken gedacht sind, welches zuerst durch das Überspielen von Magnetbändern und erst später durch das Internet geschah⁵⁶, waren die Prüfsummen und feste Record-Längen (mit Spatium oder Sonderzeichen aufgefüllt) nötig. Eine Folge dieser Zielsetzung ist das für den Menschen kaum lesbare⁵⁷ Format, welches aus einem Vorspann und dem Inhalt besteht. Im Vorspann jedes Datensatzes sind Prüfsummen (z. B. die Länge des Satzes) und einige einleitende Steueranweisungen enthalten, die sicherstellen sollen, dass der Inhalt unbeschädigt ist und richtig interpretiert wird. Der Inhalt besteht aus mehreren Zeilen, die pro Zeile ein Feld enthalten. Der Typ des Feldes wird mit einer Zahl (1-3 Zeichen bzw. 3

55 Vgl. dazu: Wikimedia Foundation Inc. (Hrsg.) : *Machine-Readable Catalog*. URL: http://de.wikipedia.org/w/index.php?title=Machine-Readable_Catalog&oldid=21799071

56 Die mit Hilfe von Magnetbändern oder über das Internet übermittelte Daten sind oft keine Dateien, sondern Datenströme (eng. Streams). Aus diesem Grund wird in dieser Arbeit der Begriff „Datenformat“ anstatt „Dateiformat“ verwendet.

57 Die „Lesbarkeit“ schließt in diesem Sinne „das Verstehen des Gelesenen“ ein. Das Verstehen der Inhalte eines Records ist z. B. bei der Fehlersuche notwendig. Da es sich hierbei um eine Ausnahme handelt, ist diese Eigenschaft nicht von Bedeutung.

Hauptsatz:

```

### 00638nM2.01000024 h
001 00005549183
002a19870612
003 20000623
004 20020110160902.0
006n3
030 a|zzazzz|||17
050 a|a
051 n
100 Vollmer, Gerhard
331 Was können wir wissen?
359 Gerhard Vollmer
410 Stuttgart
412 Hirzel
700tNAT 250f
900 0000007084x
902 |Erkenntnistheorie
904a036?040?057
905 0000009997x
907 |Naturphilosophie
909a040
910 00009601066
912 |Evolutionäre Erkenntnistheorie
912 |Aufsatzsammlung
914a041?013?083
915 00007729502
917 |Evolutionäre Erkenntnistheorie

```

Untersätze:

```

### 00279nM2.01000024 u
001 00083031170
002a19931005
004 20020110161636.0
006n0
010 00079658229
...
089 1. Die Natur der Erkenntnis : Beitr. zur
    evolutionären Erkenntnistheorie. - XXV, 337 S.
090 1

```

```

### 00273nM2.01000024 u

```

```

...
089 2. Die Erkenntnis der Natur : Beitr. zur modernen
    Naturphilosophie. - XXII, 305 S.
090 2

```

Text 3: Beispiel für MAB-Format: ein mehrbändiges Buch nach <http://www.allegro-c.de/formate/examp/examp3.htm> (Stand: 2.1.2011)

Zeichen mit einleitenden Nullen) eingeleitet. Sie gibt vor, um welchen Typ es sich handelt: Autor, Titel, Verlag o. Ä. Dieser Zahl folgt ein Trennzeichen und Bezeichnung des Teilfeldes.

```
00842nam 2200253 a 4500
001 11914984
003 CU-UC
005 20020110002535.1
008 970718s1998 nyu b 001 0 eng
010 $a 97033271
020 $a0387946470 (alk. paper)
050 00$aQA76.9.C66$bC73 1998
082 00$a303.48/34$d221
100 1 $aCrandall, Richard E.,$d1947-
245 12$aA network orange :$blogic and responsibility in the
    Computer Age /$cRichard Crandall, Marvin Levich ; with
    foreword by Howard Rheingold.
260 $aNew York :$bCopernicus, $cc1998.
300 $axvi, 130 p. ;$c25 cm.
504 $aIncludes bibliographical references and index.
650 0 $aComputers and civilization
650 0 $aComputers$xMoral and ethical aspects
700 1 $aLevich, Marvin
```

Text 4: Beispiel für MARC21-Format: ein Buch mit zwei Autoren nach
<http://www.allegro-c.de/formate/examp/examp2.htm> (Stand: 12.10.2006)

An dieser Stelle gehen die Konzepte von MARC und MAB auseinander. Während bei MARC nur Teilfelder (Maskiert durch ein Sonderzeichen z. B. „\$“ und einen Buchstaben, der den genaueren Typ vorgibt) folgen, sieht das MAB an dieser Stelle mehrere aufeinander bezogene Felder vor (100=“Name der 1. Person“, 102=“Verweisungsformen zur 1. Person“, 103=“Körperschaft, bei der die erste Person beschäftigt ist“), die untergliedert bzw. typisiert werden dürfen („vollständiger Name der Person“ oder ihr Pseudonym).

Ein weiterer Unterschied besteht in der Wiederholung von Feldern: MARC erlaubt die Wiederholung einiger Felder (z. B. Autor), MAB besitzt für den Autor und die Körperschaft drei gleichwertige Felder mit unterschiedlicher Nummerierung⁵⁸.

58 Eine genaue Auskunft über Nummerierung gibt <http://www.allegro-c.de/formate/mab.htm> (Stand: 2.1.2011). Die offizielle Dokumentation des MAB-Titel-Formats befindet sich unter: <http://www.d-nb.de/standardisierung/txt/titelmab.txt> (Stand:

Band 1:

```

100 10 $aVollmer, Gerhard, $d1943-
245 10 $aWas können wir wissen? / $cGerhard Vollmer.
260 0 $aStuttgart : $bHirzel, $c1985-
300 $av. : $bill. ; $c21 cm.
504 $aBibliography: v.1, p. [324]-327.
505 1 $aBd. 1. Die Natur der Erkenntnis : Beiträge zur
evolutionären Erkenntnistheorie / mit einem Geleitwort von
Konrad Lorenz.
650 0 $aKnowledge, Theory of.
650 0 $aThought and thinking.
650 0 $aHuman evolution.
700 10 $aLorenz, Konrad, $d1903-

```

Band 2:

```

100 1 $cVollmer, Gerhard, $d1943- <
245 10 $cWas können wir wissen? /$cGerhard Vollmer.
250 $c2., durchgesehene Aufl.
260 $cStuttgart : $bS. Hirzel, $c1988-
300 $cv. : $b ill. ; $c21 cm.
504 $cIncludes bibliographical references.
500 $cIncludes indexes.
505 1 $cBd. 1. Die Natur der Erkenntnis -- Bd. 2. Die
Erkenntnis der Natur.
650 0 $cKnowledge, Theory of.
650 0 $cPhilosophy.

```

Text 5: Beispiel für MARC21-Format: ein mehrbändiges Buch nach
<http://www.allegro-c.de/formate/examp/examp3.htm> (Stand: 12.10.2006)

Anders als MARC ist MAB hierarchisch organisiert und kann ein mehrbändiges Werk mit Hilfe von Datensatztypen (Hauptsatz, Untersatz oder Nachsatz) einander zuordnen und so einzelne Bände aufeinander beziehen, ohne Angaben zum Autor wiederholen zu müssen. Dieses Vorgehen vermindert Redundanzen und sorgt für Einheitlichkeit der Angaben und Vernetzung von Datensätzen. (UNI)MARC bietet eine Reihe von Feldern, die der Verlinkung der Datensätze untereinander dienen (sog. "linking entry block", der mit einer "4" anfängt) z.B.: 430 = Fortsetzung von, 440 = wird fortgesetzt durch, 454 = Übersetzung von, 453 = übersetzt als. In Unterfeldern kann man den Titel, den Identifikator (der in selber Datei

vorhanden sein muss) oder ein anderes Merkmal als Ziel angegeben werden.

Beide Formate wurden von großen nationalen Bibliotheken entwickelt. Sie verfügen deshalb über eine „Ausdruckstärke“, die vielen anderen Formaten für bibliographische- oder Metadaten fehlt. Diese Vielfalt an Datenfeldern ist mit hoher Komplexität erkauft. Die Schreibweise der Namen (die Ansetzungsform) hängt immer von den lokal angewandten Regelwerken ab, was ein unkontrolliertes Übernehmen fremder Datensätze ohne Verlust der Qualität der Daten ausschließt⁵⁹. Die Verwendung bestimmter Zeichen als Steuerzeichen macht den Umstieg auf Unicode unter Umständen problematisch, so dass die ursprüngliche Schreibweise der Namen nicht mitgeliefert werden kann. Einige Inhalte werden in der Sprache des Erfassenden (bzw. der Bibliothek, die erfassen lässt) mangels spezieller Datenfelder eingegeben, wodurch diese Daten an eine Sprache gebunden bleiben („pages“ bzw. „p.“ statt „Seiten“ bzw. „S.“, Stichwörter oder Anmerkungen). Auch die Redundanz im MARC-Format stellt unter Umständen ein Problem dar, da Unterschiede in der Schreibweise das Retrieval erschweren. Die Begrenzung der Autoren- bzw. Körperschaftenzahl auf drei entspricht zwar den RAK, begrenzt das MAB jedoch auf den Austausch der Katalogdaten und ist somit für automatische Generierung von Quellenangaben nicht geeignet. Diesbezüglich ist die Struktur der beiden Formate insgesamt etwas zu grob, um sie zu jedem beliebigen Format verarbeiten zu können.

Eines darf man bei allen Schwächen beider Formate nicht vergessen: es sind Formate, die in ihrem Verwendungskontext hohe Akzeptanz genießen und an der internen Praxis der Bibliotheken ausgerichtet sind. Ihren (begrenzten) Zweck erfüllen sie seit dreißig Jahren. Man könnte sie als den ersten Schritt

59 Vgl. auch Abschnitt: „1:1-Konvertierung unmöglich“ (URL: <http://www.allegro-c.de/formate/formneu.htm> Stand: 12.10.2006) In: Eversberg, B.: *Was sind und was sollen Bibliothekarische Datenformate?*. Braunschweig : Univ.-Bibliothek der TU, 1994. - ISBN 3-927115-21-5. Als Online-Version mit Ergänzungen (2003) URL: <http://www.allegro-c.de/formate/> Stand: 2.1.2011

in Richtung elektronische Datenverarbeitung im bibliographischen Bereich bezeichnen. Zusammen mit dem Protokoll Z39.50 können sie als das erste professionelle und Betriebssystem-unabhängige Retrieval-System des Computerzeitalters erachtet werden. Die Entwicklungen wie MAB-NOTAT oder MAB-PND weisen in die "richtige" Richtung: Auslieferung möglichst vieler Zusatzinformationen und die Vernetzung bzw. Gruppierung von Werken durch Klassifikation.

2.2.2.1.2. RIS

Research Information System als ein alternatives Format für bibliographische Daten steht hier stellvertretend für mehrere einfache Text-basierte Formate wie PubMed/MedLine, ISI oder Refer. Alle diese Formate entstanden in den 90er Jahren und zeichnen sich durch einen einfachen Aufbau aus. Die gemeinsame Grundidee besteht darin, einzelne Datenfelder eines Datensatzes durch Zeilenumbruch zu trennen und mit einer Kürzel am Anfang eines jeden Datensatzes die darauf folgenden Daten einem bestimmten Typ zuzuordnen. Die Kürzel werden in einigen Formaten durch ein Sonderzeichen eingeleitet und durch eine bestimmte Zeichensequenz wie Spatium-Bindestrich-Spatium oder ein einzelnes Spatium von den Daten getrennt.

Das RIS-Format hat unter den oben genannten Formaten (PubMed ausgenommen) die größte Ausdruckstärke⁶⁰. Jeder Datensatz wird unter Angabe des Typs (Kürzel „TY“) eingeleitet. RIS kennt 35 Ressource-Typen, zu denen sowohl Bücher oder Abstracts als auch Landkarten, Videos gehören. Jeder Datensatz endet mit der Kürzel 'ER - '. Die Reihenfolge anderer Felder ist nicht vorgegeben. Die Felder dürfen wiederholt werden. Für die Angabe mehrerer Autoren gibt es je nach Zusammenhang zwischen der Person und dem Werk verschiedene Datenfelder: „AU“ oder „A1“ für den ersten bzw. wichtigsten Autor, „A2“ oder „ED“ meint den Herausgeber bzw. den zweiten Autor (en. „author secondary“) und „A3“ bezeichnet sonstige

⁶⁰ Das Handbuch des RIS-Formats befindet sich unter: http://www.refman.com/support/risformat_intro.asp (Stand: 14.10.2006)

beteiligte Personen (en. „author series“). Jedes dieser Felder darf nur einen Personennamen beinhalten und ist auf 255 Zeichen begrenzt. Die Namen dürfen entweder in der gebräuchlichen Form oder in der Reihenfolge Nachname-Komma-Vorname-Spatium-Vorname-Suffix geschrieben werden, wobei Vornamen abgekürzt geschrieben werden dürfen und Suffixe wie „Sr.“ oder „Jr.“ erlaubt sind. Das Fehlen von strikten Ansetzungsregeln kann zu Problemen führen, da das Programm die Suffixe nicht syntaktisch richtig interpretieren kann und historische Namen unter Umständen während des Zitierens durch das Programm umformatiert werden.

```

TY - JOUR
A1 - Baldwin,S.A.
A1 - Fugaccia,I.
A1 - Brown,D.R.
A1 - Brown,L.V.
A1 - Scheff,S.W.
T1 - Blood-brain barrier breach following cortical
contusion in the rat
J0 - J.Neurosurg.
Y1 - 1996
VL - 85
SP - 476
EP - 481
RP - Not In File
KW - cortical contusion
KW - blood-brain barrier
KW - horseradish peroxidase
KW - head trauma
KW - hippocampus
KW - rat
N2 - Adult Fisher 344 rats were subjected to a unilateral
impact to the dorsal cortex above the hippocampus at [...]
ER -

```

Text 6: Beispiel für die Verwendung des RIS-Formats bei einem Artikel aus einer Fachzeitschrift nach: http://www.refman.com/support/risformat_sample_02.asp (Stand: 2.1.2011)

Dem RIS-Format fehlt ähnlich wie dem MARC-Format eine eindeutige Zuordnung in der Relation Teil-Ganzes, wie es z. B. bei Artikeln einer Zeitschrift der Fall ist. Diese Verbindung kann vom Computer hergestellt werden, indem der Umweg über den „Schlüssel“ genommen wird: Die

Abkürzung oder der volle Name eines Periodikums kann als Gruppierungsmerkmal herangezogen werden. Dies ist aber nur in sehr einfachen Fällen möglich d. h. wenn es keine Abweichungen in der Schreibweise gibt. (Wenn eine Zeitschrift ihren Namen ändert, ist die Zuordnung einzelner Aufsätze dieser Reihe nicht mehr möglich.)

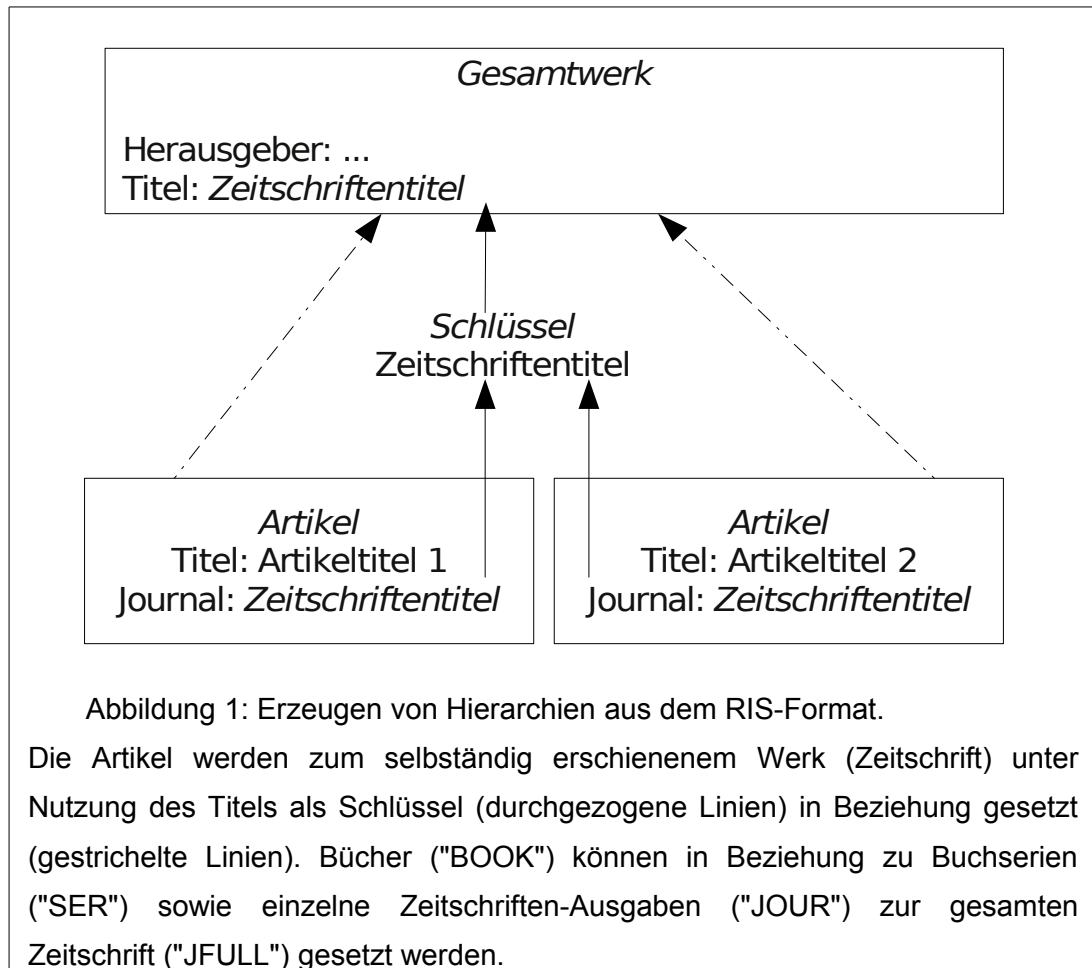
```

TY  - CONF
A1  - Catania,J.
A1  - Coates,T.
A1  - Kegeles,S.
A1  - Peterson,J.
A1  - Marin,B.
A1  - Fullilove,M.
T1  - Predicting risk behavior with the AIDS risk reduction
model (ARRM) in a random household probability sample of San
Franciscans: the "AMEN" study
Y1  - 1990/6th Annual
VL  - 6
SP  - 318
EP  - 318
RP  - Not In File
CY  - Detroit MI
KW  - risk
KW  - AIDS
KW  - models
KW  - sexual behavior
KW  - HIV
KW  - condoms
KW  - heterosexual
KW  - bisexual
KW  - ethnicity
KW  - women
T3  - International Conference on AIDS 6
Y2  - 1990/6/20
M1  - 1
N1  - OBJECTIVE: Data from the AIDS In Multi-Ethnic
Neighborhoods survey are used to test Stages 1 & 3 of ARRM
(a three stage process model of sexual risk behavior change;
Catania, Kegeles, & Coates, 1990). Stage 1 analyses examine
predictors of labeling one's sexual...
ER  -

```

Text 7: Beispiel für die Verwendung des RIS-Formats für einen Konferenzband
nach: http://www.refman.com/support/risformat_sample_04.asp (Stand: 2.1.2011)

Anders als viele andere Formate besitzt RIS Kürzel, die Vergabe von Schlüsselwörtern („KW“) und Abstracts („NX“) durch den Benutzer erlaubt. Exportiert ein Benutzer seinen Datenbestand, werden seine persönlichen Ergänzungen abgespeichert.



Das RIS-Format ist ein einfach zu implementierendes Dateiformat für die meisten bibliographisch relevanten Werke. Durch die starke Verbreitung vom Reference Manager, dem wohl am weitesten fortgeschrittenen Literaturverwaltungssystem für den PC, fand es unter Benutzern eine hohe Akzeptanz, zumal es als Datei abgespeichert und von vielen Programmen verarbeitet werden kann. Die Redundanz der Daten in ähnlichen Datensätzen und die mangelnde Vernetzung einzelner Werke und deren Teile kann nur durch ein leistungsfähiges Programm wie Reference Manager ausgeglichen werden und zu akzeptablen Suchergebnissen führen. Der Blick

auf die Bedürfnisse der Benutzer ist wohl der Grund des Erfolges: Einfachheit, Leistungsfähigkeit und Aufnahme von Zusatzinformationen aus der Sacherschließung machten dieses Format populär.

2.2.2.1.3. BibTeX

Ein weiteres zwanzig Jahre altes und weit verbreitetes Format ist das BibTeX. BibTeX bezeichnet sowohl das Dateiformat als auch das verarbeitende Programm selbst. Beides ist ein Teil des TeX (Tau Epsilon Chi), eines fast dreißig Jahre alten Textsatzsystems, welches von Donald Knuth von der Stanford Universität entwickelt worden ist. Sein Ziel war, den Wissenschaftlern ein Werkzeug an die Hand zu geben, mit dem sie ihre komplexen mathematischen Formeln erstellen könnten, ohne eine fehlerhafte Darstellung in der Print-Ausgabe befürchten zu müssen⁶¹. Auf den Erfolg von TeX baut **LaTeX** von Leslie Lamport auf, welcher die TeX-Befehle als Macro-Sammlungen zu neuen Befehlen zusammenführt und die Handhabung durch logische Gliederung des Dokumentes und den Import von beliebigen weiteren Marco-Definitionen (sog. Packages) erleichtert⁶². Diese Merkmale waren für den vermehrten Einsatz dieser Software unter Naturwissenschaftlern ausschlaggebend. Mit Verbreitung des WWW im universitären Bereich wurde LaTeX zum beliebtesten Texteditor, da die Artikel in Form von PDF- bzw. Postscript⁶³-Dateien in den Druck gehen konnten und zugleich als eine hierarchisch nach Kapiteln gegliederte Internetseite mit Inhaltsverzeichnis und Verlinkung erstellt werden konnte. Es

61 Vgl. Kalbitz, Andreas : *LaTeX-Geschichte* URL: http://www.selflinux.org/selflinux/html/latex_geschichte01.html Stand: 2.1.2011

62 Vgl. Diller, Antoni : *LaTeX wiersz po wierszu* / Jelowiecki Jan (Übers.). Gliwice: Helion, 2001. - Originaltitel: *LaTeX Line by Line*. - ISBN 83-7197-341-1. - S. XII ff

63 Postscript (Abkürzung: PS) ist eine „Programmiersprache zur geräteunabhängigen Beschreibung von Druckseiten; vorgestellt 1985 von der Firma Adobe Systems Incorporated, entwickelte sich seitdem zu einem Standard in der Druckindustrie. Druckseiten im PostScript-Format können auf den unterschiedlichsten Ausgabegeräten in variabler Größe verlustfrei ausgegeben werden.“ (Aus: Bibliographisches Institut & F. A. Brockhaus AG (Hrsg.) : *Postscript*®. URL: <http://lexikon.meyers.de/index.php?title=PostScript%C2%AE&oldid=41521> Stand: 25.08.2006, zuletzt online: 23.03.2009)

war zu erwarten, dass LaTeX um ein „Werkzeug“ für die Verarbeitung von Literaturquellen erweitert werden würde. Dies geschah 1995, als Oren Patashnik die erste Version von BibTeX herausgegeben hat.

```
@inbook{NockeBusse,
  author = {Franz-Joseph Nocke},
  editor = {Theodor Schneider},
  title = {Buße},
  chapter = {Spezielle Sakramentenlehre},
  publisher = {Patmos Verlag},
  pages = {306-334},
  year = {2002},
  volume = {2},
  address = {Düsseldorf},
  edition = {2},
  crossref = {HdD},
  booktitle = {Handbuch der Dogmatik}
}
@book{HdD,
  editor = {Theodor Schneider},
  title = {Handbuch der Dogmatik},
  publisher = {Patmos Verlag},
  year = {2002},
  address = {Düsseldorf},
  edition = {2}
}
```

Text 8: Beispiel für einen Artikel in einem Handbuch mit Querverweis (crossref) - erfasst im BibTeX-Format

Eine internationale Gemeinschaft von Benutzern und Programmierern sorgte dafür, dass BibTeX keine Interessen einer einzigen Nation oder Zielgruppe verfolgte und die Schnittstelle zur Generierung von Quellenangaben für unterschiedliche Ausgabearten offen hielt und bis heute hält. Mit Hilfe einer eigenen Macro-Sprache ist es möglich, lokalisierte (entsprechend der verwendete Sprache) und Zitierstil-gerechte (verwendete Norm) Angaben zum Zitat zu generieren. Diese Macro-Sprache ist jedoch sehr kompliziert, wodurch die Erstellung eines neuen Stils für einfache Benutzer nicht möglich ist. Durch die Bemühungen von Klaus Lorenzen (von der Hochschule für Angewandte Wissenschaften in Hamburg) entstand ein deutschsprachiger Stil⁶⁴, welches die Normen des DIN 1505-2 und 1505-3 umsetzt.

⁶⁴ Richtiger wäre die sog. Stile als Skript- oder Macro-Sammlung zu bezeichnen. Der Begriff „Stil“ bzw. die englische Entsprechung „style“ hat sich jedoch in der LaTeX-

Die BibTeX-Dateien sind Textdateien, in denen den Zeichen „@“ und „{“ bzw. „()“ eine besondere Bedeutung zukommt. Mit „@“ am Zeilenanfang wird ein Datensatz eingeleitet. Darauf folgt die englischsprachige Bezeichnung des Typs wie⁶⁵: article, book, booklet, conference, inbook, incollection, inproceedings, manual, masterthesis, misc, phdthesis, proceedings, techreport oder unpublished. Dahinter kommen die (geschweiften) Klammern, die den Inhalt umschließen. Hinter der öffnenden Klammer wird eine beliebig wählbare Kurzbezeichnung für den Inhalt angegeben. Sie wird im Text der LaTeX-Datei zur Identifizierung des zu zitierenden Werkes verwendet (im Englischen oft als „citekey“ bezeichnet). Der Inhalt besteht aus Schlüssel-Wert-Paaren, die durch Gleichheitszeichen verbunden werden, wobei der Wert in Anführungszeichen oder in Klammern stehen sollte. Diese Datenfelder müssen voneinander durch Komma getrennt werden – Formatierung durch Zeilenumbrüche dient zwar der besseren Übersicht, spielt in der Verarbeitung jedoch keine Rolle.

[1] NOCKE, FRANZ-JOSEPH: *Buße*.
 IN: SCHNEIDER, THEODOR (Hrsg.): *Handbuch der Dogmatik*,
 Band 2, Kapitel Spezielle Sakramentenlehre, S. 306-334.
 Patmos Verlag, Düsseldorf, 2. Auflage, 2002

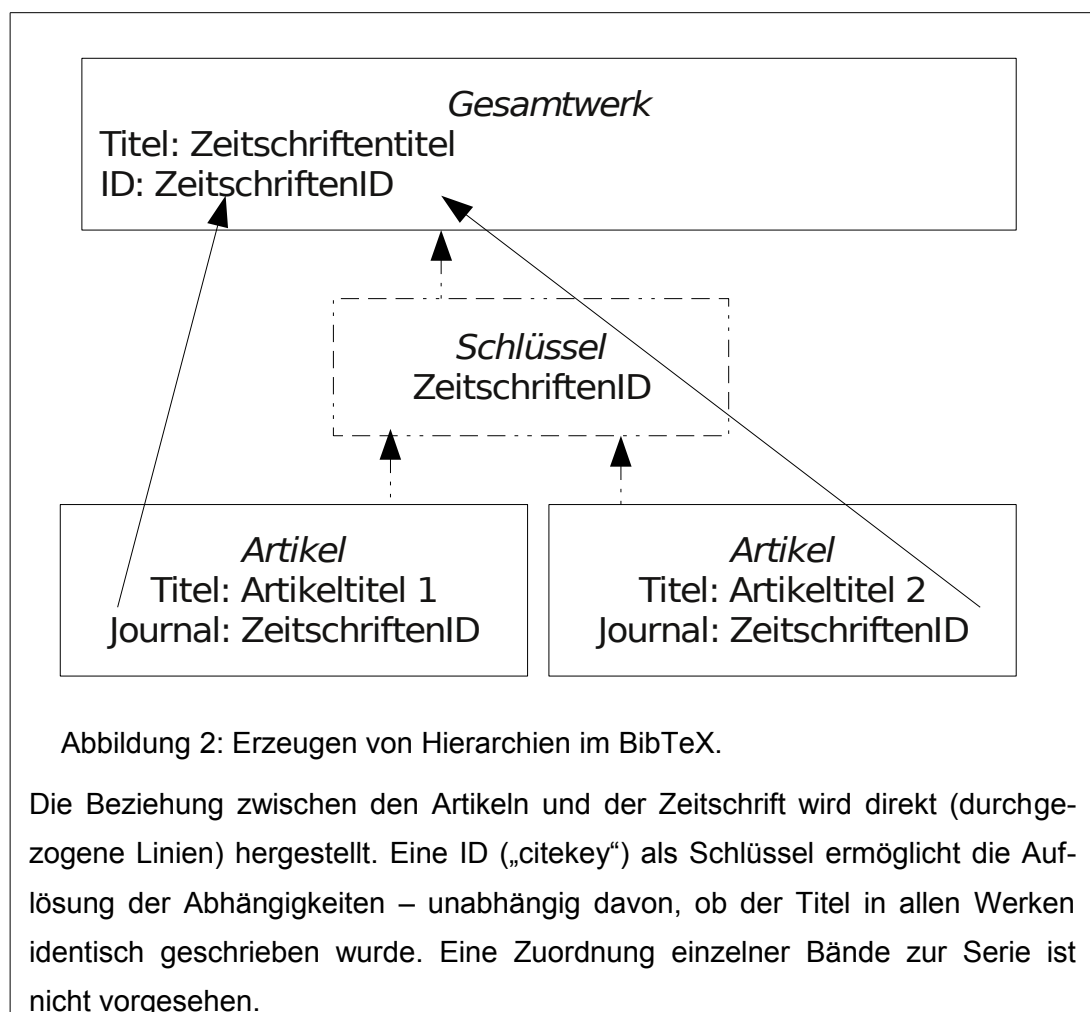
Text 9: Ausgabe von BibTeX als Teil eines nummerisch geordneten Literaturverzeichnisses

Die Datenfelder können je nach Veröffentlichungstyp unterschiedliche Typen beinhalten. Dazu zählen: address (Verlagsort bzw. Ort der Universität), annote (Bemerkungen – werden nicht in die Quellenangaben übernommen), author (der Autor), booktitle (Titel des ganzen Buches), chapter (Kapitel), crossref (Schlüssel des selbständig erschienenen Sammelwerkes, das einen Artikel enthält), edition (Bezeichnung der Ausgabe), editor (Herausgeber), institution (Organisation, die ein Raport publiziert hat), journal (Zeitschrift, in

Gemeinde eingebürgert und kann nicht ersetzt werden, ohne für Missverständnisse zu sorgen.

65 Die meisten Angaben in diesem Abschnitt stammen aus: Diller, Antoni : *LaTeX wiersz po wierszu / Jelowiecki jan* (Übers.). Gliwice: Helion, 2001. - Originaltitel: *LaTeX Line by Line*. - ISBN 83-7197-341-1; Kapitel 6: *Erstellung von Bibliographien*

der ein Artikel publiziert wurde), key (zusätzlicher Sortierschlüssel), month (Monat), note (Zusatzangaben zur Publikation – erscheinen in der Bibliographie), number (Nummer), organisation (Organisation, die eine Konferenz veranstaltet bzw. eine Benutzeranleitung verfasst hat – entspricht in etwa einer Körperschaft), pages (Seitenangaben), publisher (Verlag), school (die Hochschule, an der eine Magisterarbeit oder Promotion angefertigt wurde), title (Titel), type (Typ des Reports), volume (Bandangabe), year (Jahr).



Da die Felder nicht wiederholt werden dürfen, müssen Angaben zu mehreren Verfassern oder Herausgebern entweder unformatiert oder als Nachname-Komma-Vornamen-Wert in einer Zeile durch „and“ verbunden geschrieben werden. Bei Königs-, Päpste-Namen oder mehrteiligen Pseudonymen kann

der Wert vor einer Umformatierung (während der Generierung von Quellenangaben) durch das Einschließen in Klammern geschützt werden. Dieser "Geheimtip" ist den meisten Anwendern unbekannt und wird in Programmen zur Erstellung und Verwaltung von Literaturquellen in BibTeX-Dateien, den sog. Frontends als besonderes Merkmal gar nicht angeboten. Ähnlich wie Reference Manager formatiert BibTeX die in üblicher Weise angegebenen Namen nach Vorgaben des Stils um und nimmt an, dass der letzte Bestandteil des Namens den Nachnamen darstellt (wenn kein Komma im Namen vorkommt). Dies ist im Fall von „Benedikt XVI.“ oder „Johann Strauss Sohn“ nicht der Fall. Anders als bei RIS wird der Rang der Autoren nicht unterschieden, so dass die Forderung des DIN 1505-2, weitere beteiligte Personen hinter dem Titel anzugeben, nicht erfüllt werden kann.

Einen entschiedenen Vorteil gegenüber RIS wird in der Verarbeitung von Artikeln aus Konferenzbänden und Zeitschriften sichtbar. Mit Hilfe von crossref ist es möglich, ein Artikel durch Angabe des Schlüssels des Gesamtwerkes diesem zuzuordnen. Dabei entsteht keine Redundanz, da das Gesamtwerk nur einmal vorkommen muss. Dieses Merkmal ist beim Zitieren von unselbständig erschienenen Einheiten besonders wichtig: laut DIN 1505-2 soll einem solchen Werk nach „In: “ das Gesamtwerk angegeben werden. Dies ist aber nur dann möglich, wenn das Verhältnis Teil-Ganzes dem Programm bekannt ist. Auch die Autorennamen oder häufig wiederkehrende Titel können einem Schlüssel fest zugeordnet werden, der an ihrer Stelle angegeben wird (sog. Aliase).

2.2.2.1.4. XML – eXtensible Markup Language

Für die Abbildung unterschiedlicher Datentypen in Textdateien wurden viele Methoden entwickelt, die in Form von Programmen bzw. deren Teilen (den sog. Funktionsbibliotheken) implementiert wurden. RIS und MARC sind Beispiele dafür. Sie können bibliographische Daten unterschiedlicher Werke erfassen und typisieren. Diese Daten kann jedes Programm mit Hilfe jener Funktionsbibliotheken lesen und im Speicher auf entsprechende Datentypen abbilden. Dies ist fast ausschließlich bei Programmen zur Literaturverwaltung

der Fall, da die Formate im Hinblick auf Literatursammlungen oder Abgleich von Bestandsdaten hin entwickelt wurden. Extensible Markup Language (XML) ist – ähnlich wie **Standard Generalized Markup Language (SGML)** – eine Syntax für Beschreibungssprachen, die jegliche Inhalte (z. B. bibliographische Daten als auch den Volltext eines Werkes oder Kommentare zu diesem) in einer Datei enthalten können⁶⁶. Diese Universalität und Einfachheit der Syntax machte XML zum weltweit populärsten Datenformat. Jeder Entwickler kann auf einer abstrakten Ebene die Syntax und die Semantik nach Bedarf modellieren und als ein **Schema** bzw. **Document Type Definition** in einer Datei ablegen. Für das Abbilden dieser Daten im Speicher und deren Verarbeitung gibt es zwei sehr verbreitete und zugleich einfach einzusetzende Methoden, die zum Bestandteil der meisten Betriebssysteme geworden sind und von jedem Programm genutzt werden können: das DOM (Document Object Model), der XmlReader oder das SAX (Simple API for XML).

Die XML-Dateien, die für bessere Lesbarkeit mit Zeilenumbrüchen und Tabulatoren formatiert werden können, aber auch ohne diese Formatierung dieselbe Bedeutung haben, besitzen eine baumartige Struktur. Dieses Merkmal wird von XPath benutzt, um direkt einzelne Teile (Elemente oder „Zweige“) der Datei auszulesen oder zu verändern. Das XML kann sowohl für narrative Dokumente (wie Artikel oder Essays) als auch für Datenbank-ähnliche Datenstrukturen eingesetzt werden. Diese Eigenschaft von XML wird z. B. im TEI-Format⁶⁷ eingesetzt, um die Akte eines Dramas oder die Metrik einzelner Strophen eines Gedichtes als Metadaten im Text zu

66 Vgl. Harold, Elliotte R. ; Means, W. Scott : *XML in a Nutshell* / Deutsche Übersetzung von Katharina Udemadu, Kathrin Lichtenberg und Lars Schulten. 3. Auflage. Köln : O'Reilly, 2005. - ISBN 3-89721-339-7. S. 3 f

67 Text Encoding Initiative (TEI) ist eine Initiative, die sich zum Ziel gesetzt hat, Texte mit Hilfe von XML-Auszeichnungen um zusätzliche (oft interpretationsrelevante) Informationen zu erweitern. Dazu entwickelte das Text Encoding Initiative Consortium ein XML-basiertes Dateiformat, welches besondere Merkmale eines Werkes (wie z. B. Strophen und Verse in Gedichten) beschreiben kann. Vgl. dazu <http://www.tei-c.org/index.xml>

kodieren. Mit **RDF**⁶⁸ gibt es eine XML-Anwendung, die die Relationen zwischen einzelnen Elementen darstellen kann. Es wird vermehrt als „Methode“ der Erfassung von Metadaten nach Dublin Core verwendet.

Neu an XML – gegenüber MARC oder RIS – ist die Möglichkeit, es um neue Elemente zu erweitern und in einem eigenen Kontext zu verwenden⁶⁹. Der Kontext (en. Namespace) wird am Anfang der Datei unter Angabe der URL zum Schema bzw. zur DTD definiert. Ein validierender Parser kann die eine XML-Datei durch einen Abgleich mit der DTD bzw. dem Schema auf Gültigkeit prüfen, ohne sie (semantisch) "verstehen" zu müssen. Es ist für den Computer ebenfalls möglich, die Daten in ein anderes Format zu überführen (wie XHTML, XML oder PDF). Diese Methode nennt man **Extensible Stylesheet Language (XSL)**. Auf diese Art und Weise kann der Inhalt der Auszeichnung „Titel“ in der Quellenangabe wie ein Titel (kursiv) formatiert werden. Für jede Zitiernorm kann mit Hilfe von XSL oder **Cascading Style Sheets**⁷⁰ (**CSS**) eine eigene Stilvorlage erstellt werden. Dies setzt voraus, dass die Auszeichnungen gemäß einer einheitlichen Namenskonvention verwendet werden. Die Semantik ist bei der Nutzung der Daten entscheidend: Kennt ein Programm die Bezeichnung „Untertitel“ nicht, kann er sie nicht verwenden. Aus diesem Grund wird angestrebt, XML-Formate zu standardisieren. Mit **MODS** – der auf der Seite 122 näher beschrieben wird – existiert ein solcher Standard, der als Austauschformat zwischen Bibliothekskatalogen und Benutzern verwendet wird.

68 **Ressource Description Framework** ist ein XML-basiertes Format, welches von W3C entwickelt wurde. Es stellt ein besonders einfaches Datenmodell dar, welches mit Hilfe von Tripeln Ressourcen beschreibt (Tripel ist ein dreigliedriger Satz in Form von: "*Ressource X hat die Eigenschaft Y mit einem Wert Z*"). Vgl. dazu: Schütz, Thomas : *Dokumentenmanagement*. In: KSS. S. 341

69 Vgl. *XML in a Nutshell*. S. 65-67

70 Das CSS – eine weitere Entwicklung des W3C – ist „eine deklarative Sheet-Style-Sprache für sturkturierte Dokumente wie HTML“ (Entnommen aus: *CSS (Cascading Style Sheet)* In: InfoWissWiki. URL: http://server02.is.uni-sb.de/courses/wiki/index.php?title=CSS_%28Cascading_Style_Sheets%29&oldid=6098 Stand: 13.08.2011). Sie definiert die Darstellung von Elementen solcher Dokumente, ohne sie in ein anderes Format zu überführen (wie XSL).

Neben den Namen von Elementen, die sich auf eine definierte semantische Konvention berufen, gibt es die sog. Attribute, die weitere Angaben zum Typ des Inhalts machen können. Damit ist es möglich, dem Computer mitzuteilen, ob z. B. der Name des Autors vordefiniert ist, nicht verändert werden darf etc. Der Computer kann diese Angaben berücksichtigen, wenn sie der Namenskonvention entsprechen (d. h. der Computer "weiß", was ein Attribut bedeutet) und zur verwendeten Syntax dieser Konvention passen (d. h. nur dort vorkommen, wo sie vorkommen dürfen).

Ein weiterer Vorteil von XML ist die Verlinkung zwischen Elementen innerhalb derselben Datei mit Hilfe von ID- und IDREF-Attributen. „Ein Attribut vom Typ IDREF verweist auf ein Attribut vom Typ ID eines Elementes im Dokument. ... IDREF-Attribute werden üblicherweise dazu eingesetzt, um Beziehungen zwischen Elementen zu realisieren, wenn einfache Enthalten-in-Beziehungen nicht ausreichen.“⁷¹ IDREFS-Attribut ermöglicht sogar die Angabe mehrerer durch ein Leerzeichen getrennter Identifikatoren. Mit XML ist es also möglich, die Daten intern so zu vernetzen, dass keine Redundanz entsteht und die Inhalte der Elemente unterschiedlicher Zweige aufeinander verweisen. Für ein bibliographisches Datenformat würde dies folgendes Bedeuten: Der Autor wird nur im ersten seiner Bücher (oder auch unabhängig davon) mit einer ID angegeben. Alle weiteren Werke dieses Autors verweisen mit Hilfe von IDREF im Autor-Element auf seine ID. Dieser Ansatz erinnert an die Struktur der Daten in einer relationalen Datenbank oder an einen Thesaurus.

XML wird gegenwärtig aus vielen Gründen als das mächtigste Datenformat angesehen. Es kann von jedem Programm mit Hilfe der hier vorgestellten Methoden gut verarbeitet werden. Ein großer Nachteil von XML ist der sog. Daten-Ballast (ein ungünstiges Verhältnis von Datenmenge zu Auszeichner), welcher je nach Inhalt stark variieren kann. Dadurch verlangt die Verarbeitung großer Datenmengen im XML-Format mehr Speicher und

71 *XML in a Nutshell*. S. 47

Berechnungszeit, als es bei binären oder textbasierten Datenformaten der Fall ist.

2.2.2.2. Binäre Datenformate

Wie bereits erwähnt, sind binäre Datenformate aufgrund der Verwendung nichtdruckbarer Zeichen schneller zu verarbeiten als Textformate. Die Bedeutung solcher Zeichen muss genau definiert sein und das Vorkommen solcher Zeichen in abzuspeichernden Daten mit Hilfe von sog. Escape-Sequenzen⁷² umgangen werden. Binär formatierte Daten enthalten weniger Ballast. Sie können in den meisten Fällen nur von speziellen Programmen verarbeitet werden. Sie sind nur für Computer lesbar. Beispiele dafür sind Bild-, MS-Word- oder Datenbank-Dateien. Anders als bei Textdateien kann die Änderung eines einzigen Bits die ganze Datei unlesbar machen.

Im bibliothekarischen Umfeld kommen die binären Datenformate fast ausschließlich in Datenbanken vor, die die Grundlage der elektronischen Kataloge bilden. Eine seltene Ausnahme stellt hier das OpenOffice dar, welches die eigene Literaturdatenbank aus historischen Gründen in einem solchen Format abspeichert.

1. (Relationale) Datenbanken

Die Datenbanken sind eine relativ alte und im Computerzeitalter sehr verbreitete Art der Speicherung von Daten. Aus der Sicht der Informationswissenschaft ist eine Datenbank eine Sammlung von Daten, deren Organisationform beliebig sein kann (jedoch konsequent verwendet wird). In der Informatik versteht man unter einer Datenbank „eine bestimmte Architektur von Programmen und Datenbeständen, die bestimmte formale Anforderungen erfüllen muss“⁷³. In diesem Abschnitt wird Datenbank in diesem Sinne verstanden.

Die Relevanz der Datenbanken – die nicht speziell auf die Verarbeitung von bibliographischen Daten ausgelegt sind – ist für diese Arbeit insofern

72 Vgl. Wikimedia Foundation Inc. (Hrsg.): *Escape-Sequenz*. URL: <http://de.wikipedia.org/w/index.php?title=Escape-Sequenz&oldid=19939353>

73 Lang, Elke : *Datenbanken und Datenbank-Management-Systeme*. In: KSS. S. 409

gegeben, als diese Art der Organisation von Daten für die dauerhafte Speicherung in elektronischen Katalogen weit verbreitet ist. Der Verwendungskontext der Datenbanken ist die Speicherung und das Retrieval.

Im Folgenden sollen einige Datenbanktypen (Modelle) vorgestellt werden, die einen Einblick in die Art und Weise der Handhabung der Daten gewähren und zugleich für das Verständnis des zu entwickelnden bibliographischen Datenmodells unverzichtbar sind.

In der Informatik wurden unterschiedliche Modelle der Verarbeitung der Datenbankdaten erarbeitet. Dazu zählen vor allem: das relationale Modell, das objektorientierte Modell und das hybride objektrelationale Modell. Im relationalen Modell werden die Daten in unterschiedlichen Tabellen abgespeichert und können mit Hilfe der **Structured Query Language (SQL)** durch die Identifikatoren so zueinander in Verbindung gesetzt werden, dass die gelieferten Daten vollständig sind, ohne dass sie redundant abgespeichert werden müssen. Die Einhaltung der Regeln für die Strukturierung der Daten – der sog. Normalformen – hängt mit der Art und Weise der Speicherung. Diese werden in Tabellen abgelegt. Eine Spalte darf nur ein Attribut beinhalten. Jede Entität darf nur einmal in einer Tabelle vorkommen. Jedes Merkmal einer Entität wird einzeln gespeichert. Das wichtigste und eindeutige Merkmal einer jeden Entität ist die ID. Kommt ein Merkmal mehrfach vor (1-zu-M-Relation), wird es in eine zweite Tabelle ausgelagert und mit Hilfe des Identifikators zugeordnet. Merkmale, die vielen Entitäten zukommen, müssen aus Gründen der Redundanzsenkung ebenfalls ausgegliedert werden. Besitzt eine Entität mehrere Attribute, die zugleich anderen Entitäten zukommen, spricht man von einer M-zu-N-Relation.

Diese Vorgehensweise der Datenverwaltung ist sehr effektiv: Mit Hilfe von Formeln können die nicht redundanten Daten so zusammengestellt werden, dass sie eine bestimmte Sicht auf die Dinge bieten. (Eine solche Zusammenführung von Informationen wird als „View“ bezeichnet.) Eine

Suche kann sehr exakt formuliert werden und führt zu einer sehr hohen Präzision bei einem niedrigen Recall.

Das objektorientierte Datenbankmodell basiert auf dem aus der Programmierung stammenden Prinzip, dass die zu beschreibenden Objekte anhand ihrer Fähigkeiten (Methoden) und Eigenschaften (Attribute) gruppiert werden können. Solche abstrakten Gruppen nennt man **Klassen**⁷⁴. Die einzelnen **Objekte** sind konkrete Realisierungen (**Instanzen**) einer Klasse. Ein Objekt der Klasse „Buch“ hat demzufolge prinzipiell: zumindest einen Autor oder Herausgeber, einen Titel, einen Verleger und eine ISBN. Diese Eigenschaften bleiben trotz unterschiedlicher Werte als Merkmale von Büchern konstant d. h. ein Werk ist nur dann ein Buch, wenn es alle diese Merkmale besitzt.

Diese Gesamtschau dieser Merkmale, die eine Klasse ergeben, ist den relationalen Datenbanken fremd, entspricht jedoch eher der menschlichen Denkweise. Das objektorientierte Datenbankmodell verarbeitet die Daten so, dass jede Entität immer mit allen ihr zukommenden Merkmalen dargestellt wird. Dies hat den Vorteil, dass die nach dem Paradigma der objektorientierten Programmierung erstellten Programme die zu verarbeitenden Daten ohne weitere Anpassungen direkt in der Datenbank ablegen oder ebenso aus dieser lesen können. Man spricht bei dieser Vorgehensweise von einer persistenten Speicherung⁷⁵.

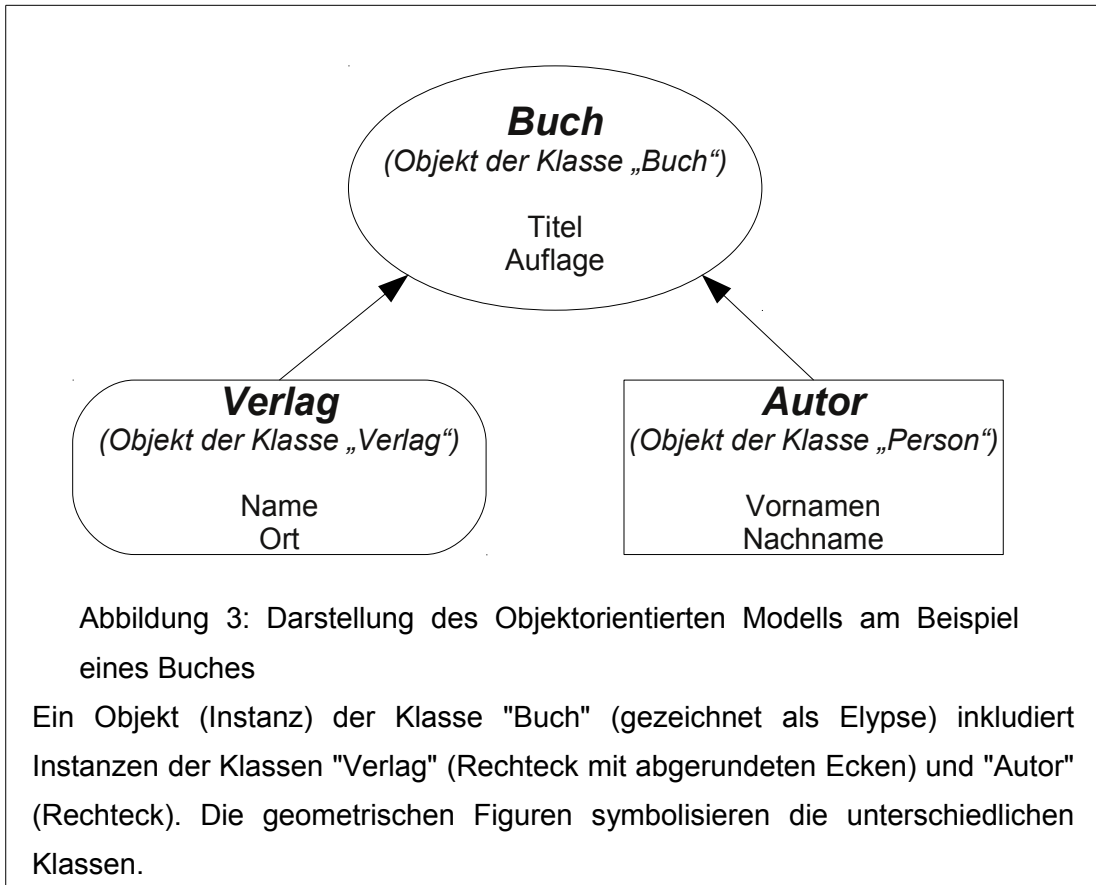
Dies trifft ebenfalls auf die sog. Vererbung⁷⁶ zu – die Übernahme der Eigenschaften der Objekte höherer Hierarchie auf die von ihnen abgeleiteten. Die Vererbung kommt in Bibliotheken in Form von unselbständig erscheinenden Werken vor. Diese erben von den selbständig erschienenen Werken, in

74 vgl. Panyr, Jiri : *Thesauri, Semantische Netze, Frames, Topic Maps, Taxonomien, Ontologien - begriffliche Verwirrung oder konzeptionelle Vielfalt?* In: *Festschrift Zimmermann*. S. 142

75 vgl. Lang, Elke : *Datenbanken und Datenbank-Management-Systeme*. In: *KSS*. S. 413

76 vgl. Panyr, Jiri : *Thesauri, Semantische Netze, Frames, Topic Maps, Taxonomien, Ontologien - begriffliche Verwirrung oder konzeptionelle Vielfalt?* In: *Festschrift Zimmermann*. S. 145

welchen sie abgedruckt worden sind, die meisten Eigenschaften wie z. B. das Erscheinungsjahr, den Verlag etc.



Die Datenbanken – ob als relationale, objektorientierte oder hybride objektrelationale – haben gegenüber reinen Textformaten viele Vorteile. Die sehr effiziente – auf das Retrieval hin ausgerichtete – Art und Weise der Speicherung wird vor allem bei großen Datenbeständen benötigt. Für die Übertragung ist dieses Datenformat nicht geeignet. Zu diesem Zweck wird die textbasierte Abbildung der Daten (meist SQL) eingesetzt.

2.2.3. Austauschprotokolle

Die verschiedenen Datenformate, die für die Speicherung und Verarbeitung bibliographischer Daten eingesetzt werden, können in für den Austausch dieser Daten benutzt werden. Je nach Quelle und Ziel der Daten werden drei Kontexte unterschieden:

- vom Katalog zum Katalog,
- vom Katalog zum Benutzer hin
- und unter Benutzern selbst

Beim Austausch zwischen Katalogen wird in Deutschland das MAB2-Format für den Datenabgleich verwendet. Die Auslieferung der Daten eines elektronischen Kataloges an den Benutzer wird je nach Bibliothek mit Hilfe von MARC, MAB oder anderen Formaten realisiert. Ein Austausch zwischen den Benutzern geschieht in der Praxis über RIS oder BibTeX.

Hinsichtlich der Art der Übermittlung kann man drei Typen unterscheiden:

- Austausch über Dateien oder Magnetbänder auf physikalischem Weg,
- Austausch von bibliographischen Dateien über das Internet (beliebige Applikationsschicht des OSI-Modells)
- und den an einen bestimmten Protokoll gebundenen Austausch von bibliographischen Daten.

Der Unterschied zwischen dem zweiten und dritten Typ besteht darin, dass der zweite nur eine Verlagerung des physikalischen Weges auf eine Übertragung über das Internet darstellt (der Empfänger einigt sich mit dem Lieferanten auf eine „protokolllose“ Art und Weise über die benötigten Datensätze und deren Lieferung – wie z. B. per E-Mail) während beim dritten Typ die Anfrage nach Datensätzen den Regeln eines technischen Protokolls genügen muss.

Das **Austauschprotokoll** definiert nicht **was** (welche Daten oder Dateien), sondern **wie** ausgetauscht wird (wie muss eine Anfrage oder eine Antwort aussehen, damit sie von beiden Teilnehmern der „Konversation“ verstanden wird).

2.2.3.1. Z39.50

Das bekannteste Protokoll für den Information Retrieval im bibliographischen Bereich stellt das Z39.50 dar. Es ist ein nordamerikanischer (ANSI/NISO Z39.50) und internationaler (ISO 23950) Standard, welcher seit 1984 von der Library of Congress entwickelt und ständig verbessert wurde (Version 1: 1988; Version 2: 1992; aktuelle Version: 2003)⁷⁷. Es ist ein Protokoll der

⁷⁷ vgl. National Information Standards Organization (Hrsg.): *Information Retrieval*

siebten Ebene des OSI-Modells⁷⁸, welches auf dem TCP/IP-Protokoll aufbaut. Eine durch ein Z39.50-Protokoll-fähiges Programm erstellte Anfrage wird nach der Authentifizierung auf dem Server an ein Program übergeben, welches die gewünschten Daten in der Datenbank sucht, sie entsprechend einem Datenformat (genauer „Übertragungsformat“ wie z. B. UNIMARC) zusammenstellt und an den Benutzer über dieses Protokoll ausliefert.

Wolfram Schneider⁷⁹ definiert das Protokoll wie folgt:

„Es ist ein Protokoll zur Kommunikation zwischen bibliothekarischen Datenbanksystemen (Server) und Zugriffsprogrammen (Clients). Z39.50 erlaubt die Suche in heterogenen Datenbanken aus der gewohnten lokalen Programmumgebung. Die Verwendung des Z39.50 Protokolls führt zu einer Unabhängigkeit von der Datenbank, der lokalen Abfragesyntax, dem eingesetzten Betriebssystem und der Hardware. Man kann sich das Z39.50-Protokoll als ein Art Datenbank-Esperanto vorstellen, das jedem Client ermöglicht, mit jeder Datenbank einen Dialog zu führen.“⁸⁰

Z39.50 ist sehr komplex und vielschichtig. Deshalb realisieren die meisten Anbieter solcher Software nur einen Teil (das sog. „Core“) dieses Protokolls. Mit Ausnahme von „YAZ Toolkit“ sind fast alle Programme kostenpflichtig⁸¹.

(Z39.50): *Application Service Definition and Protocol Specification*. Bethesda: NISO Press, 2003 - ISBN: 1-880124-55-6. S. 1

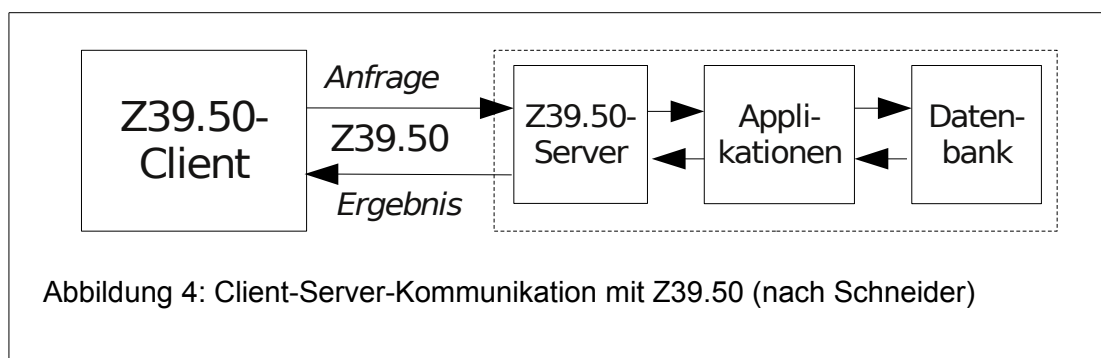
78 Payer, Margarete: *Computervermittelte Kommunikation*. – Kapitel 113. *OSI-Schicht 7: Application Layer – Anwendungsschicht*; Teil 3: *Information-Retrieval-Protokoll Z39.50*. URL: <http://www.payer.de/cmcl/cmcs1303.htm> Stand: 2.1.2011

79 Schneider, Wolfram: *Ein verteiltes Bibliotheks-Informationssystem auf Basis des Z39.50 Protokolls*. Berlin, Technische Universität Berlin, Fachbereich Informatik, Dipl.-Arb., 1999. URL: http://www.zib.de/groetschel/students/diplom_schneider.pdf Stand: 2.1.2011

80 ebd. S. 11

81 Vgl. Taube, Anke: *Konzept und Implementierung eines Moduls zum Abgleich lokaler Zeitschriftenbestände am DKFZ mittels Z39.50 Protokoll*. Darmstadt, Fachhochschule Darmstadt, Fachbereich Informations- und Wissensmanagement, Dipl.-Arb., 2004. S.

Ein minimaler Satz von sog. Diensten, die für eine funktionsfähige Software notwendig sind, besteht aus: **INIT**, **SEARCH** und **PRESENT**. Sehr nützliche, aber optionale Dienste sind: **DELETE** und **SCAN**⁸². Der INIT-Dienst initialisiert eine Verbindung und übermittelt die für die Nutzung eines Gateways notwendigen Authentifizierungsdaten (Datenbank, Benutzername und Passwort) und verhandelt die von beiden Seiten unterstützten Parameter aus (Zeichensatz, Übertragungsformate etc.). Der SEARCH-Dienst führt eine Anfrage aus und liefert die Anzahl der Treffer. Mit Hilfe des PRESENT-Dienstes werden die Datensätze vom Server abgeholt und zur weiteren Verarbeitung auf dem Rechner übernommen. Der SCAN-Dienst ermöglicht eine Suche innerhalb der Treffer, wodurch man die Abfrage genauer formulieren (einschränken bzw. präzisieren) kann. Diese geschieht sinnvollerweise zwischen SEARCH und PRESENT.



Die Anfrage wird mit Hilfe von BIB-1 Attributensatz (en. „Attribute Set“) formuliert, in die Übertragungssyntax (en. „Transfer Syntax“) RPN übersetzt (datenbankunabhängige Syntax für Suchanfragen) welche vom Z39.50-Gateway durch einen Suchanfrageparser (en. „Query Syntax Parser“) unter Nutzung von CCL (Common Command Language) auf die Abfragesyntax der Datenbank abgebildet wird. Mit BIB-1 wird definiert, was gesucht wird (ganzes Wort, rechts- oder links-trunkiertes Wort, Datum etc.) und wo gesucht werden soll (überall im Feld, im Teilfeld etc), in welcher Relation (kleiner als, größer als, gleich, ungleich) sowie welcher Datentyp gesucht

15. URL: http://www.glycosciences.de/spec/dipl_diss/anke.taube.dipl.pdf - Stand 6.2.2011)

82 ebd. S. 23

wird (Personenname, Titel oder ISBN)⁸³. Mit CCL kann man z. B. logische Operationen an Treffermengen definieren und so die Präzision erhöhen (Festlegung von Reihenfolgen oder Verbindung von Einzelergebnissen mit Hilfe von booleschen Operatoren).

Die Ergebnisse der Suchanfrage werden entweder in einem von Abstract Syntax Notation One (ASN.1) abstammenden Format (SUTRS, OPAC Record Syntax, SUMMARY Record Syntax oder Generic Record Syntax One [GRS-1]) oder in einem bibliographischen Format an den Benutzer ausgeliefert. Um diese Daten verarbeiten zu können, muss das Programm des Benutzers dieses Format unterstützen. Ein passendes Format kann zwischen dem Programm und dem Server in der Initial-Phase ausgehandelt werden, wird aber meist in der Konfiguration der Verbindung vorgegeben. In den meisten Fällen kommt einer der MARC-Formate zum Einsatz.

Das Z39.50 ist zwar universell und kontextunabhängig, zugleich sehr komplex und vielschichtig, weshalb seine Umsetzung in der Praxis lückenhaft ist (als optional deklarierte Operationen). Z39.50 wurde als Protokoll für den Information Retrieval definiert, wird aber ausschließlich in elektronischen Bibliothekskatalogen verwendet. Die Entwicklung in der EDV zeichnet sich gegenwärtig durch stärkere Nutzung einfacherer und allgemein akzeptierter Protokolle (wie des HTTP) aus. Die ZING⁸⁴-Initiative der Z39.50-Agency versucht mit SRW / SRU diesem Trend zu folgen. Es ist zu vermuten, dass Z39.50 langfristig ganz abgelöst wird.

2.2.3.2. MedLine (PubMed)

PubMed ist die weltweit größte und umfangreichste Datenbank medizinischer Veröffentlichungen (meist Artikel) aus verschiedenen Zeitschriften. Sie ist heute aus der wissenschaftlichen Arbeit in dieser Disziplin nicht

83 Genauere Angaben mit Tabellen zum Bib-1-Attributensatz gibt Anke Taube im Kapitel 2.6.2.1. sowie die Z39.50-Agency unter: <http://www.loc.gov/z3950/agency/defs/bib1.html> (Stand: 12.12.2008)

84 ZING steht für „Z39.50 International New Generation“ - ein Nachfolgemodell, welches an die Funktionalität von Z39.50 anknüpft. Seit 2009 hat sich der Begriff "SRU/W" anstelle des ZING durchgesetzt.

Anfrage:

```
GET /entrez/utils/pmqtty.fcgi?
term=Hamacher+J.&dispstart=0&db=pubmed&field=ALL&dispmax=5&
opt=d
```

Antwort des Servers:

```
<Title>QueryResult</Title>
<Body>
16937331<Br>
16823647<Br>
16815057<Br>
16518805<Br>
16450002<Br>
</Body>
```

Anfrage an den Server nach Details zu diesen PMID:

```
GET /entrez/utils/pmfetch.fcgi?
report=medline&db=pubmed&mode=text&id=16937331,16823647,1681
5057,16518805,16450002
```

Die vom Server generierte Antwort im PubMed-Format (gekürzt und auf eines begrenzt):

```
PMID- 16937331
OWN - NLM
....
DP - 2006
TI - Heterotopic cardiac xenotransplantation in rodents:
report of a refined technique in a hamster-to-rat model.
FAU - Schramm, Rene
AU - Schramm R
FAU - Schafers, Hans-Joachim
AU - Schafers HJ
FAU - Hamacher, Jurg
AU - Hamacher J
...
JT - Microsurgery.
PST - ppublish
SO - Microsurgery. 2006;26(6):484-5.
Text 10: Ablauf der Kommunikation zwischen Pybliographer und PubMed-Server
bei der Suche nach fünf Veröffentlichungen von "Hamacher J."
```

wegzudenken. Sie wurde entwickelt vom National Center for Biotechnology Information (NCBI). PubMed ist die kostenfreie Version der MedLine-Datenbank, die über eine Web-Oberfläche als auch mit Hilfe von speziellen Programmen durchsucht werden kann. Bei MedLine handelt es sich um eine bibliographische Referenzdatenbank mit Formal- und Sacherschließung mit dem Schwerpunkt der Tier- und Humanmedizin, des öffentlichen

Gesundheitswesens, Psychologie, Genetik und Biologie. Vom United States National Library of Medicine (dem Anbieter von PubMed) werden ebenfalls andere Datenbanken gepflegt wie z. B. TOXNET⁸⁵ oder Clinical Queries, die die PubMed-Datenbank (aber auch andere Datenbanken) nach bestimmten Aspekten durchsucht (in unserem Beispiel: auf dem Gebiet der Toxikologie bzw. nur klinische Studien). Sie verfügt über ein Thesaurus, der bei der Auflösung von allgemeinen Begriffen auf medizinische Termini dient (**Medical Subject Headings - MeSH**). Des Weiteren bietet sie die Möglichkeit einer schnellen „Dereferenzierung“ – der Auflösung von Quellenangaben aus Fußnoten auf entsprechende Datensätze. PubMed unterstützt Open Access⁸⁶ und bietet entweder die Volltexte oder Links zu diesen in der Detailansicht der Ergebnisse an. Mit PubMed Central OAI service (PMC-OAI) gibt es eine Implementierung des Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) – eines Standards für das Retrieval in Metadaten elektronischer Datenbestände.

Die Suche in der PubMed-Datenbank kann auf zwei Wegen geschehen: entweder in einer Web-Oberfläche unter <http://www.ncbi.nlm.nih.gov/entrez/> oder mit Hilfe von Programmen zur Bibliographieverwaltung. Je nach Art der Suche weichen die Schritte voneinander ab. Bei einer Suche im WWW-Browser folgt der Eingabe und dem Absenden des Suchbegriffs an ein CGI⁸⁷-Programm die Ausgabe einer Liste der Ergebnisse begrenzt auf eine vom Benutzer zu definierende Zahl. Weitere Ergebnisse können durch das Klicken auf entsprechende Fläche ('Next') bzw. durch die Eingabe der Ergebnis-Seitennummer erreicht werden. Dabei wird die Anfrage erneut ausgeführt und auf die gewünschten Sortiernummern begrenzt. Durch das Anklicken eines Links in der Liste kann die Detailansicht mit weiteren Daten erreicht werden.

85 Die Web-Oberfläche zu TOXNET ist zu finden unter: <http://toxnet.nlm.nih.gov/>

86 Der Zugang zur Liste der erfassten Open-Access-Zeitschriften befindet sich unter <http://www.pubmedcentral.nih.gov/about/openftlist.html> (Stand: 2.1.2011)

87 Common Gateway Interface (CGI) ist eine Schnittstelle zwischen dem Webserver (z. B. Apache) und dem Programm, welches die empfangenen Daten verarbeitet und die Ausgabe generiert, die vom Webserver an den Empfänger ausliefert.

Abbildung 5: Gefilterte Ergebnisliste einer MedLine-Abfrage

Die ersten hundert Ergebnisse der Suche nach Veröffentlichungen von "Hamacher" in MEDLINE, gefiltert nach dem Begriff "Vena". Hier am Beispiel des Open-Source-Programms Pybliographer.

Bei Bibliographieverwaltungsprogrammen wie Reference Manager oder Pybliographer wird zunächst auf der Benutzer-Seite aus den Suchbegriffen eine Suchanfrage generiert, die per HTTP an das CGI-Programm von PubMed geschickt wird. Wurde keine Begrenzung der Ergebnisse auf gewisse Sortiernummern festgelegt, antwortet der Server mit Identifikatoren⁸⁸ aller betroffenen Datensätze und der Benutzer wird aufgefordert, eine Auswahl zu treffen. Danach werden alle ausgewählten Datensätze mit allen Details in Form einer Textdatei im PubMed-Format ausgeliefert.

Das Programm des Benutzers muss diese Datei parsen und entsprechend graphisch aufbereiten. Zumeist wird das dadurch erreicht, dass das Fenster in oberen und unteren Teil aufgeteilt wird und die Ergebnisse als Liste mit

88 PubMed identifiziert die Datensätze mit Hilfe von Identifikationsnummern, die sowohl in der Kommunikation zwischen Benutzerprogramm und Server als auch in der Ergebnis-Datei zu finden sein wird. Sie trägt die Bezeichnung „PMID“.

Titel, Autorennamen und Datum oben erscheinen. Nach dem Anklicken eines von Ihnen werden weitere Informationen im unteren Fensterbereich eingeblendet. Per Doppelklick wird meist eine Kopie des Datensatzes lokal zum Editieren geöffnet.

Da das Retrieval in Pubmed durch das sehr verbreitete HTTP-Protokoll geschieht, ist es möglich, mit jeder Programmiersprache dieses Protokoll umzusetzen. Die Einfachheit des Formats der PubMed-Ergebnisdateien erfordert vom Programmierer nur wenig Aufwand beim Erstellen einer Funktion für die Verarbeitung der Daten (des sog. Parsers). Am Anfang jeder Zeile steht eine Kürzel⁸⁹, die aus zwei bis vier Großbuchstaben besteht und die darauf folgende Angabe typisiert. Die Quellenangaben sind genauso umfangreich wie die möglichen Qualifikatoren. Diese zwei Merkmale trugen – neben dem Umfang der Datenbank – zur einer hohen Anzahl an Software-Lösungen zur Suche in PubMed bei. Im Unterschied zu Pybliographer kann die Suche auf der Internetseite und im Reference Manager viel feiner definiert werden, z. B., indem der eingegebene Autorennamen mit dem Index (oder Personennamendatei) von PubMed verglichen und eine Auswahl möglicher Namen angeboten wird. In allen Varianten der Suche besteht die Möglichkeit, die Suche abzuspeichern bzw. die Ergebnisse in die lokale Datenbank zu übernehmen. PubMed ist eine bibliographische Datenbank für Medizin und Biologie. Dementsprechend einfach ist die formale Erfassung (PubMed kennt nur zwei Veröffentlichungstypen: Bücher und Zeitschriftenartikel). Dies unterscheidet PubMed von elektronischen Katalogen der Universitätsbibliotheken, die vielfältige Veröffentlichungen in ihrem Bestand besitzen. Diese besondere Aufgabenstellung von Pubmed macht das Konzept der Klassifizierung und einer Retrieval-gerechten Speicherung der Daten dementsprechend einfach. Auch die inhaltliche Erschließung, die teils vom Verlag (z. B. Abstract) mitgeliefert teils von PubMed erarbeitet wird („Thesaurus-gebundene“ Verschlagwortung), unterscheidet diese Datenbank von anderen wissenschaftlichen Bibliotheken,

89 Die vollständige Liste aller Kürzel wird dokumentiert unter <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helppubmed.table.pubmedhelp.T44> (Stand: 2.1.2011)

die außer der formalen Beschreibung nur Schlagworte und (seltener) Klassifikation als Inhalt beschreibende Daten erfassen. Ähnliche fachspezifische Datenbanken werden für verschiedene Disziplinen von vielen – meist aus dem wissenschaftlichen Umfeld stammenden – Anbietern gepflegt.

Eine interessante Entwicklung bei PubMed stellt das eigene XML-Format für Metadaten⁹⁰ dar, welches den Verlagen ermöglicht, per FTP Dateien mit Metadaten zu Veröffentlichungen auf den Server zu laden und in die Datenbank zu übernehmen. (Dies kann sogar in mehreren Schritten – je nach Vollständigkeit der Daten – geschehen.) Die Nutzung des MedlineCitationSet – eines sehr detailreichen Formates für Artikel-bezogene Metadaten – scheint noch sehr gering zu sein: Im Jahr 2010 wurde dieses Format von keinem Literaturverwaltungsprogramm verarbeitet. Die Metadaten in diesem Format kann man durch eine einfache Recherche auf den Internetseiten des MEDLINE (die Möglichkeit einer „XML“-Darstellung in eben diesem Format) oder eine direkte Anfrage an ein CGI-Programm auf eben diesem Server⁹¹ beziehen.

Abschließend kann festgestellt werden, dass das von PubMed verwendete Protokoll sehr gut durchdacht ist. Es ist einfach zu implementieren, ermöglicht eine Begrenzung der Anzahl der Ergebnisse (was bei langsamen Internetverbindungen z. B. in Entwicklungsländern technisch notwendig sein

90 Eigentlich handelt es sich hier um zwei XML Formate, die einander sehr ähnlich sind: ein Format für die Titelmeldung (von MEDLINE selbst entwickelt) und eines für die Repräsentation der Daten der „National Library of Medicine“ (darunter auch PubMed). Eine für Menschen lesbare Beschreibung des ersten („Journal Publishing Tag Set“) befindet sich unter <http://www.ncbi.nlm.nih.gov/entrez/query/static/spec.html> (Stand: 2.1.2011). Das auf die Erstellung von Quellenangaben ausgelegte „MedlineCitationSet“ wird genauer beschreiben unter: http://www.nlm.nih.gov/bsd/licensee/data_elements_doc.html (Stand: 2.1.2011). Beide Formate kennen nur einen bibliographischen Datentyp: den Artikel („Journal“ ist dabei nur eine 'Eigenschaft' des Artikels). Wegen dieser Begrenzung sollen diese beiden Formate nur erwähnt aber nicht genauer beschreiben werden.

91 Die URL lautet: <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi>

kann), bietet Hilfe bei der Eingabe der Suchbegriffe in Form von Vorschlägen aus dem Thesaurus und der Personennamendatei, enthält – sofern vorhanden – Links zu den Volltexten und nutzt ein einfach zu verarbeitendes Datenformat. Trotzdem kommt es als Ersatz für das Z39.50-Protokoll bei den meisten Universitäts- und wissenschaftlichen Bibliotheken nicht in Frage, da es zu spezifisch ist und nur wenige Veröffentlichungsarten kennt. Der Erfolg dieses pragmatischen Vorgehens wird – wie im Folgenden ersichtlich wird – zum Vorbild für viele andere Formate und Protokolle.

Name	Art	Nutzungsgebiet und Funktion	Redundanz	Datentypenanzahl	Erweiterbarkeit	Unicode-Einsatz	Akzeptanz und Verbreitung
MARC	textbasiert, spezifisch	Kataloge: Replikation und Austausch / Information	sehr hoch	hoch	nicht erweiterbar (Zahl der Datenfelder ausgeschöpft)	möglich (Probleme in der Praxis)	hoch, international
MAB	textbasiert, spezifisch	Kataloge: Replikation, seltener Information	hoch	hoch	nicht erweiterbar	möglich (Probleme in der Praxis)	gering, nicht international
RIS	textbasiert, spezifisch	persönliche Literaturverwaltung: Austausch, Zitate	hoch	niedrig: nur zitatbezogene Daten sowie Schlüsselwörter und Anmerkung	theoretisch möglich	möglich (Probleme in der Praxis)	hoch (besonders in Sozialwissenschaften)
BibTeX	textbasiert, spezifisch	persönliche Literaturverwaltung: Zitate, seltener Austausch	hoch	sehr niedrig: nur zitatbezogene Daten sowie Anmerkung	theoretisch möglich	ja	hoch (in Naturwissenschaften und Informatik)
XML	textbasiert, allgemein	unspezifisch: Speicherung, Austausch	gering	(beliebig viele definierbar und erweiterbar)	beliebig	ja	sehr hoch, international
Datenbanken	binär, allgemein	unspezifisch: Speicherung und Retrieval	sehr gering	(beliebig viele definierbar)	beliebig	ja	sehr hoch (besonders relationale DBMS)

Tabelle 1: Zusammenstellung wesentlicher Merkmale vorgestellter Datenformate im wissenschaftlich-bibliographischen Bereich

3. Vorhandene bibliographische Softwaresysteme

Wenn man von bibliographischen Systemen spricht, bezieht man sich auf einen nicht näher definierten Begriff. Dieser reicht von Karteikarten-Katalogen über Bibliographien bis hin zu Z39.50-Gateways und Software auf den Desktop-Rechnern der Benutzer. Die Funktion aller bibliographischen Systeme ist die Einordnung der einzelnen bibliographischen Einheiten (wie: Buch, Artikel oder Fortsetzungswerk) anhand vorgegebener Regeln, die das Wiederfinden (sowie weiteren Umgang damit) erleichtert. Der Nachweis der Existenz eines Werkes ist in einem solchen System wichtiger als die in RAK-WB formulierte Funktion des Nachweisens, ob ein Werk (sowie dessen bestimmte Ausgabe) in Bestand einer Bibliothek vorhanden ist⁹². Die Regeln für die Einordnung können unterschiedlich sein und werden je nach Nutzungsgebiet in unterschiedlicher Reihenfolge miteinander kombiniert. So werden zum Beispiel Bücher entsprechend ihrem Inhalt in ein Regal eingeordnet. Als Einordnungsregel zweiter Reihe kann das Erscheinungsjahr oder die Signaturennummer verwendet werden. Dies erscheint dem Benutzer oft selbstverständlich – zumindest zu dem Zeitpunkt, an dem er ein Werk

92 Vgl. Deutsches Bibliotheksinstitut (Hrsg.): *Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken*. 2., überarbeitete Ausgabe. Berlin, 1993. - ISBN 3-87068-436-4, § 101.

nicht dort findet, wo er es erwartet hatte. In Wirklichkeit ist das die große Leistung der Bibliothekare, die in einem eigenen Wissenschaftszweig diese Regeln erforschen und erstellen. Eine solche Systematik ist zum Beispiel die von Dewey im XIX. Jahrhundert entworfene und bis heute weiterentwickelte Dezimalklassifikation⁹³. Sie teilt das ganze Wissen sowie jedes darunter eingeordnete Wissensgebiet in Gruppen ein, die mit Hilfe einer Zahl repräsentiert werden. Trotz ihrer Schwächen wie der starken Monohierarchie, die vor allem bei interdisziplinären Publikationen ersichtlich wird, ist sie in der Lage, die unterschiedlichsten Werke sehr genau einem bestimmten Sachgebiet zuzuordnen. Damit wird DDC (DK) zu einem homogenen und verlässlichen System, der sowohl aus der Sicht eines Benutzers als auch eines Bibliothekars viele Vorteile bietet.

Der alphabetische Katalog bot dem Benutzer bis zur Einführung von OPACs den besten Einstieg in die Suche im Bibliotheksbestand. Sein Vorteil besteht in der Inhalt-unabhängigen Einordnung eines Werkes. Die dem Benutzer bekannten Angaben wie der Titel oder der Name des Autors erlauben ein schnelleres Finden eines Exemplars unter tausenden von Werken. Die davon abgeleitete elektronische Form erleichtert die Suche zusätzlich, da hier die Haupt- und Nebeneintragungen⁹⁴ keine Rolle mehr spielen.

Während die Entwicklung der „analogen“ bibliographischen Systeme stehen blieb, gewann die elektronische Form dieser durch den rasanten Fortschritt

93 Dewey Decimal Classification (DDC, <http://www.oclc.org/dewey/>) gewann durch die Anhängeszahlen die Möglichkeit einer eingeschränkten Facettierung und ist durch ihre sehr feine Gliederung sowie allgemein verständliche Ziffernotation international akzeptiert worden. Die deutsche Übersetzung befindet sich unter: <http://www.ddc-deutsch.de> und wird unter dem Patronat der Deutschen Nationalbibliothek gepflegt. Vgl. dazu *Dewey-Dezimalklassifikation* sowie *Dezimalklassifikation* In: *KSS-Glossar*. S. 25

94 Die Haupt- und Nebeneintragungen sind ein Relikt der Zeit der Karteikartenkataloge. Diese werden zunehmend durch elektronische Datenbanken verdrängt, bei welchen die Zahl der Autoren, ihre Reihenfolge oder die Vollständigkeit des Titels nicht von Bedeutung sind. Somit scheint ein großer Teil der Regeln der RAK obsolet geworden zu sein.

im Computer- und Netzwerkbereich immer mehr an Bedeutung⁹⁵. Mit der im Jahre 1988 von der NISO⁹⁶ verabschiedeten Norm Z39.50 entstand ein erstes elektronisches Informationssystem, welches die Suche in mehreren unterschiedlichen Datenbanken erlaubte. Dieser Standard wurde schon sehr früh von der Library of Congress umgesetzt und als elektronischer Katalog benutzt. Heute wird Z39.50 ausschließlich zu diesem Zweck verwendet. Als Schwachstellen des Z39.50 gelten: die mangelnde Unterstützung der internationalen Zeichensätze (wie Unicode), die Austauschformat-bedingte Begrenzung von Datenfeldern und die kaum verwendete Möglichkeit der Verknüpfung einzelner Werke untereinander. Mit ZING⁹⁷ entstand eine Initiative, welche sich die Weiterentwicklung und die Anpassung des 20 Jahre alten Standards an die heutigen Protokoll- und Dateiformate zum Ziel gesetzt hat. Die Neuentwicklung hielt unkritisch an zwei Prinzipien fest: dem starken Bezug auf den Bibliotheksbestand und dem unidirektionale "Informieren" an Stelle eines echten bidirektionalen Austausches und einer Personalisierung⁹⁸.

95 Zur Entwicklung des Rechercheverhaltens wissenschaftlicher Bibliotheksnutzer seit dem Jahr 2000 (und der Internetnutzung im Besonderen) vgl. Havemann, Frank; Kaufmann, Andrea: *Der Wandel des Benutzerverhaltens in Zeiten des Internet – Ergebnisse von Befragungen an 13 Bibliotheken*. In: *Festschrift für Walther Umstätter*; CD-ROM

96 Die National Information Standards Organisation (NISO, <http://www.niso.org>) ist Teil des American National Standards Institute (ANSI, <http://www.ansi.org>) und entspricht in etwa dem deutschen Normenausschuss für Bibliotheks- und Dokumentationswesen (NABD, NA009, <http://www.nabd.din.de>) des Deutschen Instituts für Normung e. V. (DIN, <http://www.din.de>). Ihre Aufgabe besteht in der Erstellung von Normen für die Identifikation und Verwaltung von Informationen.

97 ZING steht für die Initiative: Z39.50 International Next Generation, die vom Betreuer des Projektes – der Library of Congress – in SRU (Search / Retrieve via URL, <http://www.loc.gov/standards/sru/>) umbenannt worden ist.

98 Bestandsunabhängige Meta-Kataloge wie Worldcat.org oder Europeana.eu ähnlich wie das Verzeichnis Lieferbarer Bücher (VLB) in Deutschland setzen bereits die Neuentwicklungen ein, bleiben dennoch dem Server-Client-Modell (Authentizität-Prinzip) treu.

An dieser Stelle knüpfen persönliche bibliographische Softwaresysteme wie Reference Manager, EndNote oder Citavi an. Sie bieten dem Benutzer die Möglichkeit, die gesammelte Literatur lokal auf dem Desktop-Rechner zu speichern und zu verwalten, unabhängig davon, ob man im Besitz eines Buch-Exemplars ist oder nicht. Mit Hilfe von Schlagwörtern und kurzen Notizen kann auch eine größere Menge an Quellen sehr effektiv verwaltet werden. Die meisten Programme dieser Art beschränken sich jedoch auf die Übernahme bibliographischer Daten aus Z39.50-Gateways und Erstellung von Quellenangaben. Die Qualität dieser Daten schlägt sich in der Qualität der Quellenangabe nieder: Wurden sie fehlerhaft (falsch oder unvollständig) erfasst, so muss der Benutzer eingreifen und sie korrigieren.

In diesem Teil der Arbeit werden Software-Lösungen dargestellt, die in Interaktion mit dem Benutzer die bibliographischen Metadaten verarbeiten. Der Begriff „Softwaresysteme“ ist nicht unbedingt mit dem Begriff „Programm“ gleichzusetzen. Bei Zotero handelt es sich z. B. um ein sog. Plugin (optionaler [unselbständiger] Bestandteil eines anderen Programms). In anderen Fällen handelt es sich um eine Lösung, die aus Dateien und Hilfsprogrammen besteht oder um eine Ansammlung von Serverprogrammen, die in der Gesamtheit ein solches System ergeben. Für bessere Übersicht wird in dieser Arbeit zwischen „lokalen“ Systemen (d. h. solchen, die keine Daten über das Internet übertragen und sich auf die Verarbeitung lokal vorhandener Daten begrenzen) und Systemen, die Daten aus dem Internet beziehen können, unterschieden. Beide besitzen zwei Eigenschaften: die Verwaltung von Quellen und die Verarbeitung dieser zu Quellenangaben. Da dies auf die elektronischen Kataloge (Z39.50-Gateways der Bibliotheken) nicht zutrifft, werden sie hier nur als „Datenlieferanten“ behandelt.

3.1. Lokale bibliographische Softwaresysteme

Als lokale bibliographische Softwaresysteme wird in dieser Arbeit jede Art von Software bezeichnet, die bestimmte Arten von lokalen (d. h. auf dem

Rechner des Benutzers befindliche) Literaturquellen-Datenbanken verarbeitet. Dies ist auch bei solcher Software der Fall, die die Daten von den elektronischen Katalogen bezieht. Wichtig ist die ausschließliche Speicherung und Verarbeitung auf dem Rechner des Benutzers (z. B. PC). Ob die Daten in einer binären Datenbank (wie bei Reference Manager oder OpenOffice) oder in einem Textformat abgelegt werden, ist hier nicht von Bedeutung.

In diesem Kapitel sollen die Fragen zu dieser Art von Software zunächst theoretisch, später am Beispiel von BibTeX und OpenOffice, diskutiert werden. Es geht um die grundsätzliche Frage: Welche Möglichkeiten bietet diese Art von Software dem Benutzer? Wo sind die Grenzen? Hat sie Vorteile oder Nachteile gegenüber einer Speicherung und Verwaltung auf einem Server?

3.1.1. Vorteile lokaler Speicherung und Verarbeitung

Seit Personal Computer (PC) durch immer höhere Leistungsfähigkeit die Mainframe-Terminal-Lösung für das wissenschaftliche Arbeiten überflüssig machten, wird immer stärker auf lokale Datenverarbeitung gesetzt. Gegenwärtig ist es üblich, die Daten der Benutzer auf dem PC (oder einem Netzlaufwerk) zu speichern und zu verarbeiten. Die Verarbeitung geschieht lokal d. h. auf einem bestimmten Rechner mit der auf diesem PC installierten Software. Dies hat den Vorteil, dass Berechnungen, die viel Zeit benötigen, nur auf dem Rechner des Benutzers ausgeführt werden und die anderen Benutzer nicht benachteiligt. Dies kann der Fall sein, wenn eine größere Menge bibliographischer Daten durch die Software neu geordnet werden muss.

Der größte Vorteil einer lokalen Speicherung besteht in der permanenten Verfügbarkeit über die Daten. Der Autor einer wissenschaftlichen Arbeit ist zu jedem Zeitpunkt unabhängig von der Internetverbindung und kann auf die Literaturliste zugreifen. Je nach Möglichkeiten des Programms und dem Format für die Aufbewahrung der Daten kann er die Literatur sortieren, gruppieren (klassifizieren oder verschlagworten) und mit Notizen versehen.

Es ist ihm jederzeit möglich, neue Werke in die Sammlung aufzunehmen – auch solche, die in den elektronischen Katalogen nicht erfasst werden: Wie z. B. Artikel oder Briefe aus dem Nachlass eines Schriftstellers. Des Weiteren ist es möglich, an übernommenen Daten Korrekturen vorzunehmen und sie dauerhaft zu speichern.

Bei einer lokalen Verarbeitung der bibliographischen Daten zu Quellenangaben kann es notwendig sein, den Zitierstil einer gewissen Norm anzupassen. Man realisiert es meistens durch eine Änderung eines vorhandenen Stils oder durch Erstellung eines neuen. Dies ist bei fast allen lokalen Programmen der Fall⁹⁹. Solche Vorlagen bzw. Skripte werden als eigene Dateien abgelegt. Dies ist bei einer Server-seitigen Lösung nur selten möglich. Will man DIN 1505-2 gerecht zitieren, muss man bei fast allen Programmen (mit Ausnahme von BibTeX und Citavi) eine passende Vorlage erstellen. Diese Tatsache hängt damit zusammen, dass die proprietäre Software meist aus dem englischsprachigen Bereich stammt und sich an den internationalen (englischen) wissenschaftlichen Normen¹⁰⁰ orientiert. Die speziellen Bedürfnisse der deutschen Anwender werden dabei nicht berücksichtigt.

Ein weiterer Vorteil der lokalen Verarbeitung ist die Einbindung der Literaturverwaltungssoftware in den Texteditor. Der Prozess des Verfassens einer wissenschaftlichen Arbeit wird auf diesem Wege unterstützt. Die Literatur kann mit wenigen Mausklicks bereits in dieser Phase in Form von Fußnoten oder sog. Cite-Key (später vollständig in die Bibliographie) übernommen werden. Wurden die Angaben zu den Werken in der

99 Vgl. dazu Wiegand, Dorothee : *Gut zitiert ist halb geschrieben*. In: *c't* (2006), Nr. 7, S. 160-165

100 Wie bereits erwähnt meint eine „Norm“ im Sinne einer Zitiernorm eine in der Gemeinschaft der Wissenschaftler als richtig empfundene und von allen verwendete Art und Weise der Einordnung der bibliographischen Angaben. Diese enthält oft Abkürzungen für Serien-Titel oder Definitionen zur Erstellung der Schlüssel (cite-key). In den meisten Fällen stammt sie aus einem Journal, der auf gewissem Gebiet federführend ist. Die im Reference Manager verwendete Vorlagen werden nach diesen Zeitschriften benannt.

Literatursammlung verändert, können unter Umständen auch die Fußnoten und das Literaturverzeichnis mit diesen neuen Angaben neu erstellt werden. Bei der Übernahme vorformatierter Quellenangaben aus dem Internet (z. B. von Wikipedia) ist es notwendig, jeden betroffenen Eintrag nachträglich im Texteditor zu korrigieren. (Dies trifft jedoch nur auf die Übernahme von Zeichenketten zu, die den Inhalt nicht weiter typisieren.)

Im Hinblick auf den Austausch von Literatursammlungen kann man sagen, dass der Unterschied zwischen der lokalen und der Server-seitigen Speicherung und Verarbeitung immer geringer wird. Server-Lösungen werden gegenwärtig um die Möglichkeit erweitert, Daten von elektronischen Katalogen in private Sammlungen zu übernehmen und bei Bedarf in eine Datei zu exportieren¹⁰¹. Über solche Export-Filter verfügten noch bis vor wenigen Jahren nur die PC-Programme. Die Export- und Import-Funktionen der Literaturverwaltungssoftware erleichtern die Zusammenarbeit mehrerer Autoren, indem sie zur Erstellung eines Pools gemeinsam genutzter Veröffentlichungen verwendet werden. Je nach Vorgehensweise und verwendetem Dateiformat kommt es beim Austausch solcher Sammlungen zu erheblichen Datenverlusten (z. B. durch fehlende Felder für Notizen, Schlagwörter oder Abbildung von Hierarchien).

3.1.2. Schwachstellen lokaler Speicherung und Verarbeitung

Die Möglichkeit, eine private Sammlung zusammenzustellen, sie zu bearbeiten, zu korrigieren und anderen Menschen zur Verfügung zu stellen, birgt einige Schwachstellen. Sie kann zu einer sog. Insellösung führen, in der wichtige Metadaten nur auf dem Rechner genutzt werden können, auf dem sie hinzugefügt worden sind. Solche Metadaten geben die Benutzer oft im Verlauf der Sichtung und Sortierung der Quellen ein. Einige von ihnen wie z.

¹⁰¹ Diese Aussage trifft in geringem Maße auf die Import-Funktionen solcher Server-seitigen Online-Lösungen zu. RefWorks beherrscht lediglich drei: RIS-, SciFinder-Tagged- und das eigene RefWorks-Format. Mangels eines Import-Filters für BibTeX wird RefWorks in Arbeitsgruppen der Naturwissenschaften wohl kaum Verwendung finden. (Stand: Juli 2009, RefWorks Version 1)

B. Klassifikation oder Serien (als virtuelle Ebene für die darin erschienenen Einzelwerke) können von keinem der üblichen Datenaustauschformate aufgenommen werden. Damit wird das Datenformat zum „Flaschenhals“, der viele für das Retrieval interessanten Daten zurück hält. Die geordnete, mit verschlagworten und Kommentaren versehene Literaturliste, kann während der Übertragung unvorhersehbar auf rudimentäre Datenfelder gekürzt werden. Um dieser Situation vorzubeugen, müssten neue bibliographische Datenformate erfunden werden, die nicht nur die üblichen Titeldaten im Blick haben, sondern auch die Bedürfnisse der Benutzer.

Der Austausch des PCs oder des Datenträgers kann ebenfalls zu einer Situation führen, in der die erfassten Daten möglichst vollständig übertragen werden müssen. In den meisten Fällen reicht das Kopieren der Benutzerdaten. Dies führt praktisch nur dann zum Erfolg, wenn die verarbeitende Software in derselben oder einer kompatiblen Version vorliegt.

Die meist kostspielige Literaturverwaltungssoftware¹⁰² kann je nach Lizenz auf einem oder mehreren Rechnern installiert werden. Besitzt man eine sog. Einzelplatzlizenz, darf man dieses Produkt auf keinem anderen Rechner installieren. Diese Begrenzung kann für den Benutzer nachteilig sein, wenn der Computer mit dem Programm defekt wird oder wenn der Benutzer jedes Mal einen anderen Arbeitsplatz zugewiesen bekommt. Dieser Nachteil tritt meist nur bei proprietärer Software auf. Die Open-Source-Software¹⁰³ kann beliebig oft installiert und weitergegeben werden. Die Verwendung der

102 Zu Lizenzmodellen, Funktionsumfang und Preisen vgl. Kerschis, Annett : *Literaturverwaltung und Wissensorganisation im Vergleich. Das Angebot von Literaturverwaltungsprogrammen und Social Bookmarking in Bezug auf die Benutzbarkeit in Bibliotheken*. Potsdam, Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Diplomarbeit, 2007. Da keine größeren Veränderungen auf diesem Gebiet zu verzeichnen sind, bleiben diese Zahlen auch 2011 repräsentativ.

103 „Open source is a development method for software that harnesses the power of distributed peer review and transparency of process. The promise of open source is better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in.“ (aus: Open Source Initiative (Hrsg.) : *Mission | Open Source Initiative*. URL: <http://www.opensource.org/> Stand: 13.08.2011)

Formate BibTeX und RIS ist in dieser Situation vorteilhaft, da sie von vielen Open-Source-Programmen verarbeitet werden können – die binären Datenbanken von Reference Manager oder EndNote können hingegen nur von eben diesen Programmen gelesen und bearbeitet werden.

3.1.3. Umsetzung

Am Beispiel von BibTeX und OpenOffice soll gezeigt werden, wie lokale Verarbeitung bibliographischer Daten realisiert wurde. Dabei soll der Aspekt der Handhabung berücksichtigt werden.

3.1.3.1. BibTeX

Das Wort BibTeX meint sowohl das Programm als auch das dafür verwendete Datenformat. Es kann je nach gewünschter Ausgabe-Form unterschiedlich verwendet werden: als Bestandteil von LaTeX innerhalb eines Textdokumentes eingebunden (in den Fußnoten oder für die Erstellung des Literaturverzeichnisses) oder als selbständige Version mit Ausgabe in eine HTML-Datei (bzw. -Dateien). In beiden Fällen handelt es sich um Programme, die auf der Kommandozeile (explizit oder implizit z. B. durch LaTeX) aufgerufen werden, aber keine graphische Oberfläche besitzen – sie können vom Benutzer unbemerkt ablaufen. Diese Tatsache hat auf die hier darzustellenden Themen keinen Einfluss, ist jedoch als spezifisches Merkmal der Vollständigkeit halber zu erwähnen.

3.1.3.1.1. Speicherung und Bearbeitung

Beim Aufruf von BibTeX wird dem Programm als sog. "Argument" der Dateisystem-Pfad zur Datei übergeben, die die Quellensammlung enthält. Sie wird indexiert, um einen schnelleren Ablauf und die Auflösung von Abhängigkeiten sicherzustellen. Die BibTeX-Dateien können in einem Texteditor erstellt werden. Trotz der geringen Zahl der Veröffentlichungstypen ist die Kenntnis der möglichen bzw. notwendigen Felder für den Benutzer eine Hürde. Aus diesem Grund entstanden mehrere Programme, die genau diese Aufgabe erleichtern sollen. Nicht alle bieten ausreichend Hilfe bei der Einordnung des Typs und die wenigsten prüfen die

Vollständigkeit (die Existenz aller Pflichtfelder, Auflösung von Referenzen bzw. Abhängigkeiten). Aufgrund hoher Popularität dieses Formats ist es sogar möglich, in fast allen Literaturverwaltungsprogrammen mit diesen Dateien zu arbeiten (entweder durch Öffnen und Speichern oder durch Import und Export). Das Zusammenfassen mehrerer Sammlungen ist hingegen sehr einfach: es reicht ein einfaches Zusammenfügen solcher Dateien (sog. Konkatenation) oder Import mehrerer Dateien in das Bearbeitungsprogramm. Auch die gleichzeitige Verwendung mehrerer BibTeX-Dateien im LaTeX ist möglich. Ein Problem tritt erst dann auf, wenn ein Schlüssel mehrmals vorkommt. In diesem Fall – ähnlich wie beim Fehlen von referenzierten Werken – gibt ein Aufruf von BibTeX sehr präzise Fehlerwarnungen aus und bricht die Verarbeitung ab.

Bei der Erfassung von Literatur kann der Benutzer einen einfachen Texteditor einsetzen. Seit einigen Jahren kann LaTeX und BibTeX Unicode verarbeiten – die Escape-Sequenzen, die man zur Eingabe von Sonderzeichen in eine ASCII-kodierte Datei verwendet hat (z. B. '\u' für 'ü'), können jedoch nur dann ausgelassen werden, wenn der Texteditor Unicode beherrscht und die Datei so kodiert abgespeichert wird. Neben der Unvollständigkeit der Angaben und dem Fehlen der referenzierten Werke, gibt es die Probleme der Tipp-Fehler und der unterschiedlichen Schreibweisen von Personennamen, die das Retrieval erschweren. Diese Schwachstelle ist ein grundsätzliches Problem der Erfassung. Eine praktikable Lösung bietet z. B. Reference Manager, der die Eingaben mit einer Personennamendatei (bereits erfasster Autoren) abgleicht.

3.1.3.1.2. Retrieval

Die Suche nach Personennamen oder Titeln ist bei BibTeX-Dateien sehr einfach. Die Präzision kann vor allem aufgrund unterschiedlicher Schreibweisen sehr niedrig ausfallen. Die Verwendung von regulären Ausdrücken (Platzhaltern) kann den Recall erhöhen. Einige wenige Texteditoren, die diese Art der Suche beherrschen, unterstützen den Benutzer bei der Erstellung passender Suchanfragen. Der Ausdruck „Ha[m]

+acher J.+“ findet alle Vorkommen der Nachnamen „Hamacher“ und „Hammacher“ mit einem Vornamen, der mit „J“ anfängt. Eine Suche über mehrere Felder hinweg ist nur eingeschränkt möglich. Zum Beispiel erlaubt „author\s*=\s*\{.*\sHammacher\s.*\}\s*\n?\r?year\s*=\s*\{s*2005\s*\}“ eine Suche nach einem Datensatz, der im Feld „author“ den Wert „Hamacher“ besitzt und in der darauf folgenden Zeile das Feld „year“ mit dem Wert „2005“ enthält. Diese Art der Suche ist exakter als eine Trunkierung. Sie ist einem durchschnittlichen Benutzer jedoch kaum zuzumuten, weshalb ein Retrieval mit regulären Ausdrücken nur Spezialisten vorbehalten bleibt. Durch die Verwendung von Unicode ist eine einfache Suche nach Namen in nichtlateinischer Schrift hingegen ein Vorteil, den proprietäre Programme erst in den neusten Versionen bieten.

Die Gruppierung von Werken anhand von Schlüsselwörtern kann durch die Beschlagwortung im Feld „keyword“ geschehen. Das Feld "key" ist dafür vorgesehen, zusätzliche Sortierschlüssel aufzunehmen (wenn ein solches nicht automatisch generiert werden kann oder die Reihenfolge der Sortierung beeinflusst werden soll). JabRef – ein leistungsfähiges und Betriebssystem-unabhängiges Programm in Java – speichert die Angaben zur Gruppe in einem Pseudo-Datensatz mit dem Namen „comment“. Zugleich bildet es Indexe anhand der enthaltenen Metadaten, indem es alle Datensätze in allen oder nur bestimmten Datenfeldern (keyword, titel, author, editor, abstract, annote, note) nach einem Begriff durchsucht. Im Gegensatz zum Speichern im comment-Bereich stehen solche Gruppen nur in diesem Programm dem Benutzer zur Verfügung und können nicht exportiert werden. JabRef bietet bisher keine Möglichkeit zur Suche mit booleschen Operatoren in mehreren Datenfeldern.

Nicht nur für das Retrieval, sondern auch für die die Dokumentation der Einträge, können die Felder „note“, „annote“ und „abstract“ verwendet werden. Vor allem „annote“ und „abstract“, die beim Zitieren nicht benutzt werden, können nützliche Informationen enthalten: Warum wird dieses Werk berücksichtigt? Welche besonders interessanten Inhalte sind darin zu

finden? Antworten auf diese Fragen enthalten meist Stichwörter, die beim Retrieval zumindest den Recall erhöhen.

3.1.3.1.3. Normgerechte Stilvorlagen

Wie bereits erwähnt, besitzt BibTeX eine Macro-Sprache¹⁰⁴, die dazu verwendet wird, Stilvorlagen zu erstellen. Diese enthalten Übersetzungen solcher Begriffe wie „seite“, „in“, „Hrsg.“, Platzhalter für Daten bestimmter Datenfelder und Verarbeitungsbefehle wie z. B.: „der zweite Vorname wird abgekürzt“ oder „vor jedem weiteren Autor oder Herausgeber wird ein Semikolon als Trenner verwendet“.

Wenn man BibTeX in einem LaTeX-Dokument verwendet, kann man den Zitierschlüsseltyp durch Angabe einer Stil-Datei festlegen. Die Einträge der Bibliographie können sortiert oder unsortiert, abgekürzt oder ausgeschrieben ausgegeben werden. Dieses Verhalten wird von einer solchen Datei realisiert. „Plain“ steht für numerische Schlüssel, welche nach Namen des ersten Autors und dem Erscheinungsjahr eingeordnet werden. „Unsort“ belässt die Quellen in der Reihenfolge, in der sie in der BibTeX-Datei vorkommen. „Harvard“ erzeugt Quellenangaben nach dem in Harvard (und mittlerweile international akzeptierten) üblichen Stil „(Nachname Jahr)“. Unter der Federführung von Prof. Klaus Lorenzen entstanden auch DIN 1505-2 (und -3) konforme Stilvorlagen. Sie tragen die Namen „geralpha“, „gerplain“, „gerabbrv“ und „gerunstr“. Ebenso existieren Stile für die polnische, dänische, norwegische aber auch juristische Art und Weise der Zitierung.

Da die BibTeX-Dateien, die in einem LaTeX-Dokument verwendet werden, mehr Literatur enthalten können, werden im Literaturverzeichnis nur solche Werke aufgelistet, die auch im Dokument ausdrücklich zitiert worden sind. Wurde ein Werk nicht zitiert, aber beim Verfassen der Arbeit verwendet, kann es mit „\nocite{*Schlüssel*}“ in dieses Verzeichnis aufgenommen werden.

104 Diese Macro-Sprache ist sehr leistungsfähig (z. B. automatisches Generieren von Zitierschlüsseln mit entsprechender Sortierung des Literaturverzeichnisses) und dadurch auch sehr komplex.

In einer Gesamtschau fällt bei BibTeX auf, dass seine Möglichkeiten noch nicht ausgeschöpft worden sind. Ein gut durchdachtes und nahezu universelles Format eröffnet mit einem leistungsfähigen Editor und einer ebenso starken Macrosprache einen Weg, eine mittelgroße Quellensammlung – wie sie oft in Diplomarbeiten vorkommt – mit kleinstem Aufwand in eine wissenschaftliche Arbeit einzubinden. Die Austauschbarkeit der Daten, die Popularität des Formats und die vielseitige Verarbeitung steht dabei der Redundanz der Daten, der Komplexität der Datenfelder und dem mangelhaften Retrieval entgegen. Als kostenfreie Open-Source-Alternative kann sie von jedem weiterentwickelt und angepasst werden.

3.1.3.2. OpenOffice

Das OpenOffice¹⁰⁵ ist ein Paket von Büroprogrammen, welches zum größten Konkurrenten von Microsoft Office wurde. Es ist – wie BibTeX – OpenSource und setzt seit der Version 2 auf das XML-basierte Format „Open Document“. Als kostenlose Alternative mit einer aktiven Entwicklergemeinschaft wird OpenOffice im universitären Umfeld eingesetzt.

Die aus StarOffice stammende Datenbank-Komponente besitzt eine Schnittstelle zum Texteditor, welche es ermöglicht, eine binäre Datenbank als Speicherort für die eigene Literatursammlung zu nutzen. Die Datenbank (genannt Datenquelle) kann frei gewählt werden. Neben der eingebauten Datenbank steht eine zusätzliche Schnittstelle zur Einbindung anderer Datenbanken zur Verfügung. Die bibliographischen Daten können z. B. in einer MySQL-Datenbank über ein Netzwerk abgespeichert werden¹⁰⁶. Auf diese Art und Weise ist eine gemeinsame Nutzung der Sammlung möglich. In den meisten Fällen wird eine lokale Datenquelle verwendet d. h. OpenOffice benutzt die mitgelieferte HSQL-Datenbank (Bestandteil von

105 Früher StarOffice der Hamburger Firma StarDivision. Seit der Übernahme der Firma Sun durch Oracle existieren zwei Versionen dieser Software: OpenOffice und LibreOffice, die sich im Funktionsumfang wenig unterscheiden.

106 Vgl. dazu: OpenOffice.org (Hrsg.): *OpenOffice Developers Guide*. Kapitel 12.2: Data Sources in OpenOffice.org API. URL: <http://api.openoffice.org/docs/DeveloperGuide/Database/Database.xhtml> (Stand: 12.12.2008)

OpenOffice Base – der Alternative zu Microsoft Access) namens „Bibliography“. Sie besteht aus einer einzigen Tabelle, die folgende Spalten enthält: Identifier (die ID), Type (Art der Veröffentlichung), Address (Adresse), Annote (Anmerkungen), Author (Autor), Booktitle (Buchtitel), Chapter (Kapitel), Edition (Ausgabe), Editor (Herausgeber), Howpulish (Ausgabeart), Institutn (Institution), Journal, Note (Vermerk zur Ausgabe), Organizat (Organisation), Pages (Seiten), School (Hochschule), Series (Serie), Title (Titel), RepType (Art des Berichts), Volume (Band), Year (Jahr), ISBN, URL und Customer1-5 (fünf Felder für den Benutzer). Es fehlt die ISSN und die Zeitschriftennummer, obwohl andere zeitschriftentypische Felder vorhanden sind.

Für die Autoren und Herausgeber sieht OpenOffice jeweils nur ein Feld vor. Sie müssen deshalb so angeordnet werden, dass sie an passender Stelle eingefügt werden können. Obwohl dies bei BibTeX auch der Fall ist, wird dort versucht, die Autoren anhand des Trennwortes „and“ zu erkennen und die Namensbestandteile so anzuordnen, dass sie dem Zitierstil entsprechen. Bei OpenOffice werden die Tabelleninhalte nicht nachbearbeitet. Das Erstellen eines Literaturverzeichnisses in OpenOffice ist vergleichbar einfach, wie beim BibTeX. Da beim OpenOffice keine Vorlagen für Zitierstil existieren, muss der Benutzer bestimmen, in welcher Reihenfolge die einzelnen Felder erscheinen sollen und durch welche Trennzeichen sie voneinander getrennt werden. Für jeden Erscheinungstyp gibt es eigene Regeln, die unter Umständen angepasst werden müssen. Die vorinstallierte Anordnung von Feldern ist an Vorgaben des ISO 690 ausgerichtet.

Ein großer Vorteil der Verwendung einer echten Datenbank als Basis für die Literatursammlung in OpenOffice ergibt sich bei der Suche. Ein Filter-Assistent erlaubt eine mehrdimensionale Suche nach Werken unter Benutzung verschiedener Konditionen (gleich, kleiner, größer, wie, nicht wie, leer, nicht leer) auch im Zusammenhang mit regulären Ausdrücken, welche mit Hilfe boolescher Operatoren verknüpft werden. Das Ergebnis wird als Liste aller Treffer in der „Datenquellen“-Ansicht eingeblendet. Die direkte

Verwendung dieser Ergebnisse für das Zitieren ist nicht möglich. Dies geschieht auf folgendem Weg: im Menü „Einfügen“ wählt man „Verzeichnisse“ und darin „Literaturverzeichniseintrag“ aus. In einer Liste wählt man den Bezeichner (entspricht dem Eintrag der Spalte „Identifizier“ der Tabelle) aus. Eingefügt wird ein Zitierschlüssel (wahlweise in runden oder eckigen Klammern; als Bezeichner oder als Ziffer), der automatisch im Literaturverzeichnis erscheint.

Die zitierten Werke werden im Dokument abgespeichert. Eine spezielle Auszeichnung¹⁰⁷ dieser Quellen sorgt dafür, dass sie mit diesem ausgetauscht werden und im selben Dokument wiederverwendet werden können (man fügt wie oben beschrieben einen Literaturverzeichniseintrag ein, wobei nicht „Aus Literaturdatenbank“ sondern „Aus Dokumentinhalt“ ausgewählt wird). Dieser Ansatz der Übernahme der Metadaten (als solcher) in das Dokument gibt dem Benutzer die Möglichkeit, auch ohne die Datenbank an der Datei arbeiten zu können. Dies ist umso wichtiger, weil der Export und Import der bibliographischen Datenbank relativ umständlich ist.

Das OpenOffice eignet sich als Paket aus Texteditor und bibliographischer Datenbank für die wissenschaftliche Tätigkeit bei einfachen strukturierten Quellen besser als andere Lösungen. Eine mehrdimensionale Suche mit booleschen Operatoren, regulären Ausdrücken und arithmetischem Vergleich von Zahlen sowie die Einfachheit des Zitierens und der Erstellung einer Bibliographie sind seine Stärken. Die Schwächen in der Erstellung von Quellenangaben (Stil) sowie die Ergonomie (Trennung zwischen Suche und Einfügen von Quellen) stellen unnötige Hürden für den (wissenschaftlichen) Benutzer dar. Mit CiteProc¹⁰⁸ – einem neuartigen Ansatz, der in der Suche über SRU/W und der Verarbeitung der empfangen XML-Dateien anhand von

107 Die Metadaten werden in einer Bibliographie im Element "text-index-entry-bibliography-attrs" gespeichert, dessen Attribut "bibliography-data-field" die Daten näher beschreibt. Das optionale Attribut "style-name" spezifiziert einen Zitierstil. Vgl. <http://docs.oasis-open.org/office/v1.1/OS/OpenDocument-v1.1-html/OpenDocument-v1.1.html> Stand: 1.2.2007

108 CiteProc wird auf Seite 130 näher erläutert.

XSLT-Vorlagen besteht – wird der alte und nicht mehr zeitgemäße in einer der nächsten Versionen abgelöst.

3.2. Bibliographische Softwaresysteme mit Datenübertragung über das Internet (nach Server-Client-Modell)

Zusätzlich zu den Funktionen der lokalen Literaturverwaltungssysteme (Speichern, Öffnen, Editieren, Zitieren) bieten einige Programme die Übernahme von bibliographischen Daten aus elektronischen Katalogen an. Solche Kataloge sind z. B. die Z39.50-Gateways oder der PubMed-Server. Diese Funktionalität steht nur während einer Verbindung mit dem Internet zur Verfügung. Je nach verwendetem Protokoll ist die Präzision beim Retrieval unterschiedlich. Gleiches trifft auf die Datenqualität im Bezug auf das Übertragungsformat und den Datenlieferanten zu. So einfach und vorteilhaft dieses Vorgehen zu sein scheint, ist es in der Praxis nicht. Von den Z39.50-Servern kann man meist nur Angaben zu selbständig erschienenen Werken erhalten. Die Pflege von detaillierten Artikelverzeichnissen wäre theoretisch denkbar, wird jedoch aus Kostengründen¹⁰⁹ nicht realisiert. Die Übertragung der Daten geschieht gemäß den Protokolldefinitionen nur in Richtung des Benutzers, so dass die Benutzer kein Recht haben, die Angaben zu einzelnen Artikeln oder Abstracts hinzuzufügen. Dafür gibt es zwei Gründe: Zum Ersten handelt es sich bei diesen Servern um die „Darstellung“ des Bestandes einer Bibliothek, zum Anderen ist das aus Open-Source (bzw. Wikipedia) stammende Community-Prinzip beim Entwurf der Schnittstellen weder bekannt gewesen, noch eine Pluralität erwünscht. Damit bleibt der Server der Datenlieferant, der für die Richtigkeit der Daten steht, und der Client der Empfänger, der Korrekturen lediglich lokal speichern darf.

Neben dem echten Client-Server-Modell (wie bei Z39.50 oder PubMed) ist auch ein ähnliches Verfahren zu berücksichtigen, welches in der Auslieferung von Titeldaten über Suchmaschinen auf Basis von HTML besteht. Da hier der Benutzer im WWW-Browser über eine Suchmaske auf

109 Zum Thema „Erschließungskosten unter Verwendung von RAK-WB“ siehe S. 18

diese Daten zugreift, diese übernimmt und in eigenen Datenbestand einfügt, handelt es sich nicht um ein echtes Client-Server-Modell. Dieses Modell sieht vor, dass Programme miteinander die Parameter aushandeln und Daten austauschen können. Der Umweg über die menschliche Interaktion ersetzt den Automatismus, so dass man von einem „unechten Client-Server-Modell“ sprechen kann (Die Funktionen des Datenlieferanten (Server) und des Empfängers – WWW-Browser bzw. Literaturverwaltungssoftware – bleiben erhalten.) Dieses Verfahren wird oft bei Artikel-Datenbanken im Bereich der Naturwissenschaften eingesetzt. Die Angaben zum Werk werden meist als BibTeX-Datensätze innerhalb einer HTML-Seite ausgeliefert. Sie können direkt in die BibTeX-Datei eingefügt werden oder von einer graphischen Oberfläche importiert werden. Mangels eines eigenen Protokolls für BibTeX ist dieses Verfahren der einzige Weg des Austauschs von bibliographischen Daten in diesem Format und ist daher in diesem Kontext zu berücksichtigen. Analog trifft es ebenfalls auf andere weniger populäre Text-basierte bibliographische Datenformate zu.

3.2.1. Vorteile einer Übernahme bibliographischer Daten

Die Qualität und die Zuverlässigkeit der Angaben bei Daten, die von einem Bibliotheksserver bezogen werden, gehört zu den wichtigsten Vorteilen dieser Lösung – vorausgesetzt, der Server selbst greift auf homogene Daten von hoher Qualität zurück. Die Bibliothekare, die eine solche Datenbank pflegen, müssen nach einheitlichen Kriterien (für Homogenität) und mit sehr viel Hintergrundwissen (das Wissen des Fachgebietes, teilweise Kenntnis bestimmter Fremdsprachen) katalogisieren. Innerhalb der deutschen Bibliotheksverbände werden die Regeln der RAK-WB jedoch unterschiedlich interpretiert¹¹⁰. In den meisten Fällen geht es um zweitrangige Unterschiede, so dass gewöhnlich die Übernahme derselben Daten von verschiedenen Quellen keine große Auswirkungen haben sollte. Für deutschsprachige Werke und deutsche Übersetzungen fremdsprachiger Werke besteht die

110 Vgl. Hilberer, Thomas : *Aufwand vs. Nutzen : Wie sollen deutsche wissenschaftliche Bibliotheken künftig katalogisieren?* In: Bibliotheksdienst 2003, Nr. 6, S. 755

Möglichkeit, den Dienst der Deutschen Nationalbibliothek zu nutzen und die Titeldaten der Deutschen Nationalbibliographie zu übernehmen¹¹¹.

Ein weiterer Vorteil der Übernahme der Daten aus dem Internet wird sichtbar, wenn man die Komplexität und die Vielzahl der bibliographischen Ausdrücke im Bezug auf die Erfassung betrachtet. Vor allem bei ungeübten Benutzern können Begriffe wie „Zusatz zum Titel“ oder „beteiligte Personen“ zu Missverständnissen und falschen Angaben führen, da solche Felder in den Erfassungsmasken der Literaturverwaltungssoftware entweder nicht sichtbar sind oder nicht genügend erklärt werden. Diese Frage der Usability wird in der Zeitersparnis sichtbar.

Die Suche nach Personennamen oder anderen Begriffen, die die Angaben zum gewünschten Werk vom Server abrufen soll, kann einen hohen Recall erzeugen. Diese Unschärfe ist mit einem interessanten Nebeneffekt verbunden, dass der Empfänger dieser Daten darunter andere Werke findet, die ihm nicht bekannt sind aber für seine Arbeit relevant sein können.

3.2.1.1. Deutsche Einheitsaufnahme (CIP)

Die Deutsche Nationalbibliothek erfasst in Zusammenarbeit mit den Verlagen die Daten der erscheinenden Bücher¹¹² noch vor der Veröffentlichung. Dieses Verfahren nennt man „Cataloging in Publication“. Library of Congress definiert es wie folgt:

„A Cataloging in Publication record (aka CIP data) is a bibliographic record prepared by the Library of Congress for a book that has not yet been published. When the book is published, the publisher includes the CIP data on the copyright page thereby facilitating book processing for libraries and book dealers.“¹¹³

111 Diese stehen nur „zahlenden Benutzern“ in einem für Maschinen lesbaren Format (MARC) zur Verfügung. Ein kostenfreier Zugang erlaubt nur die ISBD-ähnliche SUTRS Darstellung.

112 CIP erfasst nur Bücher. Eine Datenbank mit normalisierten Angaben zu Zeitschriften (ZDB) befindet sich unter: <http://dispatch.opac.ddb.de/> (Stand: 20.03.2010)

113 Library of Congress, Cataloging in Publication Division (Hrsg.) : *The Cataloging in*

Die Titeldaten des Werkes werden bereits in der Preprint-Phase von der Deutschen Nationalbibliothek in die zentrale Datenbank eingegeben, so dass sie im Werk abgedruckt werden können. Sie sind ausschlaggebend für die Erfassung dieses Werkes in Bestände anderer Bibliotheken. Auf diesem Wege soll unter anderem garantiert werden, dass das Werk in jeder Bibliothek identisch erfasst wird. Solche Datenbanken werden ausschließlich auf der nationalen Ebene gepflegt und können bei der Erfassung dieser Daten in Bibliotheken anderer Länder problemlos übernommen werden, da sie meist im internationalen ISBD-Standard abgedruckt werden. Ob diese Hilfestellung genutzt wird, hängt jedoch von der Bibliothek (bzw. dem Verbund) und dem Bibliothekar selbst ab.

Die CIP-Daten werden von der Deutschen Nationalbibliothek ebenfalls in der Deutschen Nationalbibliographie genutzt. Diese Daten werden meist von Verlagen und Buchhändlern für die Abwicklung von Bestellungen vor und nach der Veröffentlichung genutzt. Ihre Relevanz für die Literaturverwaltungssoftware ist somit eher gering (reine Titelmeldung). Ihre Zuverlässigkeit und internationale Ausrichtung durch die Nutzung des ISBD-Formats ist vorbildlich. Der direkte Zugriff auf den Z39.50-Server der Deutschen Nationalbibliothek (ILTIS¹¹⁴) ermöglicht die Übernahme qualitativ hochwertiger Daten aus der deutschen Nationalbibliographie.

3.2.1.2. Akzeptanz

Die Bedeutung des Internets für Wissenschaftler und ihre Tätigkeit erschöpft sich nicht in Web-Publikationen und neuen Kommunikationswegen wie Email oder Chat. Immer mehr Dozenten und Studenten benutzen dieses Medium zur Recherche der Literatur¹¹⁵ und Erstellung von Quellenangaben. PubMed

Publication Program. <http://cip.loc.gov/> (Stand: 20.03.2010)

114 Die Datenbank ILTIS der Deutschen Nationalbibliothek beinhaltet neben 5 Millionen Titeln der Deutschen Nationalbibliographie (nach 1945) Musikalien und Tonträger. Eine direkte Übernahme dieser Daten ist nur aus dem MARC-Format möglich.

115 Die sog. "SteFi-Studie" ("Studieren mit elektronischen Fachinformationen" - Ergebnisse zusammengestellt unter http://www.ub.uni-konstanz.de/fileadmin/Dateien/Informationskompetenz/Oberwolfach/wlb_ennen_stefi_vortrag_01.pdf abrufbar) aus

gilt unter Medizinern nicht zuletzt aufgrund seines Umfangs als das bibliographische Informationssystem. Der Erfolg dieser Datenbank liegt darin, dass diese Disziplin vom wissenschaftlichen Austausch auf internationalem Niveau profitiert und Englisch als Fachsprache sehr früh Anerkennung fand. Jede Publikation, die auf Englisch verfasst wurde, ist in der Datenbank vertreten. Der Verzicht auf eine Recherche in PubMed bedeutet in den meisten Fällen mühsame Suche in Katalogen und eine lückenhafte Bibliographie. Alternativ zur speziellen Software für eine Recherche in diesem System kann die Web-Oberfläche verwendet werden. Die auf diesem Wege gewonnenen Daten können zwar nicht zur automatischen Generierung von Bibliographien verwendet werden, sind jedoch für die Verfasser von Artikeln oder Doktor-Arbeiten die Quelle der Wahl im biomedizinischen Forschungsbereich. Mit einem geeigneten Programm wie Reference Manager oder Pybliographer, die später genauer betrachtet werden sollen, kann dies noch effizienter geschehen, indem die gefundenen und für die Arbeit relevanten Werke auf dem Rechner des Benutzers abgespeichert werden und bei Bedarf als Quellenangabe bzw. Bibliographie direkt übernommen werden können. Die größte Hürde für einen weit verbreiteten Einsatz dieser Software sind die Lizenz-Kosten und die verschiedenen Zitierstile und -Normen.

Z39.50 gilt bei den deutschen Universitätsbibliotheken als De-facto-Standard. Programme, die einen Zugriff auf solche Katalog-Gateways bieten sind verhältnismäßig teuer. Kostenlose Versionen dieser Programme werden wegen der begrenzten Funktionalität kaum wahrgenommen. Trotz dieser Hindernisse wird das zwanzig Jahre alte Protokoll zum Informationsaustausch für „gut genug“ befunden und verwendet.

Allgemein kann festgestellt werden, dass der Trend zum Einsatz von Software zwecks Vereinfachung von Abläufen und der Zeitersparnis auch in

2002 weist auf eine verstärkte Recherche nach wissenschaftlicher Literatur im WWW hin. Die Fachdatenbanken (die eine Web-Oberfläche für die Recherche frei im Web anbieten wie z. B. PubMed) und die sog. Discovery-Services (wie Ebsco) werden in dieser Arbeit als Teil des WWW verstanden.

diesem Bereich zur Akzeptanz der beiden Protokolle sowie der Vorgehensweise im Allgemeinen zunehmend beitragen wird.

3.2.2. Schwachstellen einer Übernahme bibliographischer Daten

Eine einfache Recherche und Übernahme der gelieferten Daten führt nicht in allen Fällen zu Norm-gerechten Quellenangaben. Die Ursachen sind unterschiedlich. Neben der Genauigkeit der Erfassung samt Katalogisierungsregeln (wie RAK oder AACR2) spielt das Austauschformat und das verarbeitende Programm auf dem PC des Benutzers eine große Rolle. Damit kann man die Ursachen für mangelhafte Verarbeitung an zwei Orten feststellen: auf der Seite des Datenanbieters / auf dem Server (Software, Regeln für die Erfassung) und beim Benutzer (Genauigkeit der Implementierung von Standards in der Software).

Die meisten Nachteile dieser hier dargestellten Lösungen ergeben sich aus der Komplexität der Materie selbst. Das Z39.50-Protokoll wird oft nur unvollständig implementiert – seine als „optional“ gekennzeichneten Teile sind nicht in allen Programmen realisiert. Einige spezielle Zeichensätze wie ANSEL werden von einigen Programmen offensichtlich falsch verarbeitet.

Auf der Seite des Datenlieferanten sind vor allem die erschließenden Bibliothekare als Ursache für mögliche Fehler zu nennen sind: neben üblichen Schreibfehlern – die wohl am schnellsten auffallen und zu beseitigen sind – muss man mit unbewusst (aufgrund eines Mangels an Fachwissen o. Ä.) gemachten Fehlangaben zum Autor oder Herausgeber vor allem in fremdsprachigen Werken rechnen. In seltenen Fällen kann auch die Erscheinungsweise eines Werkes zu Missverständnissen führen (ein Sonderheft einer Zeitschrift ohne eigene Nummer etc.).

Für den Benutzer wird dieser Nachteil oft erst in der generierten Quellenangabe oder Bibliographie sichtbar. Die Zeitersparnis durch eine direkte Übernahme der bibliographischen Daten von einem Server sinkt, wenn der Benutzer immer eine Korrektur vornehmen muss und dazu einige Minuten pro Datensatz aufwendet. Ein uneingeschränktes Vertrauen in die

Richtigkeit der generierten Bibliographie kann zu unerwarteten Fehlern führen wie z. B.: Schreibweise der Autorennamen, falsche Angaben zum Autor, Titel oder Band-Nummern. Im Text Nr. 11 wird ein Ausschnitt einer Bibliographie dargestellt, in dem solche Fehler aufgetreten sind. Das Programm RefWorks (in Version 1), welches von der Saarländischen Universitäts- und Landesbibliothek für die Angehörigen der Universität lizenziert wurde, sollte auf der Grundlage der Daten des Z39.50-Gateways der Library of Congress eine Literaturliste von Büchern mit dem Titel „Pacem in terris“ erstellen. Die Fehlerquote betrug in diesem Falle 70%! Dieses Ergebnis ist untypisch, demonstriert aber beispielhaft, welche Fehler auftreten können.

- 1) Anonymous *Encyklika o pokoju między wszystkimi narodami opartym na prawdzie* 1969
- 2) Anonymous *Pacem in terris : encyclical letter of Pope John XXIII on establishing universal peace in truth, justice, charity, and liberty* April 11, 1963 1963
- 3) Anonymous *Pacem in terris. Peace on earth* 1963
- 4) Catholic Church. Pope, 1958-1963 (Joannes XXIII), Barrata, Giuseppe, . *Pacem in terris* 1968
- 5) Catholic Church. Pope, 1958-1963 (Joannes XXIII), Utz, Arthur F. *Die Friedenszyklika Papst Johannes XXIII* 1963
- 6) Catholic Church, Catholic Church. Pope, 1693-. *Pacem in terris. Enzyklika* 1966
- 7) Weiler, Rudolf, Weinbacher, Jakob and Österreichische Kommission Justitia et Pax. *Pacem in terris : [die Friedenszyklika Johannes' XXIII. 1975*

Text 11: Auflistung fehlerhafter Datensätze der ersten Ergebnisseite (erste 10 Treffer) einer Suche nach "Pacem in terris".

Diese Literaturliste wurde von RefWorks auf Basis der VOYAGER-Datenbank der Library of Congress generiert. Die Fehler reichen von 'fehlenden Buchstaben' (1) bis hin zu fehlerhaften Autorenangaben (alle).

Durch die Verwendung eines Thesaurus liefert der Server neben Büchern mit dem gesuchten Titel weitere Werke, bei welchen es sich entweder um Übersetzungen oder kommentierte Übersetzungen handelt. In 30% aller Datensätze wurde der Autor (Papst Johannes XXIII. [mit bürgerlichem Namen Giuseppe Barrata]) gar nicht genannt, in einem Fall wurden die

Herausgeber als Autoren angegeben. Dies kann entweder mit dem Datenformat (USMARC) zusammenhängen (welches syntaktisch nicht zwischen Autoren und Herausgebern unterscheidet) oder mit der Tatsache, dass mangels Fremdsprachenkenntnisse viele ausländische Werke falsch erfasst wurden.

In dem angeführten Beispiel ist vor allem die Datenqualität Grund für die große Abweichung von der Norm. Auf die Qualität der Daten haben vor allem das Datenformat, die exakte Erfassung und die Katalogisierungsregeln einen großen Einfluss. Da die Genauigkeit, mit der ein Bibliothekar bei der Erschließung vorgeht, kein wissenschaftlich relevanter Faktor ist, kann man sie in dieser Betrachtung vernachlässigen. Somit bleiben die Katalogisierungsregeln und die damit zusammenhängende Datenformate neben der Qualität der verarbeitenden Software als Ursachen für mangelhafte Bibliographien. Sie sollen im Folgenden kurz dargestellt werden.

3.2.2.1. Erschließung nach RAK bzw. AACR2 als Voraussetzung für die Datenqualität

Die Katalogisierungsregeln AACR2 und RAK-WB spielen gegenwärtig bei der Recherche von wissenschaftlicher Literatur im deutschsprachigen Raum aus zwei Gründen eine große Rolle: zum einen, weil das international verwendete AACR bzw. AACR2 auf die gesamte englische Literatur angewendet wird, zum anderen, weil auch die schweizerische Bibliotheken (Informationsverbund Deutschschweiz und Schweizerische Landesbibliothek) diese Regeln für die Erfassung ihres Bestandes verwenden¹¹⁶. Die RAK-WB sind das verbindliche Regelwerk für wissenschaftliche Bibliotheken in Deutschland. Greift der Wissenschaftler nicht nur auf die deutschen Veröffentlichungen zu, wird er auf jeden Fall mit den nach AACR2 erfassten Daten konfrontiert. Die Unterschiede werden schon bei oberflächlichen Betrachtung sichtbar. Das AACR2 verfügt über keine Ordnungsregeln, die Namen aus Sprachen mit nichtlateinischen Buchstaben werden transliteriert (ALA/LC) als bei RAK (z. B. DIN-Normen) und können sogar ohne

¹¹⁶ Vgl. Wiesenmüller, Heidrun : *Informationsaufbereitung I : Formale Erfassung* In: KSS. S. 168.

Vorzugsbenennung in englischer Sprache angesetzt werden (während RAK den gebräuchlichen – auch deutschen – oder den offiziellen Namen den Vorzug gibt und eine Verweisung vorschreibt). Aufgrund unterschiedlicher Definitionen kann in AACR2 auch derjenige als Autor genannt werden, der nach RAK als Herausgeber anzusehen wäre¹¹⁷. Geographische Bezeichnungen werden nach deutschen Katalogisierungsregeln in der Schreibweise angesetzt, in der sie gegenwärtig bekannt sind (in der Landessprache des Ortes) oder im Buch als Veröffentlichungsort angegeben werden. In AACR2 werden solche Namen in der Landessprache des Bibliothek erfasst¹¹⁸.

Bei näherer Betrachtung ergeben sich im deutschsprachigen Raum viele Schwierigkeiten. Die Ansetzung von Personennamen in der Landessprache machen die Daten deutscher Kataloge (ohne Verweisungen) international kaum nutzbar. Die Verwendung der deutschen Namen für Orte der ehemaligen deutschen Ostgebiete statt der gegenwärtigen Namen würde zu Missverständnissen führen („Heilsberg“ statt „Lidzbark Warmiński“, „Allenstein“ statt „Olsztyn“, „Breslau“ statt „Wrocław“).

Auch die RAK weisen gewisse Schwächen auf, die bei einer softwaregestützten Verarbeitung von bibliographischen Daten zu Problemen führen können. Die Schreibweise der RAK wie z. B. „MacArthur“ muss unter Umständen vom Benutzer korrigiert werden. Die Betrachtung eines Werkes mit mehr als drei Verfassern als ein anonymes Werk oder die bloß semantische Unterscheidung zwischen Autoren und „weiteren beteiligten Personen“ ist ebenfalls problematisch. Die Angabe der Funktion einer Person als Abkürzung der deutschen Bezeichnung in runden Klammern – sowie das Fehlen spezieller Felder für Personen mit anderen Funktionen in den Austauschformaten MAB und MARC führt bei Verwendung von DIN 1505-2 zu Fehlern. Die Abkürzung, die die Funktion einer Person angibt (die

117 Vgl. Payer, Margarete : AACR2 - RAK : grundsätzliche Unterschiede. URL: <http://www.payer.de/einzel/aacr2rak.htm> (Stand: 09.07.2006)

118 Vgl. Hilberer, Thomas : Aufwand vs. Nutzen : Wie sollen deutsche wissenschaftliche Bibliotheken künftig katalogisieren? In: Bibliotheksdienst 2003, Nr. 6, S. 756

genauso in AACR2 verwendet wird) wird bisher nur für die englische Sprache richtig verarbeitet (als Hinweis auf die Funktion). Die eben erwähnten Datenaustauschformate, die auf die Nutzung von RAK bzw. AACR abgestimmt sind, übermitteln zu wenige und zu schwach untergliederte Daten und tragen so zu einer suboptimalen Nutzung dieser bei.

Zusammenfassend lässt sich sagen, dass weder AACR2 noch RAK Daten ausreichender Qualität für eine zufriedenstellende Verarbeitung zu Quellenangaben liefern können. DIN 1505 als eine der komplexesten Zitiernormen ist ein geeigneter Indikator dafür. Für eine optimale automatische Verarbeitung bibliographischer Daten zu DIN 1505 konformen Bibliographien ist eine atomare¹¹⁹ und ausdrucksstarke Datenstruktur die Voraussetzung. Dieser Anforderung genügen weder diese Katalogisierungsregeln noch die veralteten Datenformate.

3.2.2.2. Schwachstellen in der Verarbeitung von Quellenangaben in unterschiedlichen Zitiernormen

Die wichtigsten Schwachstellen bei der automatischen Verarbeitung von bibliographischen Daten zu Literaturverzeichnissen ergeben sich aus der Datenstruktur (fehlende Eindeutigkeit und die Mehrdimensionalität der Datenfelder in den Austauschformaten), der Datenqualität (Eingabe- oder Erschließungsfehler, Zeichensatz-Probleme) und der Ausrichtung der Regelwerke sowie der elektronischen Kataloge wie z. B. der Z39.50-Server auf die Mensch-Maschine-Kommunikation. Die zuletzt genannte Ursache spiegelt sich zum Teil in der Datenstruktur der Austauschformate. Die elektronischen Kataloge entstanden auf der Grundlage der Regelwerke, die für Zettelkataloge entwickelt wurden und ihre Stärke in diesem Bereich bewiesen haben. Die Karteikarten waren nur für den Menschen lesbar und – eine gewisse Übung vorausgesetzt – ohne Probleme zu verstehen. Die Ergebnisse einer Suche via Z39.50 können ähnlich einer Karteikarte gelesen und vom Menschen interpretiert werden¹²⁰. Diese Tätigkeit wurde in den

119 Das Wort "atomar" wird in dieser Arbeit im Sinne der Prädikatenlogik verwendet: Eine Aussage ist dann atomar, wenn sie nicht weiter zerlegt werden kann.

120 Diese Tatsache lässt sich z. B. an der Angabe des Geburts- und Sterbejahres einer

letzten Jahren zur Domäne der persönlichen Literaturverwaltungssoftware. Anhand einiger Regeln kann ein Programm den Inhalt eines Feldes oder eines Teilfeldes interpretieren und in eine interne Darstellung überführen. In bestimmten Fällen kann es jedoch zu Fehlinterpretation kommen, nämlich dort, wo die Regeln nicht angewendet werden können: bei Anmerkungen, Unterschieden in der Schreibweise o. Ä. Der Mensch kann den Sinn der Aussage erkennen und dank seines bereits erworbenen Wissens Kürzel auflösen. Damit ein Computerprogramm eine Aussage „verstehen“ d. h. richtig zuordnen kann, muss diese möglichst eindeutig sein. „editor = {IEEE Computer Society}“ ist keine solche. Die einzige eindeutige Aussage dieses Ausdrucks ist: "IEEE Computer Society" ist der Herausgeber. Ob es sich hierbei um eine Person oder eine Organisation handelt, kann nur der Mensch mit seiner Intelligenz (sowie seinen sprachlichen Fähigkeiten) erkennen. (Ein Computer benötigt linguistisch-semantische Regeln zur Erkennung solcher Unterschiede.) Als Folge dieser Zweideutigkeit könnte die Umformatierung des Namens des Herausgebers nach den Regeln des DIN 1505-2 zu folgendem Ergebnis führen: „Computer S., IEEE“. Noch schwerwiegender als bei diesem Beispiel eines Datenfeldes in BibTeX-Format kann eine falsche Deutung eines MAB- oder MARC-Feldes sein. Darin werden Personen und Organisationen zwar unterschieden, doch es gibt kein Datenfeld für den Herausgeber (bzw. wird es beim Export nicht verwendet). Ein solcher wird mit dem Kürzel „Hrsg.“ bzw. „ed.“ (in Klammern) ausgewiesen. Um dieses Feld richtig zu interpretieren, wird ein linguistisches Verfahren angewandt, das solche Kürzel sucht. Dabei müssen dem Computerprogramm alle mögliche Kürzel, deren Bedeutungen sowie Übersetzungen bekannt sein.

Person erkennen. Solche Angaben sind für die Zuordnung von Werken zu einer Person (unter mehreren Personen mit dem selben Namen) hauptsächlich für den Menschen relevant. Die Eindeutigkeit von Entitäten wird im Computerbereich mit Hilfe von Identifikationsnummern (ID) sichergestellt. Anders ausgedrückt: Nur der Mensch kann anhand des Sterbejahres den Johann Strauss Sohn von seinem Vater unterscheiden – für eine Maschine kann diese Angabe besten Falls nur ein zusätzliches numerisches Merkmal darstellen und nur in diesem Zusammenhang relevant sein, ohne die Person zu identifizieren.

Ähnlich große Probleme können die Personennamen darstellen. Während es im Normalfall ein bis zwei Vornamen und der Nachname verarbeitet werden müssen, kommt es bei komplexeren Namen sowie sonstigen Personenbezeichnungen (wie den Namen der Päpste) zu Fehlern. Als ein pragmatischer Ausweg erwies sich die Auslieferung von formatierten Personennamen (Regelwerk- und sprachabhängig) im Datenfeld der Person (MAB, MARC) bzw. ein Merkmal, dass der Name von der Formatierung ausgeschlossen werden soll (vgl. BibTeX). Richtig und für die maschinelle Verarbeitung besser geeignet wäre, alle Bestandteile eines Namens einem Typ zuzuweisen, so dass unterschiedliche Regelwerke auf nur bestimmte Typen von Namensbestandteilen angewendet werden¹²¹.

Als weiterer signifikanter Faktor für die automatische Generierung von Literaturverzeichnissen erweist sich die Qualität der Datenbasis. Eingabefehler bei der Erfassung der Daten, falscher Zeichensatz bei der Übermittlung oder gar Fehler bei der Erschließung führt ebenfalls zu Verwechslungen und ungültigen Referenzen.

Die größte Gefahr für die Benutzer dieser Dienste besteht in einem Vertrauen in die Richtigkeit der generierten Bibliographie. Einige Fehler können schon durch die Nachbearbeitung der übernommenen Daten beseitigt werden. Andere können erst in der fertigen Bibliographie entdeckt werden. Die Korrektur erweist sich in diesem Fall meist schwieriger, da der Benutzer die Funktionsweise des Programms und die Zitiernormen genauer kennen sollte.

An dieser Stelle sollen an einigen Beispielen Schwächen der Verarbeitung von bibliographischen Daten aufgezeigt und erklärt werden.

- „*li.*, *Johannes P.*“ als Autor im Literaturverzeichnis: hier wurde der Name „*Johannes Paul II.*“ umformatiert, obwohl er nicht geändert werden dürfte.

¹²¹ Die Vorteile dieses Ansatzes sowie mögliche Typisierung der Namensbestandteile werden im Teil V dieser Arbeit behandelt.

- „Golił„ski, Zbigniew“ der Name eines polnischen Schriftstellers ist durch die falsche Zeichenkodierung nicht wieder zu erkennen (bzw. fehlende Umformatierung des Zeichensatzes).
- „Nikolay Rimsky-Korsakov“ und „Nikolaj A. Rimskij-Korsakov“ - englische und deutsche Schreibweise desselben Namens (Problem der Transliteration und Transkription)
- „Anonymous: *Pacem in terris. Peace on earth 1963*“
Erschließungsfehler bei der Enzyklika „Pacem in terris“ von Papst Johannes XXIII.

3.2.3. Umsetzung

Die Qualität der Literaturangaben hängt mit der Erfassung und den Fähigkeiten der Literaturverwaltungssoftware zusammen. Hier können die kommerziellen Anbieter – vor allem Firmen, die ihre Produkte seit mehreren Jahren entwickeln und verbessern – ihren technischen Vorsprung ausbauen. Im Gegensatz zu Open-Source-Programmen, die unter Anderem im Rahmen von Hobbys in der Freizeit entstehen, können sie Kooperationen mit großen Bibliotheken eingehen und mehr Zeit in die Entwicklung investieren. Diese Tatsache bedeutet jedoch nicht, dass dies in jedem Bereich gleichermaßen geschieht. Bei keinem der Anbieter wird der volle Funktionsumfang von Z39.50 genutzt. Hier scheinen die Bibliotheken selbst am meisten zu investieren. Der ZACK-Gateway der Universitätsbibliothek Braunschweig (programmiert von Wolfram Schneider im Rahmen seiner Diplomarbeit) bietet neben der üblichen Suche Erklärungen für die Datenformate (sogar einzelne Teilfelder), Dublettenkontrolle sowie einen Konverter für MAB2 Daten. Die Deutsche Nationalbibliothek und Library of Congress bieten ebenfalls eine Web-basierte Suche im eigenen Katalogsbestand¹²².

122 Die Library of Congress bietet neben dem Z39.50-Gateway ein Web-Oberfläche für die Recherche unter der URL <http://catalog.loc.gov/>. Die Deutsche Nationalbibliothek ermöglicht eine Online-Recherche über den gesamten Bestand unter der URL <https://portal.dnb.de>

Nicht berücksichtigt werden in diesem Kapitel Softwaresysteme, die nicht das Z39.50-Protokoll benutzen wie Zeitschriften- oder Artikeldatenbanken von Universitäten und wissenschaftlichen Einrichtungen wie z. B. SciDok der Saarländischen Universitäts- und Landesbibliothek.

3.2.3.1. Web-Gateway

Bei den sog. Web-Gateways¹²³ handelt sich um einen Z39.50-Client, der einen oder mehrere Z39.50-Server abfragt und die Ergebnisse als HTML an den Webbrowser des Benutzers ausliefert. Die Funktionalität wird von einem Programm auf dem Server (Gateway) über das HTTP bereitgestellt. Vereinfacht könnte man einen solchen Server als Z39.50-Suchmaschine bezeichnen.

3.2.3.1.1. Recherche in der Deutsche Nationalbibliographie online

Die Deutsche Nationalbibliothek stellt unter <https://portal.dnb.de> eine Web-Oberfläche für eine Recherche in eigener Katalogdatenbank ILTIS¹²⁴ (sowie einigen Datenbanken angeschlossener Institutionen und archivierten elektronischen Ressourcen) zur Verfügung. Über ein Suchformular der erweiterten Suche kann man die von der DNB erfassten Werke nach Schlagwort, Nummern (ISBN, ISSN, URN), Person (auch solche, die kein Autor ist), Personenschlagwort, Körperschaft (mit vollständigem oder nur einem Teil des Namens), Titel (sowie dessen Teilen), Verlag oder Erscheinungsort, Erscheinungsjahr, codierten Angaben, DDC, Signatur oder Sachgruppe. Ein Fragezeichen hinter dem gesuchten Begriff maskiert eine rechtsseitige Trunkierung. Beim Erscheinungsjahr ist eine Angabe in Form "von-bis" erlaubt.

123 Das Wort „Gateway“ ist verbreiteter als deren deutsche Entsprechung: „Protokollumsetzer“ und bedeutet eine Internetseite, die nur wenige eigene Informationen anbietet, aber zu solchen hinführt oder solche auf Anfrage zusammenstellt.

124 ILTIS ist das Integrierte Literatur-, Tonträger- und Musikalien-Informationssystem der Deutschen Nationalbibliothek. In seinem Zentralkatalog findet man neben den Normdaten ca. 5. Millionen Titel der Deutschen Nationalbibliographie seit 1945.

Das Ergebnis einer Suche wird als eine Liste von Titel-Ort-Verlag-Jahr-formatierten Literaturangaben dargestellt. Eine Export-Funktion in eines der bibliographischen Dateiformate ist nicht möglich. Eine solche Liste von Werken ist für eine schnelle Recherche ohne Literaturverwaltungssoftware durchaus geeignet. Dieser Internet-Dienst der DNB ist vor allem als Rechercheplattform für künftige Besucher dieser Einrichtung zu verstehen.

Ein anderes Suchformular („Globale Suche“) ermöglichte in der vorherigen Version eine Suche über mehrere Datenbanken (ILTIS, VOYAGER der Library of Congress, LIBRIS der schwedischen Nationalbibliothek, COPAC – dem gemeinsamen Katalog der Universitäts- und Nationalbibliotheken Großbritanniens und Irlands sowie den Datenbanken der GBV und BVB). Hierbei entfiel die Möglichkeit der Suche nach Personen, Erscheinungsjahr oder einigen Identifikatoren. Dies hing wohl damit zusammen, dass es sich hierbei um Suche in heterogenen Datenbanken handelte, die nach unterschiedlichen Erfassungsprinzipien erstellt wurden (RAK, AACR).

Interessanterweise ist die Suche in anderen Datenbanken deutscher Verbundkataloge trotz der gemeinsamen Erfassungsnormen (RAK und RAK-WB) sowie der für alle deutschsprachigen Bibliotheken einheitlichen Gemeinsamen Körperschaftsdatei (GKD¹²⁵), Personennamendatei (PND) sowie der Schlagwortdatei (SWD) nur mit einem minimalen Satz an Suchfeldern möglich (Autor, Titel, Sachwort, ISBN, Verlag und Körperschaft). In einem homogenen Datenumfeld dürfte man als Benutzer mehr erwarten.

3.2.3.1.2. ZACK

Der Internet-Dienst ZACK¹²⁶ entstand im Rahmen der Diplomarbeit von Wolfram Schneider. Als Zielsetzung gibt der Autor folgendes an:

125 Die GKD ging in den 70er-Jahren aus den Eintragungen der Körperschaftsnamen der Zeitschriftendatenbank (ZDB) hervor. Sie wird von der Bayrischen Staatsbibliothek, der Staatsbibliothek zu Berlin, der Österreichischen Nationalbibliothek und der Deutschen Nationalbibliothek gepflegt und aktualisiert. Sie enthält über 900 000 Datensätze. Sie wird benötigt für eine einheitliche Titelaufnahme nach RAK-WB. (vgl. <http://www.zeitschriftendatenbank.de/gkd/index.html>)

126 Dieser Dienst als HTTP-Z39.50-Gateway ist unter <http://opus.tu-bs.de/zack/> abrufbar.

„Ziel dieser Diplomarbeit ist die Entwicklung eines Bibliotheks-Informationssystems, das Bibliothekare bei der Recherche und Erfassung von Dokumenten unterstützt. Der Name des Systems ist ZACK.

Der Benutzer von ZACK kann in einer oder mehreren bibliographischen Datenbanken nach einem Dokument suchen und das geeignete Dokument in die lokale Datenbank übernehmen. Mit der Übernahme der Datensätze aus einer fremden Datenbank wird die Erfassung neuer Dokumente wesentlich erleichtert, da die Eigenkatalogisierung auf ein Minimum beschränkt werden kann. Es wird doppelte Arbeit vermieden, und die Datensätze haben eine gleichbleibend hohe Qualität.“¹²⁷

Um das zu erreichen, setzt das Programm neben einem Z39.50-Client, der gleichzeitig mehrere Z39.50-Server abfragt¹²⁸, einige Verfahren zur Vermeidung von Dubletten (sog. "Deduplizierung") ein. Dies geschieht durch Abgleich der Datenfelder (Titel, Autor, Verlag), Datensatztypen (Buch, Landkarte etc) und Normierung von den darin enthaltenen Texten¹²⁹.

Die Ergebnisse werden in Form einer ISBD-formatierten Liste unter Angabe des Datenlieferanten (der dieses Werk in seinem Bestand hat) dargestellt. Die Datenquelle wird als Hyperlink dargestellt, hinter welchem der Datensatz in seiner Auslieferungsform eingesehen werden kann. Dies kann MAB2 oder MARC sein. Die Datenfelder sind mit einer Online-Hilfe verlinkt, die die kryptischen Zahlen- und Zeichenfolgen erklärt. Zusätzliche Funktionalität, die für diese Arbeit kaum relevant aber sicherlich von großem Nutzen für die Besucher sein kann, ist die Darstellung des Umschlags des gefundenen Buches sowie ein Hyperlink zu einem weiteren Suchformular, über den frei

127 Schneider, Wolfram: *Ein verteiltes Bibliotheks-Informationssystem auf Basis des Z39.50 Protokolls*. Berlin, Technische Universität Berlin, Fachbereich Informatik, Dipl.-Arb., 1999.

128 Diese Art der Suche auf mehreren Servern gleichzeitig bezeichnet man als „verteilte Suche“.

129 ebd. S. 63 fff

verkäufliche Exemplare des Werkes (neu oder in einem Online-Antiquariat) gefunden werden können.

ZACK ermöglicht zwar weder die Übernahme der Datensätze in die eigene Sammlung noch ein Export der gefundenen Werke in einen der bibliographischen Dateiformate. Seine Stärken sind: eine genaue Suche (mit oder ohne Trunkierung) in mehreren Datenquellen, Konvertierung zwischen MAB- und MARC-Formaten sowie Dublettenkontrolle („Match & Merge“). Am Ende jeder Ergebnisliste werden die gefundenen Sachgebiete, Erscheinungsjahre und Personen in einer sog. Schlagwortwolke¹³⁰ dargestellt. Diese kann als Ausgangspunkt für die Suche nach weiterer Literatur dienen.

3.2.3.1.3. RefWorks

Das serverseitige Programm zur Recherche und persönlicher Verwaltung von Literatur ist proprietäre Software, die im Gegensatz zu ZACK oder den OPAC¹³¹ der wissenschaftlichen Einrichtungen, nicht von diesen (bzw. in deren Auftrag) entwickelt und betrieben wird. Die Nutzung dieses Web-Dienstes ist nicht kostenfrei: die Kosten trägt der Nutzer oder die Organisation, die eine Mehrbenutzer-Lizenz erworben hat.

Die hier beschriebenen Eigenschaften dieser Software beziehen sich auf die im Jahr 2008 von der SULB angebotene Version (1). Da die Zyklen der Herausgabe neuer Versionen für gewöhnlich relativ kurz sind, kann man davon ausgehen, dass zum Zeitpunkt des Verfassens dieser Arbeit eine korrigierte Version vom Hersteller angeboten wird.

130 en. „tag cloud“ - Eine Art der zweidimensionalen Visualisierung der Relevanz von einzelnen Wörtern in einem bestimmten Umfeld (z.B.: verwandte Begriffe wie das gesuchte Wort etc). Andere Bezeichnungen: Begriffswolke, Tag-Wolke.

131 OPAC en. Open Access Public Catalog bezeichnet ein öffentlich zugängliches Katalog, welches die Suche nach den im Bestand befindlichen Werken erlaubt. Es gilt als Oberbegriff für Z39.50-Gateways und sonstige Artikel- und Zeitschriftendatenbanken, die frei (d.h. jedermann kostenlos und uneingeschränkt) zugänglich sind.

Die Recherche mit RefWorks gestaltet sich ähnlich wie in ZACK: Ein Suchformular erwartet die Eingabe von Suchbegriffen, die in den typischen Datenfeldern wie Autor, Titel, Verlag und Erscheinungsjahr. Zusätzlich muss ein Datenlieferant¹³² ausgewählt werden. Es ist jedoch keine Mehrfachauswahl – wie beim ZACK – möglich. Unter den nach Titel angeordneten Ergebnissen kann man mehrere Werke auswählen und in eine der eigenen Sammlungen importieren. Es besteht ebenfalls die Möglichkeit, eine Textdatei in einem der bekannten bibliographischen Formate zu importieren. Die eigene Literatursammlung kann man als eine solche Datei (mit gelegentlichen Formatierungsfehlern) oder in Form einer Bibliographie exportieren. Zu diesem Zweck stehen mehrere hundert Ausgabe-Formate („Filter“ genannt) zur Verfügung. Der Versuch, eine Sammlung mit Werken aus unterschiedlichen Quellen (und somit verschiedenen Erfassungsweisen) als eine DIN-1505-2-Bibliographie zu exportieren, hatte eine mangelhafte Aneinanderreihung von ISBD-formatierten Quellenangaben zur Folge. Auch das BibTeX-Format war wegen fehlender Klammern mit LaTeX nicht zu verarbeiten. Der dritte und für einen Austausch geeignetste Weg ist das Exportieren der Datensätze via Web. Dies kann mit RefShare erreicht werden. Diese Daten können jedoch vom Benutzer nicht direkt in die eigene Sammlung übernommen werden.

Die Stärken von RefWorks sind die Unterstützung des Benutzers bei der Dublettenkontrolle, deutsche Sprachversion der Oberfläche, die Unterstützung des Unicode sowie die vielen Stile bzw. Filter für den Export als Norm gerechte Quellenangaben (oder als bibliographische Datenformate). Anders als die Literaturverwaltungssoftware für den PC ist RefWorks betriebssystemunabhängig und von überall mit einem Web-Browser und Internetanbindung erreichbar. Der Funktionsumfang ist größer als es bei vergleichbaren Software WriteNote von Thomson Reuters (zu welchem ISI Research - der Betreiber von Web of science)¹³³ gehört. Die

132 In Refworks kann man nicht nur Z39.50-Server, sondern auch die MedLine-Datenbank als Quelle auswählen.

133 Hobohm, Hans-Christoph: *Persönliche Literaturverwaltung im Umbruch*. In: *Information*.

äußerst dynamische Entwicklung in diesem Bereich lässt auf eine zuverlässige und qualitativ hochwertige Software hoffen.

3.2.3.2. Reference Manager als Beispiel für kommerzielle Umsetzung eines in ein persönliches Literaturverwaltungssystem integrierten Z39.50-Clients

Reference Manager bietet im Vergleich zu anderen Literaturverwaltungssystemen für den PC¹³⁴ den größten Funktionsumfang. Sie beherrscht die Suche via Z39.50 sowie auf den ISI- und PubMed-Servern. Trotz gelegentlicher Fehler, die mit der Zeichenkodierung der empfangenen Datensätze zusammenhängt, zeigt sie nur wenige Schwächen. Das Programm unterstützt den Benutzer bei der Suche, indem es (im Falle einer PubMed-Suche) eine Liste der richtig formatierten Namen zur Auswahl anbietet und somit das Risiko eines Eingabefehlers minimiert. Man kann mehrere Sammlungen anlegen, die als binäre Dateien (Datenbanken) abgelegt werden. Diese kann man per Hand erfassen oder aus Datensätzen der Antwort eines Servers zusammenstellen. Zur Auswahl stehen: PubMed, ISI (Eingabe der Login-Daten notwendig) sowie die wichtigsten Z39.50-Gateways. Eigene Z39.50 kann man anlegen, indem man in einem Formular die sog. Profil-Daten angibt oder indem man eine Datei, die diese Daten enthält, importiert.

Die gesammelten Quellen kann man in wenigen Schritten aus dem Textverarbeitungsprogramm Microsoft Word heraus zitieren. Dabei übernimmt Reference Manager die Verwaltung der sog. Cite-Key¹³⁵ und generiert an angegebener Stelle eine vollständige Bibliographie. Diese Bibliographie kann – genauso wie jedes (sog. „inline“-) Zitat – nach einer der

Wissenschaft & Praxis 56 (2005), Nr. 7, S. 386

134 Unter „PC“ scheinen die meisten Software-Hersteller nur Windows (selten auch MacOS) zu verstehen. Für Linux als alternatives Betriebssystem wird bisher keine kommerzielle Software im Bereich der Literaturverwaltung angeboten. Die vielen Open-Source-Programme können aufgrund der ehrenamtlichen Arbeit nicht mit kommerziellen Programmen im Funktionsumfang und Ausstattung verglichen werden.

135 Für eine genauere Funktionsweise der Cite-Keys vgl. S. 70. Dieses Verfahren wird auch als „cite-while-you-write“ bezeichnet. (vgl. Hobohm S. 385)

unzähligen Normen formatiert werden. Für die richtige Umsetzung von bibliographischen Daten in eine bestimmte Zitierweise sind die sog. „styles“ - Zitierstile zuständig. Nach der Installation des Programms stehen dem Benutzer einige hundert Stile zur Verfügung. Darunter befindet sich jedoch keines, das die ISO 690 oder DIN 1505 vollständig umsetzen würde. Weitere Stile können aus dem Internet heruntergeladen werden. Fehlt es an geeigneter Vorlage, kann der Benutzer selbst eine solche entwerfen. Dem Autor ist es nicht gelungen, die Norm DIN 1505-2 z.B. für Hochschulschriften, umzusetzen. Es scheint eine Frage der Ausrichtung dieser Software auf die englischsprachige Publikationspraxis und -Normen zu sein, weniger die der zu verarbeitenden Daten selbst.

Über zusätzliche Module kann man das Programm um zusätzliche Funktionen wie die Visualisierung (Textanalyse und Gruppierung) erweitern¹³⁶. Sinnvoller als solche Zusatzprogramme wäre jedoch eine bessere Übersetzung der genutzten Begriffe und Entwicklung benutzerfreundlicher Eingabemasken - wie Hobohm bemerkt¹³⁷. Ebenso wichtig wäre die Nutzung von Personennamen- oder Schlagwortnormdateien in den Suchformularen (ähnlich wie bei der Suche in MedLine) sowie deren lokalen Index, die dem Benutzer schnell eine Übersicht über die zitierten Autoren geben und für einheitliche Schreibweise deren Namen sorgen würde.

Die Anbindung an Microsoft Word und das Fehlen eines Versionsverwaltungssystems, das den Abgleich der Daten zwischen mehreren Installationen dieser Software erleichtern würde, schränken den Benutzer in der Wahl seiner Textbearbeitungssoftware bzw. seines Arbeitsplatzes deutlich ein.

Der Austausch relevanter Literatur unter den Benutzern dieser Software kann nur indirekt erfolgen. Dazu exportiert man ausgewählte (oder alle) Werke in

136 Hobohm, Hans-Christoph : *Persönliche Literaturverwaltung im Umbruch*. In: *Information. Wissenschaft & Praxis* 56 (2005), Nr. 7, S. 386

137 ebd.

eine Datei (z.B. RIS) und sendet diese per Email oder stellt sie ins Web. Auch ein Export in eine HTML-, RTF- oder Word-Datei als Bibliographie ist möglich. Diese Form kann nicht wieder importiert werden. Eine Gruppen-Funktionalität ist beim EndNote realisiert, wie der Hersteller (Thomson Reuters) auf der Internetseite des Produktes betont. (Die Ausgestaltung und der Umfang dieser wird dort jedoch nicht näher beschrieben.)

Trotz einiger Schwächen ist Reference Manager sowie EndNote die persönliche Literaturverwaltungssoftware mit der besten Integration in die Arbeitsabläufe des wissenschaftlichen Autors. Sie ist pragmatisch aufgebaut und genügt den Anforderungen eines deutschen Bibliothekars oder Informationswissenschaftlers kaum. Dennoch sind EndNote und Reference Manager das Werkzeug der Wissenschaftler, wie das Streben nach einer Campus-Lizenz dieser Software für die Universität des Saarlandes beweist¹³⁸.

3.2.3.3. Tellico als Beispiel für Umsetzung eines Z39.50-Clients im Rahmen einer Software zur Verwaltung von Sammlungen

Das Open-Source-Programm Tellico entstand mit dem Ziel, den Benutzern des freien Betriebssystems Linux ein Werkzeug an die Hand zu geben, mit dem er seine CD-, DVD-, Buch- oder ähnliche Sammlungen katalogisieren kann. Heute reichen die Datenbank-Vorlagen für einen neuen Katalog von „B“ wie Briefmarken bis hin zu „W“ wie Wein. Der Benutzer kann eine davon auswählen und somit einen Satz an Datenfeldern übernehmen (und bei Bedarf erweitern) oder selber eine definieren. Dies kann sehr nützlich sein, wenn man versucht Artikel (als unselbständig erschienene Werke) zu erfassen, die oft in kein Schema passen. Verwendet man die Vorlagen „Büchersammlung“ oder „Literaturverzeichnis“, so ist es möglich, bibliographische Daten aus RIS-, BibTeX-, BibTeXML oder MODS-Dateien zu importieren. Das eigene XML-basierte Format sowie CSV (Comma

138 Vgl. Vergünstigte Lizenzen dieser Programme für Mitglieder der UdS. URL: <http://www.its.uni-saarland.de/informationen/arbeitsplatz/verfuegbare-software-lizenzen/thomson-reuters-endnote/> (Stand: 10.06.2007)

Separated Values) können jederzeit importiert werden. Wählt der Benutzer die Option „Sammlungen verschmelzen“ statt „hinzufügen“, versucht das Programm Dubletten zu erkennen und importiert nur neue Datensätze. Die Dublettenerkennung ist nur schwach ausgeprägt, da kleinste Unterschiede in der Schreibweise der Namen oder zusätzliche Felder desselben Werkes als neues Werk interpretiert werden. Neben dem Import bibliographischer Daten, beherrscht Tellico auch einige Exportfilter für eben diese Formate, darunter: BibTeX, RIS und ONIX. Die Werke können dem Autor – oder eher seinem Namen – zugeordnet dargestellt werden. Die Identität der Person wird wie in allen Programmen vom Namen her abgeleitet, so dass „Benedict XVI.“ nicht mit „Joseph Ratzinger“ identisch sein kann.

Die Besonderheit eines Literaturverzeichnisses bzw. einer Buchsammlung besteht in Tellico darin, dass man zusätzliche Werkzeuge für die Erfassung nutzen kann. Dazu gehört ein Z39.50- und SRU-Client. Das Dialogfenster wurde im Menü der „Suche im Internet“ zugeordnet. Es gibt nur ein Eingabefeld, welches einem der fünf Kriterien zugeordnet werden kann: Titel, Person, ISBN, „Schlüsselwort“ (eigentlich: Schlagwort), UPC (nur bei Amazon-Suche) und LCCN (nur bei Z39.50-Suche)¹³⁹. Weil Tellico mehrere Profile für Z39.50-Gateways sowie die Abfrage der Amazon-Server (auch mit mehreren Länderprofilen) und des ISBNdb.com-Servers beherrscht, kann der Benutzer zwischen den Quellen auswählen und dieselbe Abfrage nacheinander an mehrere von ihnen abschicken, um die Ergebnisse in einer Liste zusammenzufassen. Die eingegebenen Begriffe werden ohne Trunkierung gesucht. Die Suche nach UPC bzw. ISBN erlaubt sogar eine Mehrfach-Eingabe. Die Ergebnisse sind mit anderen in dieser Arbeit vorgestellten Programmen vergleichbar. Die Verarbeitung von Sonderzeichen funktioniert in allen Programmen unzuverlässig¹⁴⁰. Im

139 UPC steht hier für „Universal Product Code“, den Vorgänger von „European Article Number“ (EAN) – beide sind allgemein als „Barcode“ bekannt. ISSN oder Körperschaft fehlen in der Auswahlliste. Wahrscheinlich ist es der kleinste gemeinsame Nenner für die sehr heterogenen Datenquellen.

140 Es ist nicht auszuschließen, dass diese Probleme nicht von den Programmen her kommen, sondern an der Art und Weise der Erfassung liegen: Während der Katalog

Vergleich zu Reference Manager sind beim Tellico die einzelnen Datenfelder in der Darstellung nach Zugehörigkeit zu einer Eigenschaft sortiert: Verlag, Auflage, Sprache, Seitenzahl, Erscheinungsjahr und Erscheinungsort werden im Kasten "Ausgabe" gruppiert, Titel, Untertitel und Autor hingegen unter "Allgemein". Es ist sozusagen die Trennung zwischen dem Werk als solchem und seiner physikalischen Erscheinung¹⁴¹.

Ähnlich wie die meisten Programme versucht auch Tellico eine direkte Zitation zu ermöglichen. Zu den unterstützten Programmen zählen das Textverarbeitungsprogramm OpenOffice Writer sowie der GUI-LaTeX-Editor Lyx. In beiden Fällen gelingt dies nur unzureichend. Weder Vorlagen für unterschiedliche Zitier-Normen noch eine Möglichkeit der Erstellung einer eigenen Zitiernorm sind möglich. Dies kann mit Hilfe der OpenOffice-eigenen Mittel geschehen, da Writer (bzw. das ODF) bibliographische Daten als solche rudimentär auszeichnet und im Nachhinein neu formatieren kann. (Dies ist in Microsofts Word grundsätzlich nicht der Fall. Die Quellenangaben werden nicht ausgezeichnet und wie ein formatierter Text behandelt.)

Die Bedienung von Tellico ist benutzerfreundlich. Das Fehlen von Zitierstil-Vorlagen ist auf Grund der Bindung an OpenOffice und LaTeX stellt eine Schwäche dar. Eine Funktion zum Export und Import von Z39.50-Server-Einstellungen wie bei EndNote oder Reference Manager fehlt. Es existieren jedoch einige vordefinierte Verbindungen zu den Nationalbibliotheken.

„SWB lite“ fast alle nicht-deutschen Buchstaben falsch ausliefert, sind die Ergebnisse des „voyager“ der Library of Congress nur in wenigen Fällen fehlerhaft. Es ist denkbar, dass solche falschen Eingaben mit der Codierung von Sonderzeichen auf der Betriebssystem-Ebene des Erfassenden zusammenhängt: ISO-8859-X, Windows-CP-125X oder IBM-85X beherrschen in ihrer 8-Bit-langen Darstellung "nur Westeuropäisch" oder "nur Mitteleuropäisch" - benutzen für die Repräsentation von nichtlateinischen Zeichen jedoch unterschiedliche Bits. Die Umsetzung in den intern verwendeten Zeichensatz kann diese Probleme zusätzlich verstärken.

141 Ein Buch oder eine Zeitschrift kann mehrere Ausgaben haben: mehrere Auflagen oder mehrere Medien (z. B. Print, online-Ausgabe und CDRom). Das Werk als solches existiert mit seinen Kern-Eigenschaften wie Autor, Titel oder Inhalt unabhängig davon. Diese Differenzierung wird durch FRBR genauer ausformuliert.

Anders als viele Programme für Windows bietet er jetzt schon das neue Protokoll ZING (SRU) an, die Suche und Import von bibliographischen Daten von Social-Bookmarking-Diensten wie Bibsonomy sowie Google-Scholar. Die Darstellung der Bücher und der Dialog zum Erfassen bzw. Korrigieren von Angaben zu den Werken zeigt, dass die Programmierer mit der Usability vertraut sind oder aus der Rückmeldung der Benutzer gelernt haben. Tellico bietet die Suche in PubMed an und kann mit unselbständig erschienenen Werke – jedoch ohne eine Verbindung zu den selbständig erschienenen Werken – umgehen. Viele „nette“ Funktionen, wie „mit Daten vom Server aktualisieren“ (für unvollständige Datensätze), eine einfache Ausleihe, Bewertung auf einer Beliebtheit-Skala, Marker („gelesen“, „signiert“ etc) und Verknüpfung mit dem Abbild des Umschlags sind seine Stärken. Die Schwachstellen im Bereich der Zitation sind deutlich. Für Software zur Verwaltung von Sammlungen jeglicher Art ist der Funktionsumfang im bibliographischen Bereich dennoch beträchtlich.

3.2.4. MedLine als Beispiel für eine Artikeldatenbank

MedLine ist in vielerlei Hinsicht anders als die Z39.50-Gateways: Der Informationsaustausch wird über das weit verbreitete HTTP-Protokoll realisiert; es ist kein Standard (weder ISO noch NISO) und – wohl die wichtigste Eigenschaft von PubMed – es enthält fast ausschließlich Artikel. Das Spektrum an Informationen zu einzelnen Artikeln reicht vom Autor und Titel bis hin zur genauen Angaben zu Zeitschrift, in der sie veröffentlicht wurden, dem darin behandeltem Sachgebiet (Namen von Organen, Krankheiten oder Therapiearten) sowie einer kurzen Zusammenfassung. Die heutige Forschung auf diesem Gebiet kann nach Meinung des Autors eine solche Informationsquelle praktisch nicht entbehren, da die Ergebnisse dieser Tätigkeit fast ausschließlich in Form von Artikeln erscheinen. Diese Art der wissenschaftlichen Kommunikation gewinnt auch in anderen – meist „jungen“ – Fachgebieten wie Informatik oder Psychologie an Bedeutung. Damit kann man die MedLine-Datenbank als den Prototypen der zukünftigen wissenschaftlichen Informations- und Referenz-Quellen ansehen.

3.2.4.1. Gateway von MedLine

Wie bereits erwähnt, ist die Kommunikation zwischen dem Programm des Benutzers und dem Datenbank-Server kein eigenständiges Protokoll. Es basiert auf dem Hypertext Transfer Protocol (HTTP) des WWW. Der Unterschied liegt darin, dass der Benutzer nicht auf die Hyperlinks im Hypertext klickt, sondern die Angaben zu den Werken in einem eigenen bibliographischen Format abrufen. Übertragen werden reine Textdateien (RIS-ähnlich) bzw. XML.

Der Gateway besteht aus einer Reihe von Programmen, die eine der Datenbanken (wie PubMed für Medizin oder HomoloGen für Gen-Codes) abrufen. Die Eingaben werden miteinander kombiniert, so dass eine Suche und Filterung gleichzeitig möglich ist. Die Suche kann nur eines der dreißig möglichen Felder betreffen. Der Filter kann die Ergebnisse aus der Liste entfernen, die z. B. keinen Hyperlink zum Volltext enthalten, nicht den Menschen als Forschungsobjekt betreffen, nicht in ausgewählten Sprachen verfasst worden sind o. Ä. Das Ergebnis einer Suche kann sehr umfangreich werden. Deshalb wird an das Programm des Benutzers zunächst eine Liste mit Identifikatoren der Artikel geschickt, aus welcher der Benutzer eine, alle oder nur einige auswählen kann¹⁴². Dadurch kann verhindert werden, dass eine Internetverbindung mit geringer Bandbreite überlastet wird oder der Benutzer sehr lange auf die Verarbeitung größerer Datenmengen warten muss. Dieser Zwischenschritt hat zugleich den Nachteil, dass der Benutzer nicht erkennen kann, ob die Ergebnisse für ihn relevant sind. Das Austauschformat ähnelt dem RIS¹⁴³. Anders als bei RAK- oder AACR-Normen werden in PubMed alle Autoren erfasst und ausgegeben. Ein weiterer Unterschied zu MARC besteht darin, dass Pubmed weitere nützliche Informationen wie Zusammenfassung, Sachgebiet und – sofern vorhanden – einen Hyperlink zum Volltext des Artikels.

142 Vgl. das Austauschprotokoll von PubMed / MedLine auf Seite 57 und den Auszug aus der Client-Server-Kommunikation auf Seite 58

143 Vgl. Kapitel „RIS“ auf Seite 38. Das XML-basierende Kommunikationsformat wird hier der besseren Verständlichkeit wegen ausgelassen.

Das MedLine-Protokoll ist viel einfacher als Z39.50 und funktioniert wie eine herkömmliche Internetanwendung. Die Ausgabe wird bei Anfragen aus dem Programm des Benutzers nicht als HTML sondern im PubMed Format ausgeliefert. Wird über die Web-Oberfläche des PubMed (<http://www.ncbi.nlm.nih.gov/sites/entrez/>) gesucht, werden dieselben Daten als HTML formatiert und ausgegeben. Neu und „genial einfach“ gegenüber Z39.50 ist die Nutzung der URI sowie der Recherche auf Basis von HTTP. Dieser technische Konzept ist natürlich um einiges jünger als das von Z39.50.

3.2.4.2. Web-Oberfläche

Unter der frei zugänglichen URL <http://www.ncbi.nlm.nih.gov/sites/entrez> befindet sich das Suchformular für die dreißig Datenbanken und Thesauri des NCBI. Mit der Auswahl einer von ihnen ändert sich die Gestaltung des Formulars und somit die Zahl der für die Recherche zur Verfügung stehenden Felder. Die umfangreichsten Filter- und Suchmöglichkeiten bietet PubMed – die öffentliche medizinische Datenbank des NCBI. Das Suchfeld neben der Datenbank-Auswahlliste ist keinem Datentyp zugeordnet, so dass man für eine genaue Recherche entweder Filter (Reiter mit der Bezeichnung „Limits“) oder den Reiter „Preview/Index“ benutzen sollte. Wünscht der Benutzer eine Übersicht über alle Artikel zu einem Suchbegriff, kann er diesen eingeben und die Abfrage durch Eingabetaste bzw. den Klick auf den Knopf „GO“ starten. Diese oft sehr umfangreiche Liste wird in Schritten von 20, 50, 100, 200 oder 500 Ergebnisse pro Seite angezeigt. Sie kann nachträglich gefiltert werden. Entscheidet sich der Benutzer für die gleichzeitige Suche und Filterung, so steht ihm unter „Limits“ eine große Auswahl an miteinander kombinierbaren Filteroptionen zur Verfügung. Ein besonders interessanter Bereich ist „Search by Autor“. Während der Eingabe wird per AJAX¹⁴⁴ eine Liste der Autorennamen vorgeschlagen, die mit

144 Der Kunstbegriff AJAX steht für Asynchronous JavaScript And XML und bezeichnet eine neu Art der Kommunikation zwischen dem Benutzer (dem genutzten Browser) und dem Server sowie eine neue „Verhaltensweise“ von Internetanwendungen, die ähnlich einer lokalen Anwendung Funktionen oder Bestandteile bei Bedarf nachladen, statt den

gleichen Buchstaben anfangen. Auch eine Mehrfach-Eingabe ist mithilfe mehrerer Felder möglich. In etwa ähnlich funktioniert der Bereich „Search by Journal“. Es fehlt jedoch eine ähnliche Funktionalität im Bezug auf den Thesaurus MeSH¹⁴⁵. Erst am Ende des Reiters findet man die Möglichkeit dem bisher untypisierten Suchfeld einen Typ zuzuordnen. Die im „Limits“-Reiter gemachten Angaben zur Filterung werden vom Browser in das Suchfeld übertragen und mit „AND“ zu einer Konjunktion verknüpft. Jede Eingabe wird mit der dazugehörenden Typ-Angabe in der Form („Suchbegriff“[Typ]) hintereinander im Suchfeld angeordnet. Da nur wenige um diese Syntax wissen und alle nötigen Typen als Kürzel auswendig kennen, ist dieser sehr schnelle Weg zu einer scharfen Suche eher als eine Ausnahme anzusehen.

Neben der vielfältigen Such- und Filter-Möglichkeiten stellt die Internetseite weitere nützliche Funktionen zur Verfügung. Dazu gehört die chronologische Liste der Such-Abfragen (mit Angaben zur Zeit und Anzahl der Ergebnisse), Sortierung der Ergebnisse nach Namen, Zeitschriftentitel oder Datum der Veröffentlichung, Export der Ergebnisse als Quellenangaben (vorformatierter Text), RSS oder das Kopieren der ausgewählten Werke in die Zwischenablage.

Alle diese Funktionen hinterlassen den Eindruck eines vollwertigen Programms. Die imitierte Funktionalität eines PC-Programms basiert auf modernen AJAX-Techniken, die nicht in jedem Web-Browser einheitlich umgesetzt wurden. Spezielle Software für Sehbehinderte wie Screenreader arbeitet mit solchen Lösungen noch nicht optimal zusammen¹⁴⁶. Diese Web-Oberfläche zeichnet sich durch durchschnittliche bis gute Usability, erweist

ganzen Inhalt des Fensters (alle Daten und Funktionen) neu anzufordern. Vgl. dazu: Wenz, Christian : *AJAX. schnell + kompakt*. Frankfurt am Main: EntwicklerPress, 2006. - ISBN 3-935042-92-2 S. 7

145 Vgl. dazu: *Medical Subject Headings*. In: *KSS-Glossar*, S. 81

146 vgl. Kailer, Michaela: *Gütezeichen für barrierefreies Internet : Eine theoretische und empirische Analyse*. Wien, Universität Wien, Fakultät für Wirtschaftswissenschaften, Diplomarbeit, 2009. S. 12.

sich jedoch als problematisch, wenn man die grundsätzlichen Anforderungen der Barrierefreiheit stellt.

3.3. Vorläufiges Fazit

Die bisher beschriebenen bibliographischen Datenformate und Austauschprotokolle reichen Open Source Lösungen bis hin zu Client-Server Anwendungen. Ein Mittelmaß zwischen der Leistungsfähigkeit (hinsichtlich der Recherche), der Austauschbarkeit der Daten und der zitationsgerechten Aufbereitung dieser ist nicht gegeben. Vor allem die exakte Einhaltung der DIN 1505-2 oder ISO 690 ist im Moment aus eben diesen Gründen durch Computerprogramme nicht erfüllbar. Keines der Datenformate – auch das auf Quellenangaben spezialisierte BibTeX-Format – liefert so fein strukturierte Daten, dass sie maschinell fehlerfrei verarbeitet werden könnten. Es bedarf des menschlichen Bearbeiters – vor allem des Wissens um die semantische Bedeutung einzelner Wörter, um dieses Ziel zu erreichen. Ein verstärkter Einsatz computerlinguistischer Verfahren kann dieses Problem teilweise lösen.

Die Art und Weise der Austauschbarkeit von bibliographischen Daten, die in der Einleitung dieser Arbeit anvisiert wurde, ist mit den gegebenen Austauschprotokollen und -Formaten noch nicht erreicht. Die Beziehungen zwischen den Werken gehen nach einer Übertragung größtenteils verloren. Ein direkter Austausch zwischen Benutzern ist kaum möglich. Die Daten werden fast ausschließlich durch die Server großer Bibliotheken oder Bibliotheksverbände ausgeliefert. Ein Vertrauen in die Richtigkeit dieser Daten kann zu Problemen führen. Wünschenswert wäre ein offener elektronischer Katalog, der von jedermann nach einfachen und zugleich möglichst eindeutigen Prinzipien erweiterbar sein sollte. Eine hohe Anzahl an Erschließungsregeln bzw. deren mangelnde Eindeutigkeit scheint nach Meinung des Verfassers teilweise kontraproduktiv zu sein – wie Schwachstellen von RAK-WB oder AACR2 beweisen. Auch ein Meta-Protokoll auf Basis von HTTP für eine Recherche und Austausch von Daten

zwischen Clients nach dem Vorbild eines P2P-Netzwerkes wäre denkbar. Unterschiedliche Meinungen zu einem Werk (in Form von Abstracts) können bisher nur in Social-Bookmarking-Systemen wie Bibsonomy zusammen mit den Metadaten erfasst werden.

Die Annahmen, von denen man bei der Entwicklung der Software bzw. der Standards ausgegangen ist, sind in vielen Bereichen überholt. Die Zeichenkodierung gehört ebenso zu den Altlasten wie die Orientierung am Bibliotheksbestand bzw. Print-Werken. Eine Trennung (oder eher: „Abstraktion“) von Werk und seiner Erscheinungsweise könnte nach Meinung des Autors weiterhelfen. Man denke an die einfache Frage: „Ist eine einzelne Internetseite ein selbständig oder ein unselbständig erschienenes Werk?“ oder an die neuen Möglichkeiten der elektronischen Auslieferung der Werke. Eine zu strikte Festschreibung von Regeln schränkt die Anwendbarkeit eines Systems in der Zukunft ein. Eine zu laxen Handhabung dieser fördert den Wildwuchs einzelner Interpretationen. Dies betrifft z. B. alle bisher genannten Datenformate (mit Ausnahme von XML). Sie können in den nächsten Jahren an ihre Grenzen stoßen, da sie nur eine begrenzte Anzahl an typisierten Datenfeldern beherrschen.

4. Neuere Entwicklungen im bibliographischen Bereich

4.1. Erwartungen

Mit der Idee des „Semantic Web“ – eines weltweiten Netzes, das die Aussagen so formuliert, dass Programme die Inhalte miteinander verknüpfen und dadurch neue Aussagen ableiten können – hat der Internetpionier Tim Berners-Lee im Jahr 2001 den Traum von intelligenten Maschinen neu angesprochen¹⁴⁷. Mit jedem Fortschritt in der Technik erwarten die Menschen eine Entlastung in ihrer Arbeit und träumen von künstlicher Intelligenz, die nicht auf ihre Eingaben wartet, sondern den Wunsch des Benutzers „erahnt“ und von sich aus tätig wird. Diesen Wunsch soll das semantische Netz verwirklichen: Es soll eine kurze Biographie samt allen Veröffentlichungen der Person zusammenstellen, statt nur auf Texte zu verweisen, wo möglicherweise diese Person genannt worden ist. Ein semantisches Netz soll sogar imstande sein, selbständig wahre Aussagen zu formulieren.

Analog entwickeln sich die Erwartungen im Bereich der persönlichen Literaturverwaltung. Nicht mehr eine Liste verfügbarer Werke ist heutzutage interessant, sondern auch die Volltexte zu diesen, die auf Wunsch heruntergeladen werden können. Die Verweise auf die übergeordnete bzw.

¹⁴⁷ Berners-Lee, Tim; Hendler, James; Lassila, Ora: *The semantic web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. Scientific American. 5 / 2001

darin enthaltene bibliographische Einheiten sollen dabei berücksichtigt werden. Natürlich soll die Suche nur die gemeinte Person bzw. das gesuchte Themengebiet berücksichtigen und keine Namensvetter oder gar fremde Sachgebiete ausgeben.

Es sind sehr hohe Anforderungen, die dabei gestellt werden. Dem begegnen die Informatik und die Informationswissenschaft (bzw. die Bibliothekswissenschaft) mit unterschiedlichen Konzepten. Der Idee einer intellektuellen Erschließung und Einordnung innerhalb eines Klassensystems wie Dewey Decimal Classification steht der Gedanke, dass spezifische Wörter mit einer bestimmten Quantität nur in den Texten eines Sachgebietes vorkommen werden, gegenüber¹⁴⁸. Gemeinsam ist beiden Wissenschaften, dass nur die beste Qualität der Basis-Daten zu zufriedenstellenden Ergebnissen führen kann.

Eine weitere Entwicklung der letzten Jahre, die immer mehr an Einfluss gewinnt, sind die sog. Mashups – besser bekannt als „Web 2.0“ bzw. „Mitmach-Web“. Es bezeichnet den neuen Weg der Information und Kommunikation im Internet. Jeder darf seine Inhalte anderen Nutzern zur Verfügung stellen oder diese mit Inhalten anderer kombinieren. Die Information fließt in beide Richtungen: vom Internet zum Benutzer hin und umgekehrt. Die gesellschaftlichen Veränderungen der letzten Jahre, die durch das Internet ermöglicht wurden und in einer kollaborativen Erarbeitung einer Wissensbasis (Wikipedia) münden sind das deutlichste Beispiel für den neuen Umgang mit Informationen¹⁴⁹. Neben den autoritativen Antworten der

148 Die automatische Dokumentenerschließung ist Gegenstand der Forschung von Prof. Harald H. Zimmermann. Im Report zum DFG-Projekt KASCADE wird die Relevanz dieses computerlinguistischen Verfahrens für die elektronischen Kataloge sichtbar (Lepsky, Klaus ; Zimmermann Harald H. : *Katalogerweiterung durch Scanning und automatische Dokumenterschließung . Ergebnisse des DFG-Projekts KASCADE*. In: Zeitschrift für Bibliothekswesen und Bibliographie 4/00, S. 305-316. URL: <http://is.uni-sb.de/zimmermann/pdf/2000a.pdf> Stand: 15.12.2008). Ein ähnliches – aber stark vereinfachtes Verfahren – kommt in den Suchmaschinen wie Google Scholar vor.

149 Surowiecki, James: *The Wisdom of Crowds. Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*.

älteren Medien wie Presse oder Rundfunk scheinen vielfach bestätigte Informationen mehrerer Internetquellen ebenso glaubwürdig zu sein (z.B. Blog- und Twitter-Meldungen zu Unruhen in Iran nach Präsidentenwahlen im Jahre 2009). Das Web 2.0 ist ein anarchisches (pluralistisch und unzensiert) und komplementäres Medium¹⁵⁰. Jeder darf Informationen beitragen und so die Datenbasis, aus der Wissen generiert werden kann, erweitern.

Dies trifft ebenfalls auf das Bibliothekswesen zu. Es gibt keine Bibliothek, die das ganze Schrifttum der Welt erfasst hätte¹⁵¹. Einige Werke sind zudem fehlerhaft erfasst worden (vgl. Beispiele aus Library of Congress auf Seite 86). Die wissenschaftlichen Artikel werden nur an wenigen Orten so genau und akribisch erfasst wie bei PubMed. Dennoch enthalten auch solche Fachdatenbanken nicht alle Artikel eines Sachgebietes. Da die wissenschaftliche Kommunikation seit einigen Jahrzehnten immer mehr in Form von Artikeln geschieht, stellt dieser Mangel an vollständigen Artikeldatenbanken (für Print-Ausgaben) ein immer größeres Hindernis für diese Tätigkeit dar. Im Bereich der elektronischen Publikation sollen die

Doubleday, 2004; ISBN: 0385503865. Das Buch setzt sich mit dem Phänomen der sog. Kollektiven Intelligenz auseinander und zeigt unter welchen Bedingungen das kollektive Wissen dem Expertenwissen ebenbürtig sein kann. Zugleich markiert das Erscheinen dieses Buches den Beginn einer Bewegung ("Crowdsourcing" genannt), die auf die Zusammenarbeit vieler freiwilliger Teilnehmer als zuverlässige und zudem kostengünstige Informations- und Wissensquelle setzt. Dies ist eines der Prinzipien von Wikipedia.

150 Zur Vision und Realität einer neuen, von Internet geprägten Gesellschaft, schreibt der Erfinder des Copyleft-Prinzips Laurence Lessig in seinem Buch "Freie Kultur. Wesen und Zukunft der Kreativität" (The Pinguin Press, New York 2004) Folgendes: "Diese Form der Rede führt zu mehr Demokratie. Laut Winer vor allem deshalb, weil man 'nicht mehr für eine kontrollierende Instanz, einen Torwächter arbeiten muss'. Das stimmt. Aber sie führt noch auf eine andere Weise zu mehr Demokratie. Wenn immer mehr Bürger ihre Gedanken ausdrücken und rechtfertigen, ändert das die Art, wie die Leute öffentliche Angelegenheiten verstehen."

151 Das World-Cat-Projekt (<http://www.worldcat.org/>) von Online Computer Library Center, Inc. (OCLC) ist ein gemeinsamer Katalog vieler Bibliotheken und wird manchmal als eine solche Stelle angesehen. Es beinhaltet jedoch nur Daten zu selbständig erschienenen Einheiten – entsprechend der Erfassung der einzelnen Bibliotheken.

Publikationsserver der Universitäten gemäß der Berliner Erklärung zum Open Access¹⁵² Abhilfe schaffen. Auch hier sind die Export-Funktionen für die Quellenangaben nur sehr selten vorgesehen. Aus diesem Grund ist der Austausch von Literaturquellen mit Medienbrüchen verbunden. Die für ein bestimmtes Thema aufwendig gesammelte und geordnete Literatur wird bestenfalls als ISBD- oder RIS-formatierte Liste per Email an den Mitarbeiter oder Studenten übermittelt. Dieser kann daraus eine eigene Datenbank zusammenstellen. Dieser Zustand erinnert an die „negative Utopie“ die Umberto Eco in seinem Vortrag zum fünfundzwanzigjährigen Jubiläum der Mailänder Stadtbibliothek formuliert hat. Obwohl seit dem Zeitpunkt über zwanzig Jahre vergangen sind und einige technische Hilfsmittel die damaligen Kritik heute nicht so scharf erscheinen lassen¹⁵³, ist der Umgang des (angehenden) Wissenschaftlers mit diesen Daten immer noch wenig komfortabel. Es ist somit im Sinne des Bibliotheksbenutzers (der immer öfter in Form eines Online-Rechercheurs die Bibliothek aufsucht¹⁵⁴) als auch des Bibliothekars, diese Barrieren so weit wie möglich abzubauen.

An dieser Stelle wird die nicht ausgesprochene Erwartung klarer: Ein Peer-2-Peer System, in dem man die eigenen Quellen anderen zur Verfügung stellt und in der Literatur seiner Kollegen suchen kann. Medienbrüche sollen vermieden werden. Vielmehr sollen die ausgewählten Werke (nach Möglichkeit samt Kontext) übernommen werden können. Ein solches System erfordert nicht nur ein eigenes Austausch-Protokoll sondern auch ein geeignetes Format, in dem die Zusammenhänge (Hierarchie, Einordnung, Bemerkungen etc.) erhalten bleiben. Die bisher genannten Formate sind diesen Anforderungen nicht gewachsen.

152 Vgl. *Open Access*. In: *KSS-Glossar*, S. 91f und *Budapest Open Access Initiative*. ebd S. 16

153 Wagner-Döbler, Roland: *Umberto Ecos Betrachtung einer benutzerfeindlichen Bibliothek – 25 Jahre danach*. In: *Festschrift für Walther Umstätter*; CD-ROM

154 Mehr zum Suchverhalten der Wissenschaftler in: Havemann, Frank; Kaufmann, Andrea: *Der Wandel des Benutzerverhaltens in Zeiten des Internet – Ergebnisse von Befragungen an 13 Bibliotheken*. In: *Festschrift für Walther Umstätter*; CD-ROM

4.2. Umsetzung

Die SGML-basierte Auszeichnungssprache XML besteht aus wenigen Syntax-Regeln (einer Notation) und ist universal einsetzbar (im Gegensatz zu den davon abgeleiteten Formaten, die diese Syntax um Semantik erweitern). Anders als z.B. das HTML ist XML nicht an ein bestimmtes Einsatzgebiet hin ausgerichtet. Es kann sowohl bei narrativen als auch datensatzartigen Dokumenten eingesetzt werden. Mit DTD, XMLS¹⁵⁵ (XML-Schema) oder Relax NG¹⁵⁶ (ebenfalls eine XML-basierte Schema-Beschreibungssprache) kann man festlegen, wo und wie oft ein Element vorkommen darf. Was dieser bedeuten soll, ist – mit Ausnahme von RDF – zunächst nur für den Menschen verständlich. Das Resource Description Framework – bestehend aus einigen wenigen Auszeichnern – definiert eine Syntax zur Beschreibung von Entitäten (Ressourcen). „In RDF formulierbare Aussagen ordnen einer Entität eine Eigenschaft (bzw. eine Beziehung) und einen zugehörigen Eigenschaftswert (bzw. eine in Beziehung stehende Entität) zu.“¹⁵⁷ Wenn diese Eigenschaften durch Beziehungen zu anderen Eigenschaften (wie „konträr“, „kontradiktorisch“ oder „komplementär“) beschrieben wird (z. B. in Ontologien), kann auch ein Computer aus der Beschreibung schließen, dass ein dickes Buch nicht dünn sein kann. Dies bedeutet nicht, dass der Computer diese Aussage verstehen kann. (In diesem Fall ist es eine Tautologie, die auf der Kontradiktion der Begriffe "dick" und "dünn" basiert.) Der Computer wendet nur die programmierte Logik auf die Ausgangsdaten an. Diese Art der Datenverarbeitung ergibt einen gewissen Fortschritt: das

155 Ein von W3C entwickelter Standard für die Beschreibung von XML-basierten Dateiformaten (<http://www.w3.org/XML/Schema>)

156 Relax NG (<http://relaxng.org/>) wurde unter der Beteiligung der OASIS (Organization for the Advancement of Structured Information Standards, die unter anderem an der Entwicklung des OpenDocument-Formats [ISO/IEC 26300:2006] beteiligt war) entwickelt und stellt eine Alternative zu XMLS dar. Der grundsätzliche Unterschied zwischen XML-Schema und Relax NG besteht darin, das XML-Schema sehr genaue (strikte) Vorgaben zum Inhalt von Elementen machen kann. Relax NG definiert hingegen nur wenige syntaktische Regeln und ist somit übersichtlicher.

157 *RDF*. In: *KSS-Glossar*, S. 103

Programm kann die Eingabe des Benutzers und die nicht genannten Eigenschaften oder Beziehungen miteinander kombinieren, um bestimmte Entitäten ein- oder auszuschließen. Bisher gelten RDF und Web-Ontologien als wenig "intelligent".

„RDFS ist mächtig, aber nur auf begrenztem Raum. Die damit gewonnenen Fähigkeiten, logische Schlüsse zu ziehen, ähneln eher denen eines Kleinkindes als denen eines erwachsenen Menschen.“¹⁵⁸

Die hier genannten Ideen und Formate dienen einem einzigen Zweck: dem Aufbau des besagten semantischen Netzes. Die einzelnen Entwicklungen bringen im Gegensatz zum zukünftigen Semantic Web schon heute einen großen Vorteil. So lassen sich die Daten genauer typisieren und miteinander verknüpfen. Die Verwendung von URI, XML (mit Namensräumen) oder Unicode sorgt für „Verständigung“ trotz unterschiedlicher Notationen und Zeichensätze. Vor allem aus diesen Gründen versucht man alte Formate mit Hilfe von XML abzubilden oder neue (auf XML-Basis) zu entwickeln. Zu den ersten gehören BibTeXML (eine XML-Notation von BibTeX), MARCXML oder RISXML. Zu den neuen – aber nicht immer neuartigen – Formaten zählen: MODS oder ONIX.

4.2.1. ZING: SRU (SRW)

Das ZING (Abkürzung für: „Z39.50 International: New Generation“) ist keine vollständige Neuentwicklung. Es ist ein neues Protokoll, welches auf HTTP aufbaut, die Contextual Query Language (CQL) für die Abfrage und moderne Datenaustauschformate wie MODS für die Auslieferung der Daten benutzt. SRU steht für Search/Retrieval via URL. Analog entstand der Begriff SRW (Search/Retrieve Web Service), der im Grunde dasselbe meint. Die wenigen

158 Ziegler, Cai: *Sinn oder nicht Sinn. Vom Suchen und Finden der Semantik im Web.* c't 21 / 2007. S. 173. S. 174 (Dies ist keine endgültige Beurteilung. Im Folgenden beschreibt der Autor die Ontologien, die RDF nutzen kann, um weiter reichende Schlüsse zu ziehen. Bringt aber auch diese Möglichkeit nicht ausdrücklich in Verbindung mit der künstlichen Intelligenz.)

Unterschiede zu SRU wie Authentifizierung, Einbindung von Erweiterungen und die Unterstützung von Web Services sind für diese Beschreibung irrelevant. 2006 entschied sich LOC, den alten Begriff ZING zu verwerfen und ausschließlich SRU sowie SRW zu nutzen. (In dieser Arbeit werden ZING, SRU, SRW oder SRU/W synonym verwendet.)

Die Funktionsweise von SRU/W gleicht der des Z39.50 und des PubMed. Die Gemeinsamkeiten mit PubMed sind: HTTP als Basis für Anfragen und Antworten sowie die Möglichkeit der Nutzung von CGI-Skripten als Vermittler, der gleichzeitig die Web-Oberfläche und HTML-Ausgabe auf dem Gateway bedient. Mehr Gemeinsamkeiten weist SRU mit Z39.50 auf: die Datenbank-unabhängige Abfragesprache (CQL), die Erweiterbarkeit auf jegliche Sachgebiete und die nahezu unbegrenzten Möglichkeiten der Kombination der Suchfelder. Die Kommunikation zwischen dem Server und dem Client verläuft ähnlich wie bei Z39.50: Eine Abfrage wird vom Client mit Hilfe von CQL Datenbank-unabhängig formuliert und zum Server abgeschickt. Der Server übersetzt an Hand der hinterlegten Regeln diese Anfrage in eine oder mehrere (je nach abzufragender Datenbank) interne Abfragen. Dabei kann er auch (wie ZACK) die Abfrage des Benutzers an andere Server – wie z. B. OAI-Datenbanken (Open Archive Initiative) – senden und die Ergebnisse in die eigene Ergebnis-Liste einfügen. Das Verhalten des Servers kann durch den Benutzer z. B. durch die Auswahl bestimmter Datenbanken gesteuert werden. In der Anfrage des Clients (die wie eine gewöhnliche URL [„REST“ genannt] oder SOAP formuliert sein kann¹⁵⁹) wird neben der Suchanfrage selbst die gewünschte Datenbank und das zu nutzende Datenformat definiert. Der Server übersetzt die Daten in das gewünschte Datenaustauschformat und liefert es aus.

159 REST (Representational State Transfer) bezeichnet die einfachste Art der Übermittlung der Abfrage an den Server. Dies geschieht durch die Übergabe von Variablen-Namen und deren Wert in Form von „Variable=Wert“ innerhalb einer URL. SOAP (früher als Abkürzung für Simple Object Access Protocol) ist eine fortgeschrittene und komplexe Methode der Übergabe einer Anfrage an den Server. Sie ist genauso wie XML-RPC in XML formuliert. Sie kann XML-Namensräume nutzen und Metadaten beinhalten. (Vgl. dazu XML i a Nutshell, S. 269-271).

Das SRU-Protokoll nutzt die im Web gängigen Methoden, Formate und Protokolle, um die Funktionalität bereitzustellen, die schon Z39.50 leistete. Der größte Vorteil des neuen internationalen Retrieval-Protokolls ist eben diese Vereinfachung durch Nutzung anerkannter Web-Standards. Die neuen Elemente dieses Protokolls sollen im Folgenden kurz dargestellt werden¹⁶⁰.

4.2.1.1. SRU-Operationen

Das SRU ist eine Weiterentwicklung des Z39.50. Dies beweist nicht zuletzt die Übernahme der drei möglichen Operationen, die ein SRU-Gateway ausführen kann. Diese sind¹⁶¹: Explain, Search/Retrieve und Scan. Mit Explain (de. erklären) beschreibt der Server die ihm bekannten XML-Formate, verfügbare Indexe und Datenquellen. Diese Information kann die Benutzersoftware dazu nutzen, um z. B. nur die dem Server bekannten Felder (Indexe) oder Datenbanken in der Suchmaske zur Verfügung zu stellen. Ebenso wichtig ist die Abstimmung von Server und Client auf ein Datenaustauschformat. Dies geschieht anhand der Informationen aus dem Explain-Request. Darauf folgt die eigentliche Suche und die Auslieferung der Ergebnisse. Dabei kann der Client vorgeben, welche Anzahl an Ergebnissen ausgeliefert werden soll und wie diese angeordnet werden sollen. Dies geschieht mit Hilfe der CQL. Anschließend kann sich der Client auf die bereits gestellte Abfrage berufen und diese eingrenzen oder erweitern. Auf diese Art und Weise erfährt der Benutzer, wie viele Treffer es im „Umfeld“ seiner Suche gab (z.B.: Was wäre wenn nur „Information“ statt „Informationswissenschaft“ gesucht worden wäre – und umgekehrt).

160 Da diese Arbeit die Bibliotheken im Fokus hat, werden andere Möglichkeiten der SRU-Recherche nach anderen Informationen ausgeblendet. Denkbar wäre z. B. die Suche nach Personen, technischen Informationen o. Ä.

161 Angaben nach McCallum; Sally H.: *A Look at New Information Retrieval Protocols: SRU, OpenSearch/A9, CQL, and XQuery*. (Vortrag an: World Library And Information Congress: 72nd IFLA General Conference And Council, 20-24.08.2006, Seoul, Korea) URL: <http://www.ifla.org/IV/ifla72/papers/102-McCallum-en.pdf> (Stand: 8.4.2010)

4.2.1.2. CQL als Abfrage-Sprache des SRU/W

„CQL, the Contextual Query Language, is a formal language for representing queries to information retrieval systems such as web indexes, bibliographic catalogs and museum collection information. The design objective is that queries be human readable and writable, and that the language be intuitive while maintaining the expressiveness of more complex languages.“¹⁶²

Library of Congress als Entwicklungsgremium des Z39.50 versuchte die bisherige Abfragesyntax zu „reformieren“ und einen Standard zu schaffen, der gleichermaßen für Menschen und Maschinen verständlich wäre. Dabei sollte die Unabhängigkeit dieser Abfragesprache von bestimmten Datenbanken und Kontexten gewahrt bleiben. Wie McCallum zurecht bemerkt, ist diese Reform nur zum Teil gelungen: „Thus is more powerful than a simple Google-like language.“¹⁶³ Diese Syntax ist in der Tat sehr leistungsfähig und unabhängig von verwendetem Kontext (z. B. einer Datenbank). Sie nutzt eine in der Programmierer-Welt gängige Vorgehensweise, indem der zu verwendete Kontext (auch „Namensraum“ genannt) dem Modifikator vorangestellt wird und mit einem Punkt getrennt wird (sog. Prefix). Enthält ein Modifikator keine Information zum Kontext, wird der Sprachumfang von CQL angenommen. Auf diese Weise kann der Gateway eigene Modifikatoren zur Verfügung stellen und CQL erweitern. Interessant ist diese Möglichkeit, wenn man dem Benutzer eine besondere Suche ermöglichen will (z. B. „Herausgegeben um das Jahr 2000“ statt des üblichen „Herausgegeben im Jahr 2000“).

¹⁶² Library of Congress (Hrsg.): *CQL: the Contextual Query Language: Specifications (SRU: Search/Retrieval via URL, Standards, Library of Congress)*. URL: <http://www.loc.gov/standards/sru/specs/cql.html> (Stand: 11.11.2007)

¹⁶³ McCallum; Sally H.: *A Look at New Information Retrieval Protocols: SRU, OpenSearch/A9, CQL, and XQuery*. S. 5 (Vortrag an: World Library And Information Congress: 72nd IFLA General Conference And Council, 20-24.08.2006, Seoul, Korea) URL: <http://www.ifla.org/IV/ifla72/papers/102-McCallum-en.pdf> (Stand: 8.4.2010)

Die einfachsten Suchanfragen werden wie üblich durch die Verbindung eines Index-Feldes mit einem Wert mit Hilfe einer Relation (wie „any“, „=“, „>“, „<“, „<=“, „=>“ etc) formuliert. Wird in mehr als einem Feld gesucht, müssen beide durch einen booleschen Operator miteinander verknüpft werden („and“, „or“, „not“ und „prox“). Einige von ihnen können modifiziert werden, indem man zusätzliche Angaben (wie z. B. die Entfernung der gesuchten Wörter voneinander) macht oder einen zusätzlichen Suchalgorithmus hinzuzieht. An das Ende der Suchanfrage kann eine Sortierungsklausel angehängt werden, die bestimmt, welcher Index in welcher Reihenfolge sortiert werden soll.

Mit Hilfe dieser Syntax kann man jede denkbare Anfrage in eine SQL-Abfrage-Syntax der jeweiligen Datenbank abbilden. Ob diese auch für den ungeübten Menschen verständlich oder nachvollziehbar ist, darf bezweifelt werden. Es gilt hingegen als sicher, dass diese Syntax die notwendige Ausdruckstärke für ein kontextunabhängiges und sehr präzises Retrieval hat.

4.2.1.3. XML-basierte Datenaustauschformate

Der gesamte Fortschritt des SRU fände wohl kaum Beachtung in der Computer-Welt, wenn das XML – das Datenformat der Mashups – nicht berücksichtigt worden wäre. Die Möglichkeit, einzelne Datenfelder genau zu beschreiben, ist im bibliographischen Bereich nahezu unverzichtbar. Atomare Aussagen¹⁶⁴ wie Autorennamen, Titel oder Verlagsort ergeben erst als Ganzes eine bibliographische Beschreibung. Die Möglichkeit einer starken Unterteilung und genauen Typisierung der Daten, die XML bietet, wurde zum Anlass für einen Umstieg oder eine Weiterentwicklung auf dieser Basis zu realisieren. Zu den zweiten zählen MARCXML und MABxml.

4.2.1.3.1. MARCXML und MABxml

164 In dieser Arbeit wird der Begriff "atomare Aussage" analog zum eben diesem Begriff in der formalen Logik verwendet und bezeichnet den kleinsten Bestandteil einer bibliographischen Beschreibung z.B. den Vornamen des Autors (und nicht den vollständigen Namen).

„XML ist eine Technologie, die sich bereits auf breiter Basis (insbesondere als Syntax für Austauschformate) durchgesetzt hat. Viele Software-Anwendungen wurden für XML und die damit verwandten Technologien wie XML-Schema und XSL konzipiert. So erlaubt beispielsweise das viel beachtete OAI-Protokoll ausschließlich den Austausch von Daten, die eine XML-konforme Syntax haben. So ergab sich der Wunsch, auch MAB-Daten in einer XML-Struktur zu transportieren. Mit MARCXML - einem XML-Schema für die Eins-zu-Eins-Übertragung von MARC-21-Datensätzen in eine XML-Struktur - hat die Library of Congress auf diese Entwicklung reagiert. MABxml soll eine vergleichbare Funktion für MAB2 erfüllen.“¹⁶⁵

Diese kurze Einführung in MABxml auf der Internetseite der Deutschen Nationalbibliothek skizziert sehr gut die Motivation und zugleich das Ergebnis der Bemühungen um ein bibliographisches Datenformat auf Basis von XML. Es trifft auf die Nationale bzw. Internationale Entscheidungsinstanzen im Bereich der bibliographischen Standards genau zu. Das bereits Erarbeitete soll lediglich an die neuen Standards angepasst werden. Die inhaltliche Diskussion, ob und inwieweit ein MAB2- oder MARC21-Standard der tatsächlichen Entwicklung in der Datenverarbeitung angepasst werden sollte, wurde trotz eines vielfach monierten Bedarfs von Seiten der praxisnahen Bibliothekare nicht geführt. Folgende Gründe sind für diese Situation denkbar:

- Die Daten sind intern so abgelegt, dass sie nicht anders darstellbar sind.
- Der Standard ist zuverlässig und allgemein anerkannt, so dass jegliche Änderung zur Verwirrung des Benutzers führen würde.
- Der erarbeitete Standard bedarf in seiner Vollkommenheit weder einer Revision, noch einer grundlegenden Überarbeitung.

¹⁶⁵ Deutsche Nationalbibliothek (Hrsg.): *MABXML*. URL: <http://www.d-nb.de/standardisierung/formate/mabxml.htm> (Stand: 13.8.2011)

```
<?xml version="1.0" encoding="UTF-8"?>
<datei xmlns="http://www.ddb.de/professionell/↵
mabxml/mabxml-1.xsd" xmlns:xsi="http://www.w3.org/2001/↵
XMLSchema-instance" xsi:schemaLocation="http://www.ddb.de/↵
professionell/mabxml/mabxml-1.xsd
http://www.ddb.de/professionell/mabxml/mabxml-1.xsd">
  <datensatz id="ID946451494" typ="h" status="n"
mabVersion="M2.0">
  <feld nr="001" ind=" " >946451494</feld>
  <feld nr="004" ind=" " >19980605</feld>
  <feld nr="030" ind=" " >a|liar|z|||37</feld>
  <feld nr="036" ind="a">DE</feld>
  <feld nr="050" ind=" " >a|a|||||||</feld>
  <feld nr="051" ind=" " >m|||z||</feld>
  <feld nr="100" ind="b">Ginsbach, Julia</feld>
  <feld nr="331" ind=" " ><ns>Die</ns> schÄ¶ne Lau</feld>
  <feld nr="335" ind=" " >nach dem Stuttgarter
HutzelmÄ¶nnlein von Eduart MÄ¶rike</feld>
  <feld nr="359" ind=" " >Julia Ginsbach (Bilder) ; Andrea
Liebers (Text)</feld>
  <feld nr="410" ind=" " >Leinfelden-Echterdingen</feld>
  <feld nr="412" ind=" " >DRW-Verl. Weinbrenner</feld>
  <feld nr="425" ind=" " >1995</feld>
  <feld nr="540" ind="a">ISBN 3-87181-371-0 Pp. : DM
24.80</feld>
  <feld nr="800" ind=" " >118583107 MÄ¶rike, Eduard</feld>
  <feld nr="805" ind="b"><ns>Das</ns> Stuttgarter
HutzelmÄ¶nnchen</feld>
  </datensatz>
</datei>
```

Text 12: Beispiel für eine MABxml-Datei nach: http://www.d-nb.de/standardisierung/formate/mabxml_beispiel_ebene1.xml. (Stand 20.01.2011, gekürzte Fassung)

Die fehlerhaften Zeichen deuten auf Zeichensatz-Konvertierungsfehler während des Exports des Datensatzes hin: UTF-8-Text wurde wie ISO-8859-1 behandelt und erneut in UTF-8-kodiert. (Wahrscheinlich ein Versehen, da die Konvertierung dieser beiden Zeichensätze prinzipiell problemlos funktioniert.)

Im vorhergehenden Teil dieser Arbeit wurde nachgewiesen, dass es genügend Gründe gäbe, an einer Revision zu arbeiten. Die Anerkennung, die diese Standards genießen, kann nur unter der folgenden Bedingung als Argument gelten: die erfassenden Bibliothekare sind zur Einhaltung der AACR2- bzw. RAK-Regeln verpflichtet. Auf diese Regeln sind MAB bzw. MARC21 ausgerichtet. Neue XML-basierten Formate können hingegen einen Großteil der Erfassungsregeln überflüssig machen. Die Entscheidung, ob

unter „Beethoven, Ludwig van“ oder unter „van Beethoven, Ludwig“ katalogisiert werden soll, entfällt, wenn für jeden Bestandteil des Namen ein eigenes Feld vorgesehen wäre. Für die LoC kann auch der zuerst genannte Grund nicht voll zutreffen: die mittels SRU abgefragten Datensätze können auch in neueren Formaten wie MODS exportiert werden (wenn auch nicht ganz so atomar, wie es für eine bessere Verarbeitung nützlich wäre). MARCXML bzw. MABxml stellen eine temporäre Lösung dar. Sie liefern alte Daten in einem XML-Format aus. Die Datenqualität bleibt dabei unverändert.

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01142cam 2200301 a 4500</leader>
    <controlfield tag="001"> 92005291 </controlfield>
    <controlfield tag="003">DLC</controlfield>
    <datafield tag="100" ind1="1" ind2=" ">
      <subfield code="a">Tan, Chung Lee</subfield>
      <subfield code="d">1949-</subfield>
    </datafield>
  </record>
</collection>
```

Text 13: Beispiel für eine MARC 21 XML Datei. Die Kontrollfelder und die Datenfelder sind klar voneinander zu unterscheiden. Die tag-Attribute beinhalten die Daten-Typ-Nummern, die code-Attribute die Unterfelder-Bezeichner.

Beide Formate sind ein Abbild ihrer Vorgänger. Der erste und größte Vorteil ergibt sich aus der Nutzung des Unicode: Die einheitliche Zeichenkodierung sorgt für „Verständigung“ trotz unterschiedlicher Zeichensätze. Die Transkription oder die Transliteration wird dadurch nicht abgeschafft (sie wird weiterhin für die an die Sprache angepasste Schreibweise benötigt). Der zweite Unterschied und zugleich Vorteil der XML-basierten Formate ist die Aufteilung der Teilfelder mit Hilfe von Tags: Im MARCXML als „subfield“-Element des Elementes „datafield“¹⁶⁶; im MABxml werden die Indikatoren im Attribut „ind“ zusammen mit der Feldnummer (Attribut „nr“) angegeben¹⁶⁷.

166 Vgl. Illustrierte Darstellung der DTD von MARCXML. Library of Congress (Hrsg.): *Schema MARC21slim.xsd*. URL: <http://www.loc.gov/standards/marcxml/xml/spy/✓spy.html> (Stand: 15.12.2008)

167 Vgl. Kett, Jürgen: *MABxml-1*. Dokumentation. URL: http://www.d-nb.de/✓/standardisierung/pdf/mabxml_1_dok.pdf Version 1.2, (Stand 20.01.2011) sowie

Durch die XML-Nutzung entfällt der Einsatz kryptischer Zeichenkombinationen. An ihre Stelle treten öffnende und schließende Auszeichner, die einzelne Informationen voneinander trennen. Auch eine Validierung der Datei ist vor dem Import möglich. Eine solche Syntax-Überprüfung kann mit jedem beliebigen XML-Validierungsprogramm oder -Internetdienst durchgeführt werden.

4.2.1.3.1. MODS

MODS (Metadata Object Description Schema) sowie MADS (Metadata Authority Description Schema) wurden vom Büro für MARC Standards (der Library of Congress) entwickelt. Die Motivation und der Zweck dieser neuen Auszeichnungssprache für bibliographische Daten wird auf der Internetseite des LOC wie folgt beschrieben:

„The Network Development and MARC Standards Office, with interested experts, has developed a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. As an XML schema, the "Metadata Object Description Schema" (MODS) is intended to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records.“¹⁶⁸

Grundsätzlich ist dieses Format als menschenlesbares Format für MARC-Daten anzusehen. Es ist nicht so komplex wie MARCXML und die Daten aus MODS-Dateien können in MARCXML zurück konvertiert werden – jedoch wird die daraus entstandene MARCXML-Datei nicht so detailreich sein, wie die ursprünglichen MARC-Daten¹⁶⁹. Der Grund dafür ist wohl darin zu

Kett, Jürgen: *Regeln zur Übertragung von MAB2-Datensätzen nach MABxml-1*. URL: http://www.d-nb.de/standardisierung/pdf/mabxml_1_uebertr.pdf Version 1.0, (Stand: 20.01.2011)

168 Library of Congress (Hrsg.): *MARC in XML*. URL: <http://www.loc.gov/marc/marcxml.html> (Stand: 01.12.2007)

169 Library of Congress (Hrsg.): *MODS: Uses and Features*. URL: <http://www.loc.gov/standards/mods/mods-overview.html> (Stand: 14.12.2007)

suchen, dass MODS bewusst auf die größte mögliche Einfachheit und Lesbarkeit hin entwickelt wurde. Felder und Teilfelder werden nicht nummerisch, sondern mit Begriffen wie „name“, „subject“, „typeOfResource“ oder „genre“ umschrieben.

MODS ist ein XML-basiertes Format und eignet sich zur Übertragung von Metadaten im SRU, als Erweiterung von Metadata Encoding and Transmission Standard (METS), als Format für Metadaten für Harvesting oder für die Einbettung in elektronischen Ressourcen¹⁷⁰. Es ist umfangreicher als Dublin Core und stärker auf die Bibliotheken bezogen als ONIX. Zusätzlich setzt es die Functional Requirements for Bibliographic Records (FRBR) modellhaft um. Es nutzt die Möglichkeiten von XML aus, indem es die interne Verlinkung von Elementen durch Nutzung von ID und IDREF zulässt. So werden die Redundanzen vermieden. Wie viele andere bibliographischen Formate erlaubt es die Einbindung von Schlüsselwörtern, Abstracts und Angaben zum Standort des Werkes (physikalisch oder als URL-Adresse). Es ist für die Beschreibung elektronischer Ressourcen im Vergleich zu anderen Formaten besser geeignet, da es die Quelldaten der Ressource selbst um alternative Angaben erweitert (Titel, Ort der archivierten Version etc.).

Die Vorteile von MODS werden vor allem auf dem Gebiet der Erschließung von Zeitschriften oder Artikeln sichtbar. Genaue Angaben zur Serie oder der übergeordneten (selbständig erschienenen) bibliographischen Einheit, die Relation zwischen dem beschriebenen Werk und der übergeordneten Einheit können darin genau festgehalten werden (host – part).

Die Schwachstellen von MODS sind:

- die Angaben zu Personen
- die fehlende Verlinkung zwischen den verschiedenen „Arten“ derselben „Ressource“.
- Bei Personennamen wurden nur zwei Felder vorgesehen: „namePart“ und

¹⁷⁰ ebd.

„displayForm“, die den vollständigen Namen der Person in zwei Formen darstellen. (Der Grund dafür sind die nach AACR erfassten, MARC-konformen Daten.)

Eine Aufteilung von Namensbestandteilen brächte einen nicht unerheblichen Fortschritt für die automatische Generierung von Quellenangaben. Eine Verlinkung und Nennung der Relation z. B. zwischen dem Original und der Übersetzung oder zwischen unterschiedlichen Schreibweisen des Namen von Personen (bzw. unterschiedlichen Namen derselben Person) könnten bei der Suche sehr hilfreich sein. Da dies bei der Suche in elektronischen Katalogen der Bibliotheken zum großen Teil automatisch geschieht, hat man diese Möglichkeit nicht beachtet. Damit gehen während des Exports einige Daten verloren.

An dieser Stelle sollen zwei Beispiele für MODS-Dateien dargestellt werden. Im ersten wird eine Homepage beschrieben. Im zweiten Beispiel wird ein Artikel aus einer Zeitschrift aufgelistet. Beide Beispiele stammen von der Internetseite des Standardisierungsgremiums der LOC.


```

<modsCollection xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/mods/v3"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
<mods version="3.0">
  <titleInfo>
    <title>FranUlmer.com -- Home Page</title>
  </titleInfo>
  <titleInfo type="alternative"><title>Fran Ulmer, Democratic
candidate for Governor, Alaska, 2002</title>
</titleInfo>
  <name type="personal">
    <namePart>Ulmer, Fran</namePart>
  </name>
  <genre>Web site</genre>
  <originInfo>
    <dateCaptured point="start" encoding="iso8601">20020702
</dateCaptured>
    <dateCaptured point="end" encoding="iso8601">
20021203</dateCaptured>
  </originInfo>
  <language>
    <languageTerm authority="iso639-2b">eng</languageTerm>
  </language>
  <physicalDescription>
    <internetMediaType>text/html</internetMediaType>
    <internetMediaType>image/jpg</internetMediaType>
  </physicalDescription>
  <abstract>...</abstract>
  <subject>
    <topic>Elections</topic><geographic>Alaska</geographic>
  </subject>
  <relatedItem type="host">
    <titleInfo>
      <title>Election 2002 Web Archive</title>
    </titleInfo>
    <location>
      <url>http://www.loc.gov/minerva/collect/elec2002/</url>
    </location>
  </relatedItem>
  <location>
    <url displayLabel="Active site (if
available)">http://www.franulmer.com/</url>
  </location>
  <location>
    <url displayLabel="Archived site">http://wayback-
cgil.alexandria.com/e2002/*/http://www.franulmer.com/</url>
  </location>
</mods>
</modsCollection>

```

Text 14: Gekürztes Beispiel für eine MODS-Datei mit Daten zur Homepage.
Quelle: <http://www.loc.gov/standards/mods/v3/mods-userguide-examples.html>
(Stand: 22.12.2007)

```

<modsCollection ...>
<mods version="3.0">
  <titleInfo>
    <title>Hiring and recruitment practices in academic
libraries</title>
  </titleInfo>
  <name type="personal">
    <namePart>Raschke, Gregory K.</namePart>
    <displayForm>Gregory K. Raschke</displayForm>
  </name>
  <typeOfResource>text</typeOfResource>
  <genre>journal article</genre>
  <originInfo>
    <place>
      <placeTerm type="text">Baltimore, Md.</placeTerm>
    </place>
    <publisher>Johns Hopkins University Press</publisher>
    <dateIssued>2003</dateIssued>
  </originInfo>
  <language>
    <languageTerm authority="iso639-2b">eng</languageTerm>
  </language>
  <physicalDescription>
    <form authority="marcform">print</form>
    <extent>15 p.</extent>
  </physicalDescription>
  <relatedItem type="host">
    <titleInfo>
      <title>portal: libraries and the academy</title>
    </titleInfo>
  </relatedItem>
  <originInfo>
    <issuance>continuing</issuance>
  </originInfo>
  <part>
    <detail type="volume">
      <number>3</number>
    </detail>
    <detail type="level">
      <number>2</number>
    </detail>
    <extent unit="pages">
      <start>53</start>
      <end>67</end>
    </extent>
    <date>Jan. 2003</date>
  </part>
</relatedItem>
</mods>
</modsCollection>

```

Text 15: Gekürztes Beispiel für eine MODS-Datei mit Daten zum Zeitschriftenartikel. Quelle: <http://www.loc.gov/standards/mods/v3/mods-userguide-examples.html> (Stand: 22.12.2007)

4.2.1.3.1. ONIX

ONIX (Online Information eXchange) ist ein weiteres Format für Metadaten. In dieser Arbeit soll der „Dialekt“ namens „ONIX for books“ kurz dargestellt werden. Dieses Format wurde (und wird weiterhin) entwickelt von EDItEUR in Zusammenarbeit mit Book Industry Communication (UK), Book Industry Study Group (USA) unter Beteiligung von Benutzern aus EU-Ländern sowie Australien, Kanada und der Demokratischen Republik Kongo¹⁷¹. EDItEUR beschreibt das Format wie folgt:

„The ONIX for Books Product Information Message is the international standard for representing and communicating book industry product information in electronic form.“¹⁷²

Wie man dieser Definition entnehmen kann, handelt es sich bei ONIX for books um ein einheitliches internationales Datenaustauschformat im Bereich der Verlage. Da es auf XML basiert, kann es für den Austausch bibliographischer Daten im Rahmen des SRU verwendet werden. Sein Einsatzbereich ist ausschließlich die Titelmeldung an das Verzeichnis Lieferbarer Bücher (in Deutschland) oder ähnliche Institutionen im Ausland.

Eine Meldung an ein solches Verzeichnis geschieht in Form einer „ONIX message“, die grundsätzlich aus 2 Teilen besteht: der Informationen zum Melder und der Meldung selbst („SenderId“, „SentDate“ oder „DefaultLanguageOfText“) und der Metadaten. Dabei können mehrere Werke in einer Meldung enthalten sein. Für die Übertragung einer solchen Nachricht ist kein Protokoll festgelegt worden, so dass die meisten Meldungen per Email oder FTP gemacht werden.

ONIX besitzt zwei Arten von XML-Auszeichnern: eine lesbare und für den Menschen nachvollziehbare Version (sog. „reference names“ wie „<TextFormat>“) sowie eine für den Menschen kaum lesbare Form (sog.

¹⁷¹ EDItEUR (Hrsg.): *ONIX for Books*. URL: <http://www.editeur.org/onix.html> (Stand: 22.12.2007)

¹⁷² ebd.

„short tags“ wie: „<d103>“) wobei beide gleichwertig sind. Das VLB kann mit beiden Versionen umgehen, exportiert jedoch nur die „short tags“-Version. EDItEUR stellt für die Konvertierung ein Java-Programm bereit¹⁷³. Einige Datentypen können sowohl als Daten einer Auszeichnung als auch als deren Attribute. (Z.B. das Element „Text“ kann seine Daten zwischen „<Text>“ und „</Text>“ enthalten – in diesem Fall sind formatierende Auszeichner erlaubt – oder im Attribut „Text“ bzw. „d104“)

ONIX ist sehr detailreich. Es enthält z.B. spezielle Elemente, die der Auszeichnung von religiösen Texten – wie der Bibel oder dem Koran – dienen. Ebenso sind Titel, Prefixe für Titel, Titel-Abkürzungen oder Titel von Serien vorgesehen. Für Zeitschriften und Serien wurden Elemente wie „YearFirstPublished“ oder „YearOfAnnual“ geschaffen¹⁷⁴. Für Artikel und Artikelserien¹⁷⁵ ist eine Reihe von Elementen vorgesehen, die genaue Angaben zum Zeitpunkt, Umfang, Urheberrechten, Inhalt oder Relation zu anderen Werken aufnehmen können. Dabei stehen Felder für Print- (Seitenumfang, Band-ID) als auch für elektronische Versionen (Dateigröße, Version) zur Verfügung. ONIX for Books unterscheidet in der Beschreibung seit Version 1.1 vom März 2007 zwischen einem Exemplar und einem Werk (nach FRBR).

Mit steigender Zahl an Elementen für den Einsatz in spezifischen Bereichen wird zwar die bibliographische Beschreibung genauer, jedoch verliert man um so schneller den Überblick. Es verwundert daher nicht, dass die Titelmeldungen in diesem Format nur von Programmen mit sehr umfangreichen Handbüchern (wie z.B. Anko Publishing Manager) erstellt werden können. Dennoch kommt es in einigen Fällen vor, dass die

173 Zu finden unter: <http://www.editeur.org/ONIX%20tagname%20converter%20v2.html>
(Stand: 22.12.2007)

174 Vgl. EDItEUR (Hrsg.): ONIX for Books XML-Schema. URL:
http://www.editeur.org/onix/2.1/reference/ONIX_BookProduct_Release2.1_reference.xsd
(Stand: 20.01.2010)

175 Angaben zu Artikel-Serien nach http://www.editeur.org/files/ONIX%20-%20DOI%20Serial_Article_v1.1.pdf
(Stand: 01.04.2010)

gemachten Angaben in einer nicht validen XML-Datei resultieren. Das ONIX-Schema (definiert in XML-Schema) wäre zu restriktiv – hätte man nur XML-Schema-Datentypen verwendet. Daher wird meistens der Typ „NonEmptyString“ (nicht leere Zeichenkette) verwendet, auch wenn dies an einigen Stellen zu Problemen führt. So ist es möglich, unter „EditionNumber“ „2., erweiterte Auflage“ anzugeben, obwohl an dieser Stelle ein Zahl erwartet wird und weitere Angaben unter „EditionStatement“ gemacht werden sollten. Gleiches trifft auf Datums- und Preisangaben zu.

In einem der Problemfelder – dem, der genauen Erfassung von Personennamen des Autors bzw. des Herausgebers – erfährt der Benutzer keine Hilfe. Diese Namen werden in zwei Kategorien verwendet: beim „Contributor“ und beim Besitzer der Rechte am Werk. „Contributor“ ist für die automatische Generierung von Quellenangaben ungeeignet, weil es nicht zwischen dem Autor, Herausgeber und Übersetzer unterscheidet. Die Personennamen selbst werden entweder in der üblichen Reihenfolge „Vorname Nachname“ geschrieben („PersonName“) oder in umgekehrter Reihenfolge: „Nachname, Vorname“ („PersonNameInverted“).

Das Fehlen einer Verlinkung zwischen einzelnen Daten sowie die schwache Ausprägung von Elementen für die Erfassung von beteiligten Personen machen dieses Format für die Verwendung in Bereich der privaten Bibliographien oder Generierung von Quellenangaben nur bedingt geeignet. Wird die um DOI (Digital Object Identifier)¹⁷⁶ erweiterte Version von ONIX

176 „The DOI System is for identifying content objects in the digital environment. DOI® names are assigned to any entity for use on digital networks. They are used to provide current information, including where they (or information about them) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI name will not change.“ nach The International DOI Foundation (Hrsg.): *The Digital Object Identifier System*. URL: <http://www.doi.org/> (Stand: 20.01.2011). Digital Object Identifier entstand mit dem Ziel, digitale Objekte so zu identifizieren, dass sie immer anhand dieser Art von URI gefunden werden können, auch wenn sich ihre URL verändert hat. Realisiert wird es durch die Vergabe von eindeutigen Identifikationsnummern, die mit Hilfe eines Servers (Proxy oder Resolver genannt; z. B.: <http://dx.doi.org/>) auf URL aufgelöst werden können.

verwendet (ONIX DOI Metadaten Schema), die von mEDRA¹⁷⁷ entwickelt wurde, können einzelne Werke, die Teil einer Serie, Kapitel einer Monographie oder Fortsetzungswerke darstellen, anhand ihrer DOI in Zusammenhang gebracht werden. In diesem Format können auch komplizierte Relationen, wie unterschiedliche sprachliche Versionen desselben Werkes, Ersetzungen, Fortsetzungen und Ähnliches abgebildet werden. Die Verwendung von DOI ermöglicht auch indirekt eine Verlinkung zum Volltext (über DOI-Resolver). Der Einsatz von Unicode ermöglicht die originale Schreibweise von Namen, die jedoch mangels entsprechender Felder oder Verlinkung wohl ungenutzt bleiben wird.

4.2.2. Weitere Ansätze

Neben ZING – der bedeutendsten Weiterentwicklung im bibliographischen Bereich – gibt es viele kleine Neuerungen, die die Vorteile von XML nutzen wollen. Bis auf Zotero (mit COinS) und citeproc handelt es um eine Umsetzung alter Formate auf XML. Bei Zotero und citeproc kann man von Programmen sprechen, da sie nicht bloß der Abbildung von Metadaten (Datenformat) dienen, sondern diese Daten auch verarbeiten. Sie sind Erweiterungen ("Plug-Ins" oder "Add-Ons") zu anderen Programmen, mit welchen sie zusammenwirken.

Der Einsatz von XML bringt vielfältige Vorteile:

- Die Möglichkeit einer guten Strukturierung und Typisierung von Metadaten.
- Der Einsatz eines XML-Schemas ermöglicht eine automatisierte Überprüfung der Gültigkeit einer Datei.
- Die Akzeptanz von XML und die dafür vorgesehenen Techniken und Funktionen: sie sind für jedes Betriebssystem verfügbar – so kann die Entwicklung von Software beschleunigt werden.

In den meisten Fällen führt eine bloße Umstellung auf die Auszeichnung der Daten mit Hilfe von XML nicht zu höherer Akzeptanz. Einige Formate wie

¹⁷⁷ mEDRA ist die multilinguale DOI Registrationsagentur für Europa, zu welcher auch MVB Marketing- und Verlagsservice des Buchhandels GmbH – ein Wirtschaftsunternehmen des Börsenvereins des Deutschen Buchhandels e.V. gehört.

BibTeXXML haben in der Praxis keine Bedeutung. Die interessantesten Weiter- und Neuentwicklungen sollen hier kurz skizziert werden.

4.2.2.1. BibTeX-XML (BibTeXXML)

Das BibTeX ist ein altes und in den Formal- und Naturwissenschaften sehr verbreitetes Format. Seine Klarheit und Einfachheit, die nur auf das Zitieren wissenschaftlicher Literatur ausgelegt sind, werden vor allem in den Formalwissenschaften sehr geschätzt. Die Macro-Sprache, mit welcher diese Daten verarbeitet werden, setzt Programmierkenntnisse und eine längere Einarbeitungszeit voraus, so dass nicht jeder Nutzer eigene "Stile-Vorlagen" erstellen kann. Einige Einschränkungen in der Erfassung sind ebenfalls nicht mehr zeitgemäß. Alle diese Aspekte hatten die Beteiligten des Projektes der Eidgenössischen Technischen Hochschule in Zürich vor Augen, als sie das Datenformat auf XML-Basis portiert haben. Die dabei entstandene Diplomarbeit¹⁷⁸ „Erstellung von Bibliographien auf der Basis von XML und XSL“ von Heike Müller zeigt, wie man diese XML-konforme Daten mit Hilfe von XSLT (XSL-Transformations) umformatieren kann. Da Veröffentlichungstypen beibehalten werden sollten, ist das BibTeX-XML (oder „BIBTEXML“) grundsätzlich eine XML-basierte Abbildung des BibTeX-Formats. Beibehalten wurden neben den Datentypen auch das Prinzip der Trennung zwischen dem Dokument, den bibliographischen Daten und den Stile-Vorlagen. Anders als im BibTeX-Format wurde eine erweiterte Abbildung von BibTeX-Datenstruktur eingesetzt. Darin werden die Autoren einzeln angegeben, wobei ihre Namen in Vor-, Zweit- und Nachnamen (sowie Suffix) aufgeteilt werden. Dies trägt der Tatsache Rechnung, dass eine automatische Konvertierung von Personennamen in BibTeX oft zu einer fehlerhaften Darstellung führte. Die Typisierung von Einträgen geschieht in Form des Attributes „type“ im Element „bibitem“ - dem Hauptelement der Bibliographie.

178 Müller, Heike: *Erstellung von Bibliographien auf Basis von XML und XSLT*. Diplomarbeit. Fachhochschule Stuttgart (Hochschule der Medien). 2001

Die Neuerung liegt im Umstieg von der komplexen Macro-Sprache auf XSLT¹⁷⁹. „XSLT ist eine XML-Anwendung, die Regeln festlegt, anhand derer ein XML-Dokument in ein anderes XML-Dokument transformiert wird. Ein XSLT-Dokument – das heißt ein XSLT-Stylesheet – enthält Template-Regeln. Jede Template-Regel besteht aus einem Muster und einer Regel.“¹⁸⁰ Um diese Template-Regeln anzuwenden muss ein Programm – XSLT-Prozessor genannt – diese Muster der XSLT-Datei mit dem zu verarbeitendem XML-Dokument vergleichen. Trifft ein Muster zu, wird die Regel auf das Element angewandt. Sind alle Muster abgearbeitet, werden die Änderungen in ein neues XML-Dokument geschrieben. So kann man mit wenigen Muster-Regel-Sätzen die BibTeX-XML-Datei in XHTML (im Gegensatz zu HTML ist XHTML ein valides XML-Dokument) umwandeln. Mit XSLT kann man die Reihenfolge ändern, Zeichen hinzufügen, XML-Elemente ersetzen oder Formatierungszeichen um einen Text einsetzen. XSL ist ein verbreitetes internationales Format. Diese Tatsache erlaubt vielen Menschen "Stile-Vorlagen" zu entwickeln. Diese können z.B. mit Hilfe von modernen Web-Browsern verarbeitet werden, so dass die Ausgangsdaten für eine Verwendung im Web keiner weiteren Formatierung bedürfen.

Diese Realisierung konnte sich in der Praxis nicht durchsetzen. Die wenigen Vorteile, die sich aus dem Umstieg ergeben hätten, wie bessere Namensformatierung oder Verarbeitung zu XHTML stehen in keiner akzeptablen Relation zum Aufwand, den man hätte betreiben müssen, um BibTeX-XML-Dateien zu erstellen. Die Idee selbst, XML und XSLT für die Generierung von Bibliographien zu nutzen, findet bis heute in vielen Fällen Verwendung.

4.2.2.2. CiteProc

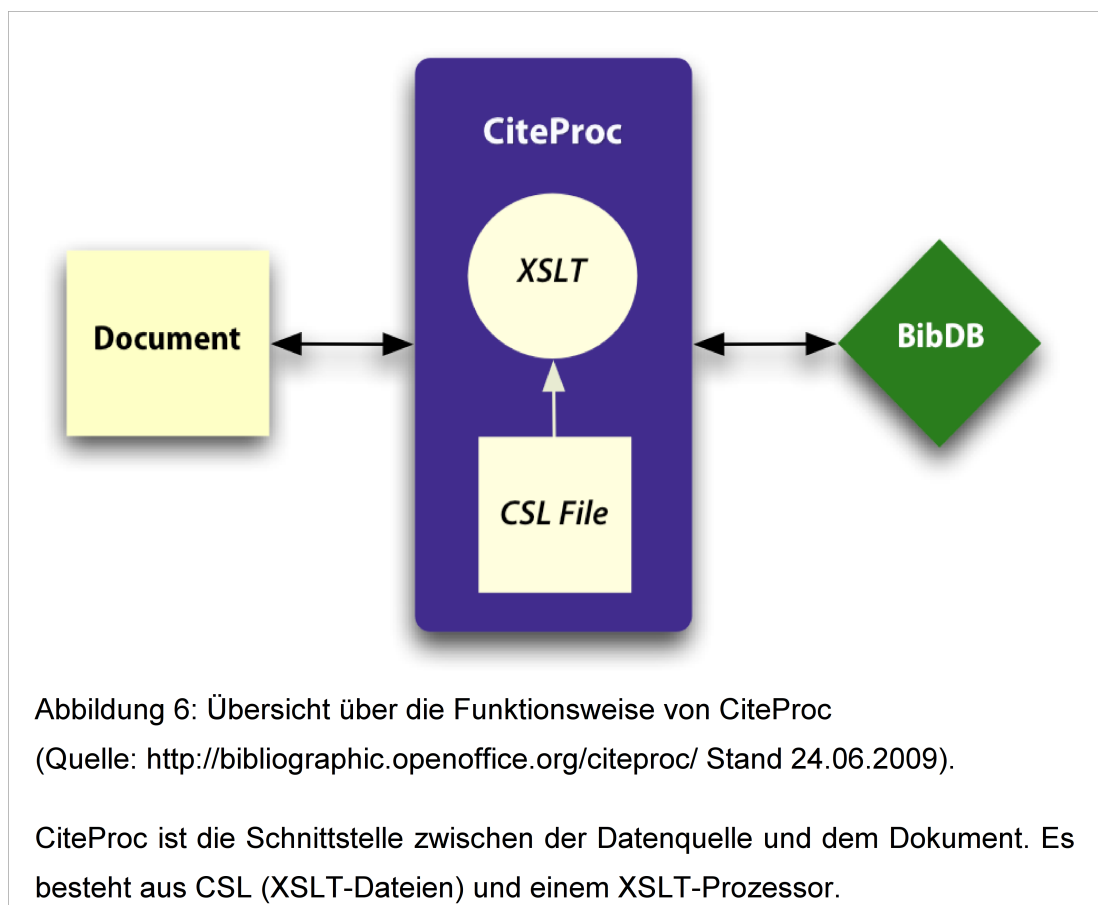
CiteProc ist keine eigenständige Software wie Reference Manage oder EndNote. Es ist eine Erweiterung für OpenOffice wie BibTeX für LaTeX.

179 XSL ist eine Abkürzung für Extensible Stylesheet Language, die aus zwei Teilen besteht: XSL Transformations (XSLT) und XSL Formatting Objects (XSL-FO).

180 XML in a nutshell. S. 146

Auch das Prinzip ähnelt dem von BibTeX-XML: die Eingangsdaten im MODS-Format werden mit Hilfe eines XSLT-Prozessors und einer Vorlage (CSL – Abkürzung für „citation style language“) verarbeitet und in das aktuelle Dokument eingefügt.

„The data store can either be a flat XML file, or a server that supports HTTP-based XQuery or SRU queries. SRU is a particularly promising new RESTful protocol that comes out of the library world, and which can provide a nice — easily implemented—standard around which a diversity of bibliographic solutions can interoperate.“¹⁸¹

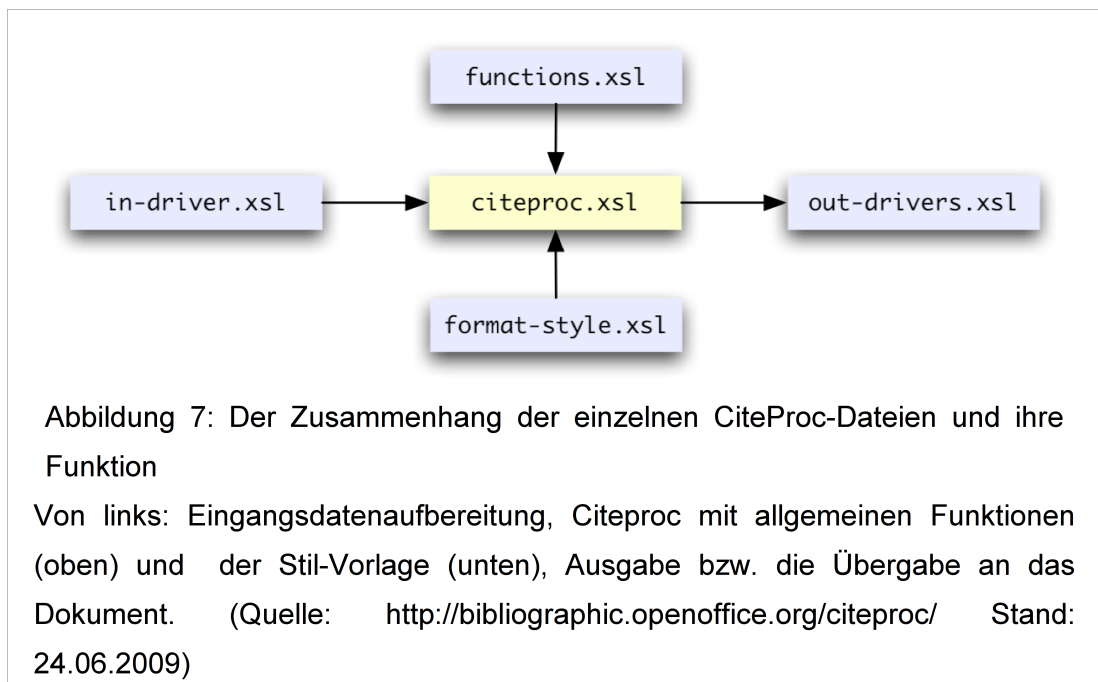


Die zu verarbeitenden Eingangsdaten können entweder als MODS-Datei vorliegen oder es findet eine Abfrage eines SRU-Servers statt. Die Abfrage

¹⁸¹ OpenOffice.org (Hrsg.) : *Bibliographic: CiteProc - Main Page*. URL: <http://bibliographic.openoffice.org/citeproc/> (Stand: 31.12.2007)

bzw. die Angabe der Datei und die Auswahl der Werke soll in Form eines Formulars realisiert werden¹⁸². Dafür sind die sog. Input-Treiber zuständig. Für einzelne Schritte der Formatierung wie das Ändern der Reihenfolge der Autoren ist die XSL-Funktionen-Datei vorgesehen. Die Reihenfolge der Elemente und deren Formatierung ist in der eigentlichen Stil-Vorlage enthalten, die ebenfalls in Form einer XSL-Datei realisiert ist.

CiteProc ist Open-Source-Software und basiert auf weiteren Open-Source-Programmen wie dem XSLT- und Xpath-Verarbeitungsprogramm (sog. Prozessor) Saxon oder (der alternativen Speicheremethode für MODS-Daten) der XML-Datenbank eXist¹⁸³.



Mit CiteProc entsteht ein eigenes „Programm“ für die automatische Zitation von Quellen in OpenOffice. Es verzichtet auf die Nutzung der OpenOffice-eigenen bibliographischen Datenbank und kann auf die Daten des SRU-Server oder einer anderen Speicherungsform wie MODS-Dateien oder XML-

¹⁸² CiteProc befindet sich im OpenOffice immer noch in einem frühen Entwicklungsstadium, so dass man nur von Prozeduren – nicht jedoch von konkreten Abläufen sprechen kann.

¹⁸³ URL: <http://exist-db.org/> Stand: 31.12.2007

Datenbank zugreifen. Der Vorteil gegenüber der bisherigen Lösung ist die direkte Übernahme bibliographischer Daten, die Nutzung üblicher internationaler Standards und somit ein enormes Potential für die Entwicklung eigener Zitier-Vorlagen. Wird keine Datenbank bzw. Datei als Datenbasis benutzt, werden lediglich einzelne zitierte Werke in OpenOffice-Datei aufbewahrt. Es ist vorstellbar, dass diese Daten aus einzelnen Dateien exportiert werden können und somit der Austausch bibliographischer Daten (statt einer formatierten Literaturliste in Textform) zwischen den Nutzern einer solchen Datei möglich sein wird.

4.2.2.3. COinS und Zotero

Zotero ist eine Erweiterung für den Open-Source-Internet-Browser Mozilla Firefox. Der Grund für die Entstehung ist in der zunehmenden Zitation von Literatur im Internet zu suchen. Diese kann vom Surfer in seine Sammlung – den Lesezeichen für Webseiten ähnlich – aufgenommen und daraus zitiert werden. Diese Sammlung kann beim Verfassen von Texten (mit Hilfe von Add-Ons für OpenOffice oder Microsofts Word) als persönliche Bibliographie benutzt werden. Diese Art des Umgangs mit bibliographischen Daten im Web ist nur dann möglich, wenn eine HTML-Seite diese als solche auszeichnet. Dies geschieht mit Hilfe von COinS¹⁸⁴. Es ist kein international anerkannter Standard und baut auf Microformat OpenURL¹⁸⁵ auf. In ein SPAN-Element der HTML-Datei werden zwei Attribute eingefügt, die geeignete Software (wie z. B. Zotero) erkennen kann. Es sind: das Attribut „class“, welches immer „Z3988“ (OpenURL NISO-Nummer) zum Inhalt hat und das Attribut „title“. In diesem Attribut werden die notwendigen Daten abgespeichert: „ctx_ver=Z39.88-2004“ - verweist zunächst auf OpenURL, „referrer_id“ ist nicht notwendig und besagt nur, wer diese Referenz generiert hat (auch die Eingabe von DOI oder URL möglich); „rft_val_fmt“ definiert, ob es sich beim zitierten Werk um ein Buch („info:ofi/fmt:kev:mtx:book“) oder eine Zeitschrift

184 Abkürzung für Context Objects in Spans.

185 Nordamerikanischer Standard für Kontext-abhängige Dienste. (ANSI/NISO Z39.88 -2004, The OpenURL Framework for Context-Sensitive Services). URL: http://www.niso.org/standards/standard_detail.cfm?std_id=783 Stand: 31.12.2007

(„info:ofi/fmt:kev:mtx:journal“) handelt. Erst „rft.genre“, „rft.issn“, „rft.aufirst“, „rft.aulast“ und weitere Kürzel beinhalten die vollständigen Daten des Werkes. Alle Daten müssen URL-konform (RFC 3986) abgespeichert werden („&“ wird durch „&“, Leerzeichen mit „%20“ ersetzt).

Nach Meinung des Verfassers ist COinS einfach zu verwenden, jedoch wenig komfortabel in der Erstellung. Auf der Seite des Projektes befindet sich ein COinS-Generator¹⁸⁶, der aus den Angaben des Benutzers entsprechende HTML-Span-Elemente generiert. Vorgesehen sind nur wenige Typen von bibliographischen Quellen: Bücher, Zeitschriften, Patente, Dissertationen und „Sonstige“. COinS ist darauf ausgelegt, mit DOI, PubMed-Identifikator und ähnlichen weltweit eindeutigen Kürzeln zu arbeiten. Es ist ebenfalls möglich, eine URL des zitierten Dokumentes anzugeben. Diese Vorgehensweise soll die Zugänglichkeit zum vollständigen Text des zitierten Werkes verbessern.

Obwohl diese Konvention kein internationaler Standard ist, wird es von Google Scholar (der Suchmaschine für wissenschaftliche Literatur), dem OCLC, der LoC, der Wikipedia aber auch vom deutschen Vascoda-Projekt unterstützt. Die Nutzung und Unterstützung von COinS kann auf zwei Wegen geschehen: als Einbindung von COinS in eine HTML-Seite (von Autoren oder von OPAC) oder als sog. Linkresolver für die Auflösung der OpenURL zu einem konkreten Medium (DOI-Auflösung oder „Find in a Library“ für eine Reservierung eines Exemplars in der Bibliothek).

Die bibliographischen Daten, die in COinS gespeichert werden können, sind im Vergleich zu den zuvor erwähnten Formaten rudimentär. Interessant und neuartig ist der Weg des Umgangs mit diesen Daten im WWW. Die Grenzen zwischen den Print- und den elektronischen Publikationen ist fließend. Eine Recherche im Web gehört zunehmend zur üblichen Vorgehensweise in der Wissenschaft (vor allem in der Informatik und den Naturwissenschaften)¹⁸⁷.

186 Zu finden unter <http://generator.ocoins.info/> (Stand: 31.12.2007)

187 Vgl. Havemann, Frank; Kaufmann, Andrea: *Der Wandel des Benutzerverhaltens in Zeiten des Internet – Ergebnisse von Befragungen an 13 Bibliotheken*. S. 86. In: Festschrift für Walter Umstätter (2006), S. 65-89.

An Stelle einer Trennung zwischen Online- und Print-Medien entsteht eine "Brücke von WWW in die Bibliothek".

Der Vorteil für den Nutzer von COinS ist die Möglichkeit, mit Plugins wie Zotero, eine eigene Literaturliste zusammenzustellen, sie in eigenen Publikationen zu nutzen und bei Bedarf diese Werke direkt in der Bibliothek vorzumerken. Zotero hilft dem Surfer auch die Internetseiten zu katalogisieren und zu Erfassen (wenn keine DublinCore-Daten vorhanden sind). Gleiches gilt für andere Werke, die der Benutzer in einer Eigabemaske des Plugins erfassen kann.

Mit COinS und Zotero eröffnet sich ein neuer Weg für bibliographische Daten: die Internetseiten werden wie Zeitschriften oder Bücher behandelt, die Bücher fast genauso einfach „abrufbar“ wie Internetseiten. Die Zotero-Plugins für die Textverarbeitungsprogramme zeigen noch einige Schwächen (wie z.B. eine sehr kleine Auswahl vorgegebener "Zitier-Stile") oder einige Fehler in der Verarbeitung der rudimentären Eingangsdaten. Export-Funktionen für viele verschiedene Formate wie MODS, RDF, RIS, BibTeX oder DublinCore können für eine Integration dieser Daten in anderen – auf das Zitieren ausgelegten – Programmen sorgen.

4.3. Vorläufiges Fazit

Der Umstieg auf Unicode, XML und andere neue Standards wie DOI bringen einen Fortschritt. Die XML-basierten Formate sind besser lesbar, können die Informationen besser organisieren, indem sie diese unterteilen oder miteinander vernetzen und sind jederzeit (auch von Dritten) erweiterbar. Reicht der sprachliche Umfang eines Formats nicht aus, kann man ihn mit Hilfe eines eigenen Namensraums erweitern. Dabei können bestehende Elemente mit neuen Elementen desselben Namens (aber mit anderen Attributen oder Kinder-Elementen) ersetzt werden. Dieser Dialekt wird für alle Programme, die dieses Format beherrschen, weiterhin lesbar. Solche unbekannteren Erweiterungen werden während der Verarbeitung ausgelassen.

Das im XML als Voreinstellung verwendete UTF-8 (ein Teil des Unicode) kann die meisten Zeichen der lebendigen, natürlichen Sprachen abbilden. Damit ist eine doppelte (die lateinisierte sowie die originale) Schreibweise möglich. Diese Möglichkeit wird bis heute in keinem Format benutzt.

Mit DOI und URI gibt es zwei weltweit akzeptierte Standards für die Vergabe von Identifikatoren, die sowohl gedruckte Werke, als auch elektronische Medien erfassen können. Diesen Vorteil nutzt COinS, um die Print-Medien im WWW erfassbar zu machen. mEDRA hat das ONIX um einen eigenen Dialekt erweitert, der unter Nutzung von DOI die Abbildung von Relationen zwischen unterschiedlichen Werken realisieren kann. Auch hier sind eine Internetseite und ein Zeitschriftenartikel in der Erfassung einander ähnlich. Dies trifft nicht auf jedes XML-basierte bibliographische Format zu. MARCXML und MABXML sind nicht darauf ausgelegt. Sie sollen das in die Jahre gekommene Format in die Welt des XML hinüber retten. MODS – das Format für MARC-Daten – ist mit dem Blick auf die Bedürfnisse der Bibliotheksnutzer entstanden. Es ist relativ einfach aufgebaut und auf MARC-Daten ausgerichtet. Untrennbare Namen wie „Gallus Anonymus“ werden von den meisten Programmen wie übliche europäische Namen verarbeitet, Namensbestandteile nicht als solche gekennzeichnet. Personennamen können untereinander nicht verknüpft werden. ONIX hingegen ist sehr umfangreich und komplex. Einige Fehlerkorrekturen (genaue Definition des Typs der Inhalte in einigen Feldern aber auch die Validierung der erstellten Dateien) und die Erweiterungen von mEDRA könnten ONIX zur hohen Akzeptanz verhelfen. Die einzige Schwachstelle des ONIX bleibt die Erfassung der Personennamen.

Keines der neuen Formate konnte die Redundanz der wiederholbaren Daten wie Personennamen, Verlage oder Organisationen relevant senken. Die meisten Formate blieben ihrem ursprünglichen Nutzungskontext treu und bilden die Datentypen nach alten Schemata ab.

Das ZING (SRU/W) genießt jetzt schon einen sehr guten Ruf. Die Umstellung auf verbreitete und weltweit anerkannte Standards zeigt ihre

Wirkung. Programme, die diesen Dienst nutzen sollen, können relativ schnell erstellt werden. Das alte Prinzip der Bibliothek als Datenlieferanten blieb erhalten (unidirektionales Protokoll). Es ist eine einfache Suche – nicht mehr. Diese Tatsache schränkt die Verwendung dieser Entwicklung im "Web 2.0"-Umfeld stark ein. Der Anspruch auf Vollständigkeit der Daten (wie Relationen zu den im Zusammenhang stehenden Werken) ist in der Planung wohl nicht berücksichtigt worden. Die Entwicklung eines „OPAC 2.0“ ist damit nicht realisierbar. Der Benutzer muss Schritt für Schritt mit vielen Anfragen an den Server seine eigene Bibliographie an einem anderen Ort zusammenstellen und ordnen. Die Übernahme von einigen interessanten Metadaten bleibt ihm wie zu Z39.50-Zeiten unmöglich.

Der Ist-Zustand und der vorsichtige Blick in die Zukunft zeigen es deutlich. Es ist noch nicht alles erreicht, was in dieser Arbeit als Ziel gesetzt worden ist: die Realisierung eines Informationsmanagementsystems für Literaturquellen in wissenschaftlichen Bereich, welches eine einfache, präzise und redundanzfreie Suche ermöglicht, die Zusammenhänge zwischen Publikationen herstellt, originale Schreibweise der Namen erlaubt und zugleich die persönliche Bibliographie so speichert, dass sie zu DIN 1505-2 konformen Quellenangaben verarbeitet werden kann.

Im Folgenden soll eine Zusammenstellung der wichtigsten Eigenschaften von BibTeXML, MARCXML, MODS und ONIX for books Auskunft darüber geben, wo die Stärken und die Schwächen liegen sowie in welchem Kontext sie ursprünglich beheimatet sind.

	BibTeXML	MARCXML	MODS	ONIX for books
Völlige Neuentwicklung	nein	nein	teilweise: an vorhandenen MARC-Daten ausgerichtet	ja
Nutzt DOI / URI	nein	nein	ja: im Identifier-Feld + Typ-Attribut	ja: im DOI-Feld
Zeichensatz	beliebig	beliebig	beliebig	beliebig
Angaben zur Sprache (außer XML-lang-Attribut)	nein; (lang-Attribut unbenutzt)	im Text eines Elementes (kein eigenes Feld)	ja: in "language" (RFC-3066, ISO-639 oder RFC-4646)	ja: LanguageOfText, OriginalLanguage
Für Bibliotheken entwickelt	nein	ja	ja	nein
Verbreitung	keine (Entwurf)	marginal	gut (wachsend)	nur im Verlagswesen
Einsatzgebiet	Persönliches Literaturmanagement	Datenaustausch zwischen Bibliotheken	Datenaustausch (Benutzer)	Titelmeldung (Verlagswesen)
Serien möglich?	ja (nur Artikel)	ja	ja	ja, als MainSeriesRecord
Serien und Serienteile in Beziehung zueinander?	ja, indirekt: collection / incollection; proceeding / inproceeding; book / inbook	ja	indirekt: nur über Titel	ja, über: SubSeriesRecord, ParentIdentifier
Teile von Personennamen ausgewiesen	ja	nein	ja: Vor- und Zunamen	nein
Original - Übersetzung in Beziehung	nein	ja	nein	nein
Qualität der Quellenang.	sehr gut	gut	gut	gut (theoretisch)

5. Entwurf einer bibliographischen Auszeichnungssprache

Die bisher definierten Formate haben den Wunsch nach einer vom Computer generierten Bibliographie oder einem zufrieden stellenden Retrieval nicht erfüllen können. Um die wissenschaftliche Arbeit zu erleichtern, braucht man unter anderem genauere Erfassung der bibliographischen Daten, einen internationalen Zeichensatz und einen Mechanismus, der Verbindungen zwischen einzelnen Teilen herstellt. Eine weitere Herausforderung für ein neues Datenformat der bibliographischen Daten sind die unselbständig erschienenen Werke (Artikel), die vor allem im Bereich der Naturwissenschaften einen immer höheren Impact-Faktor haben, aber von den Bibliotheken bisher kaum erfasst werden. Die Möglichkeit eines direkten Zitierens aus den Z39.50- oder SRU-Servern ist dadurch nicht realisierbar. Viele wissenschaftliche Artikel werden im Open-Access-Verfahren dem wissenschaftlichen Publikum zur Verfügung gestellt. Der vollständige Inhalt solcher Artikel ist unter einer DOI bzw. URN sowie einer URL einzusehen. Eine andere Form der Publikation von Artikeln ist das Ablegen von HTML- oder PDF-Dateien auf den Internetseiten der Autoren. Die Angabe solcher Quellen wird umso komplizierter.

Sowohl die International Organization for Standardization als auch deren Mitglied – das Deutsche Institut für Normung – sehen angesichts der steigenden Menge an wissenschaftlichen Quellen im Internet Handlungsbedarf im Bereich der Normung von Quellenangaben. Die ISO 690 sowie DIN 1505 sollen deshalb überarbeitet werden. Dieses Vorhaben gestaltet sich schwierig, da diese Quellenangaben oft keine permanenten Links sind und an angegebener Stelle temporär angezeigt werden oder deren Zuordnung zu einem bestimmten Publikationstyp Probleme bereitet¹⁸⁸.

Diese Fragen sollen in den Entwurf einer bibliographischen Auszeichnungssprache einfließen. Die Vereinfachung der Erfassung bibliographischer Daten, verlangt ein neues und unorthodoxes Vorgehen. Die so erfassten Daten sollen so weit wie möglich gemäß der DIN-Norm 1505-2 und 1505-3 verwendet werden können.

5.1. Anforderungen

Eine Anforderung an das neue Format ist die Einfachheit der Erfassung. Eine fehlerfreie Erfassung kann nicht einmal von Spezialisten – den Bibliothekaren – garantiert werden¹⁸⁹. Mehrdeutigkeit der Datenfelder oder zu hohe Anzahl möglicher Publikationstypen erfordern eine Deutung. Dadurch sinkt die Genauigkeit und Zuverlässigkeit der Daten. Eine kleinere Menge solcher

188 ISO 690 Teil 2 („ISO 690-2“ genannt) macht einige Vorgaben zur Zitation von elektronisch erschienenen Werken. Diese werden ähnlich den Print-Ausgaben behandelt und meistens einem dieser Typen gleich behandelt (Monographien = Datenbanken und Computerprogramme; Teile davon; Serien und Artikel; Elektronische Nachrichtensysteme = Foren, Mailinglisten, E-Mails). Der Besitzer der Internet-Domäne (wie „iso.org“) werden hierbei wie ein Verlag behandelt, obwohl sie meistens die Rolle des körperschaftlichen Urhebers einnehmen und fast immer ähnlich dem Herausgeber zu behandeln wären. (Die juristische Verantwortlichkeit für die auf dem Server publizierten Inhalte tragen die Webseitenbetreiber ähnlich den Herausgebern und Verlagen.) Aus diesem Grund werden in dieser Arbeit die Besitzer der Internet-Domäne als Herausgeber behandelt.

189 Vgl. Auszug Nr. 11 auf Seite 86

Datentypen ist sowohl während der Erfassung als auch bei der Verarbeitung zu Quellenangaben von Vorteil.

In der globalisierten Welt der Wissenschaft ist es kaum nachvollziehbar, warum nur eine – in einem bestimmten Land geltende – Schreibweise eines Personennamens zum Treffer im elektronischen Katalog führt. Die Information über weitere Schreibweisen wird nicht mitgeliefert¹⁹⁰. Die Personennamendatei der Deutschen Nationalbibliothek oder das Name Authority File der LoC sind das beste Beispiel dafür, dass die Verwendung verschiedener Namen (Schreibweisen oder unterschiedlicher Künstlernamen) einer Person kein Problem darstellen muss.

Eine bibliographische Auszeichnungssprache soll maschinell fehlerfrei zu verarbeiten sein. Sie soll so fein gegliedert sein wie nötig und gleichzeitig so einfach wie möglich. Atomare Datenstrukturen sind nicht nur die beste Grundlage für die Generierung von Literaturangaben, sondern auch für das Retrieval. Das XML ist hier das Mittel der Wahl: es kann die Daten gut strukturieren und beschreiben, es ist mit standardisierten Techniken wie XSL gut zu verarbeiten und – was für die Zukunft nicht unerheblich ist – beliebig erweiterbar. Es kann sogar in anderen XML-Dateien „eingebettet“ werden.

Die maschinelle Verarbeitung meint nicht nur die Zitierung oder eine Suche nach einzelnen Datenfeldern. Auch ein „semantisches Netz von Veröffentlichungen“ wäre denkbar. Die Voraussetzung dafür ist, dass nicht nur die hierarchischen (syntaktischen) Beziehungen, sondern auch die vertikalen (z.B. semantischen) Beziehungen erfasst werden können. Nur dadurch wäre es möglich, das Original mit der Übersetzung oder mehrere Artikel einer Serie miteinander in Verbindung zu bringen.

Nicht unbedingt notwendig aber sehr nützlich ist nach Meinung des Autors die Verwendung anerkannter internationaler Standards. Dazu zählen: URN

190 Die originale Schreibweise des Namens eines Buchautors aus Japan ist für die meisten Europäer von geringem Nutzen. Für diejenigen jedoch, die der japanischen Sprache mächtig sind, erspart dies die Suche nach der richtigen Schreibweise und ermöglicht eine Recherche in den Original-Quellen.

(für ISBN und ISSN / ISSN-L) und DOI aber auch die DDC. Auf diese Art und Weise kann man eine weltweit eindeutige Zuordnung vornehmen. Vor allem DOI und URN können dazu genutzt werden, den Benutzer vom elektronischen Katalog zu einer konkreten Print-Ausgabe oder einer archivierten (und damit unveränderten) Version einer Internet-Ressource zu führen.

5.1.1. Einfachheit der Erfassung

Die Einfachheit (bzw. Überschaubarkeit) erwies sich in vielen Projekten als wichtiger Faktor für den Erfolg eines Vorhabens. Je mehr Menschen die Regeln verstehen, desto höher die Akzeptanz eines Regelwerkes oder Standards. Viele weitere Faktoren haben ebenfalls einen nicht unerheblichen Einfluss auf das Gelingen oder Mislingen eines Projektes. Zum Beispiel ist die Vollständigkeit im Sinne des Nicht-Auslassens wichtiger Bestandteile einer Sache oder eines Sachgebietes ebenso bedeutend, wie die Einfachheit. Deshalb soll „Einfachheit“ wie folgt verstanden werden:

Es soll jede bibliographische Einheit (mit Ausnahme von Archivalien, Normen und Gesetzen¹⁹¹) erfasst werden können. Die Vorgaben müssen so knapp und zugleich eindeutig sein, dass man sie ohne eine Ausbildung im Bibliothekswesen verstehen und ohne Zweifel anwenden kann. Die Anzahl an Regeln ist zu begrenzen. Die Datenfelder müssen so eindeutig sein, dass der Benutzer problemlos die richtigen Angaben darin erfasst.

Die Geschichte des Dublin Core zeigt, dass Einfachheit nicht in Widerspruch zum Nutzen stehen muss. Die nach DC erfassten Metadaten sind in einer Bibliothek nur begrenzt nützlich. Die Verbreitung dieses Standards führte zu seiner Verwendung als Format für die Ausgangsdaten im nestor-Projekt¹⁹²

191 Die Archivalien stellen eine sehr heterogene Gruppe dar. Die Anzahl der möglichen Datentypen ist höher als die der Print-Medien. Die gleichzeitigen Nutzung von Archivalien und sonstiger Literatur kommt in der wissenschaftlichen Praxis nur in Geschichtswissenschaften vor. Normen und Gesetze unterliegen (ähnlich wie die Bibel) eigenen Zitierregeln (Bildung von Kürzeln, Fehlend der Autoren etc.). Stellen aus diesen Werken werden in wissenschaftlichen Arbeiten genannt, nicht zitiert.

192 vgl. Schwens, Ute; Liegmann, Hans: *Langzeitarchivierung digitaler Ressourcen*. In:

der Deutschen Nationalbibliothek. DC gilt im Internet praktisch als "alternativlos" und "de facto" Standard.

Diesem Beispiel folgend soll der Benutzer der einfachen bibliographischen Auszeichnungssprache nicht zu intellektueller Höchstleistung herausgefordert werden, sondern mit dem „gesunden Menschenverstand“ die Aufgabe bewältigen können. Begriffe wie „selbständig erschienenenes Werk“ oder "technischer Report" sollen im Wortschatz des Erfassers gar nicht erst vorkommen. Statt dessen sollen allgemeine Begriffe wie „Buch“ und „Periodikum“ verwendet werden. „Loseblattausgabe“ muss keine eigene Kategorie sein: Der Erscheinungsmodus (selbständig erschienen) muss dem Typ des Inhalts nicht widersprechen (meist Artikel oder Report). Eine solche kommt eher selten vor und wird fast immer von Bibliothekaren erfasst. Von Spezialisten kann man erwarten, dass sie die Regeln beherrschen und diese Art von Print-Medien einem passenden Typ zuordnen. Ein Nichtbibliothekar wird wohl eher aus der Länge und dem Inhalt auf "einen Artikel" schließen. Diese Angabe ist weniger präzise aber beim Beschaffen fast ebenso hilfreich.

Mit gleichen Datenfeldern sollen ebenfalls die Internet-Ressourcen erfasst werden. Diese spielen in der Wissenschaft eine immer größere Rolle. Dazu zählen **Online-Artikel** (Internetseiten), **Online-Bücher** (als PDF, einzelne oder miteinander verknüpfte Dateien im Web) oder **Artikel-Serien** (ein aus mehreren Teil-Artikeln bestehender Text, wobei jeder Teil meist einen anderen Aspekt beleuchtet). Sicherlich sind hier nicht alle praktisch existierenden oder denkbaren Konstruktionen aufgelistet. Es ist sehr schwierig, den unzähligen Kombinationen jeweils einen Namen zu geben. Umso schwieriger ist es, den dazu passenden Datentyp zu erschaffen. Eine Vereinfachung auf mathematisch wirkende Gleichungen kann eine praktikable Lösung darstellen. Zum Beispiel könnte als Online-Artikel gelten, was auf nur einer Internetseite (eine URL) zu finden ist (und wie ein Print-Artikel zu behandeln); ein Online-Artikel aufgeteilt auf mehrere URLs

(URL>1) würde in einem solchen Beispiel als Serie gelten. Der Bibliothekar verbindet mit dem Wort „Serie“ meist ein Fortsetzungswerk oder ein mehrbändiges Werk. Dabei bedeutet Serie nichts anderes als die imaginäre Vereinigung mehrerer Werke, die denselben Titel tragen oder während der Publikation als ein Ganzes gesehen worden sind. Diese ursprüngliche Bedeutung ist den meisten Menschen geläufig und kann so besser eingesetzt werden.

Eine solche Vereinfachung ist mit dem Verlust einiger Informationen verbunden. Um diese zu erhalten, kann einem allgemeinem Datentyp ein optionales Merkmal mit dem exakten Datentyp hinzugefügt werden. Während der grobe Raster allgemein Verständlich ist, enthält das Zusatzmerkmal die für den erfahrenen Benutzer oder Bibliothekar relevanten Informationen.

5.1.2. Internationalität

Nur wenige Wissenschaftler in Westeuropa beherrschen die slavischen oder die orientalischen Sprachen. Deshalb – so könnte man glauben – ist die „Internationalität“ keine besonders wichtige Anforderung an einen modernen Bibliothekskatalog. Dies trifft auf viele Wissenschaften zu, die Englisch als *Lingua Franca* akzeptieren (Psychologie, Ökonomie oder Naturwissenschaften).

Außer den hier erwähnten – meist „jungen“ – Wissenschaften gibt es viele andere Disziplinen, die entweder auf fremdsprachigen Quellen aufbauen oder sogar die Fremdsprache behandeln. Slavistik oder Japanologie zählen dazu ebenso wie Kulturwissenschaften oder Geschichte. Ein elektronischer Katalog, der nur lateinische Buchstaben abbilden kann, ist für diese Wissenschaftler oft nur bedingt von Nutzen. Deshalb entstanden neben den Regeln für Transkription und Transliteration Erweiterungen für die bereits genutzten Zeichensätze. Der Bibliothekar war dadurch imstande, Buchstaben der slavischen Sprachen hinzuzufügen¹⁹³. Ein anderer Weg war, den

193 Da die erfassenden Personen entsprechende Fremdsprachen meist nicht beherrschen, fügen sie oft ähnliche Zeichen ein. In den elektronischen Katalogen in Zentral- und Osteuropa ist manchmal ein „B“ an Stelle des „ß“ anzutreffen. Die Ähnlichkeit bleibt

„eigenen Zeichensatz“ wie ISO 8859-1 in Westeuropa oder den ISO 8859-2 in Zentraleuropa einzusetzen. Damit konnte man alle Werke der eigenen Sprache richtig erfassen. Die Eingabe der Zeichen eines anderen Zeichensatzes war jedoch nicht möglich.

Um solchen Problemen schon in der Planung aus dem Weg zu gehen, muss man auf einen internationalen Standard zurückgreifen, der die meisten Zeichen abbilden kann. Dabei soll die Datenmenge weiterhin möglichst klein gehalten werden, d. h. eine bestimmte Menge an Informationen soll in dieser Kodierung so wenige Bytes beanspruchen, wie sie auch in ASCII oder ISO-8859 einnähme.

Die Internationalität drückt sich aber nicht nur in der Darstellbarkeit von nicht-lateinischen Zeichen aus. Es ist ebenso die Verknüpfung von unterschiedlichen sprachlichen Versionen desselben Werkes oder unterschiedlichen Schreibweisen desselben Personennamens. So soll es eine Möglichkeit geben, das Original mit seiner Übersetzung in Verbindung zu bringen und gleichzeitig die Sprache des Werkes zu vermerken.

5.1.3. Atomare Feldstruktur

In MAB und MARC sind die Datenfelder sehr groß gewesen und mussten in Teilfelder aufgeteilt werden. Auch solche Teilfelder wurden mit zusätzlichen Daten angereichert, so dass sie ihre Eindeutigkeit verloren haben. Für den Menschen, der diese Informationen sieht, ist es kein Problem, daraus einzelne Teile zu extrahieren. Eine Maschine kann das ohne ausreichende Hilfsmittel nicht. Die Vornamen (davon kann es in einem Namen mehr geben) erkennt ein Computer anhand der Semantik (gleichet die Namensteile mit einer Liste von Vornamen ab) oder versucht es aus der Grammatik heraus zu erkennen (Nachname – Komma – Vorname – Leerzeichen – Vorname etc). In beiden Fällen ist der Computer dem Menschen unterlegen.

Dieses Problem lässt sich mit einer sehr feinen Unterscheidung der Datentypen ausgleichen. Der Vorname, der Nachname und weitere

gewahrt und kann vom Menschen korrigiert werden. Bei der maschinellen Verarbeitung kommt es jedoch zu Fehlern.

Vornamen (sowie Suffixe oder Titel) müssen eine Kategorie für sich darstellen. Jedes Feld darf nicht mehr als eine atomare Information eines bestimmten Typs aufnehmen. Erfordert ein Datenfeld eine zusätzliche (genauere) Zuordnung, muss diese als ein eigenes Merkmal aufgenommen werden. Diese Forderung steht im Zusammenhang mit weiteren Erwartungen wie der besseren "Lesbarkeit" für Maschinen oder der Eindeutigkeit der Felder während der Erfassung.

5.1.4. Maschinelle Verarbeitung

Die Informationen sollen in der bibliographischen Auszeichnungssprache so angeordnet sein, dass sie unabhängig von der späteren Darstellung abgespeichert werden. Die Schreibweise: „Vorname Nachname“ ist schon eine Art der Darstellung. Da es mehrere Arten der Darstellung geben kann, sollen die Daten mit Hilfe von Stil-Vorlagen in jede beliebige Darstellungsart gebracht werden können. Diese Vorgehensweise hat sich bewährt. Sie ist die einzige Möglichkeit die „Datenebene“ von der „Darstellungsebene“ zu trennen¹⁹⁴. Die bisher genutzten Lösungen waren sehr spezifisch und oft anspruchsvoll (BibTeX), so dass nur wenige Menschen einen neuen Zitierstil umsetzen konnten. Angestrebt wird die Nutzung eines möglichst einfachen, allgemeinen und weit verbreiteten Formats für die Umsetzung von Stil-Vorlagen.

5.1.5. Erweiterbar nach Bedarf

MARC und MAB wurden vor 30 Jahren definiert und orientieren sich an den Bedürfnissen einer bestimmten Zeit und eines gegebenen Nutzungskontextes (Datenaustausch zwischen Bibliotheken). Die Festlegung auf eine dreistellige führende Zahl bietet 999 Datenfelder an. Davon sind je nach Format nur ca. 75% belegt. Würden mehr als 250 zusätzliche Felder für neue Datentypen benötigt, müsste man das ganze Format erneuern.

¹⁹⁴ Im gegenwärtig als 'State of the art' geltendem Entwicklungsparadigma: „Model-View-Controller“ (MVC) gibt es drei sog. Ebenen: das Datenmodell und die darin enthaltenen Daten, die Darstellung (die HTML, PDF oder ein maschinell vorgelesener Text sein kann) und die Steuerung, die entsprechend der Eingabe des Benutzers die Daten aus dem Modell in die gewünschte Darstellung überführt.

Es nicht absehbar, wie, wo und in welcher Form bibliographische Elemente in der Zukunft benötigt werden. Deshalb ist es sehr wichtig, die Datenformate von heute für die künftigen Entwicklungen offen zu halten.

Zugleich soll die Möglichkeit nicht ausgeschlossen werden, dass ein Format für bibliographische Metadaten als Basis für ein viel komplizierteres System dient. Mit dem Begriff der „Bibliothek 2.0“¹⁹⁵ kam die Idee auf, dass ein Bibliotheksnutzer seine Quellen in die Datenbank seiner Universitätsbibliothek einfließen lässt (selber erfasst), während der Bibliothekar diese Angaben vervollständigt und stärker mit dem Bestand verknüpft. In dieser Situation wird der Student vom „Buch“ sprechen und ein Werk dem Typ „Buch“ zuordnen, während die Bibliothekare diesem Werk das Attribut „Monographie“ anhängen. „Mehrbändiges Fortsetzungswerk“, „Tagungsband“, „Technischer Report“ oder „Forschungsreport“ als eigene Typen nicht benötigt, da sie als Zusatzmerkmale erfasst werden können. Eine erweiterte Version (z. B. durch Heranziehung anderer Standards und Namensräume) soll ebenfalls möglich sein.

5.1.6. Verwendung in anderen Datenformaten

Bibliographische Daten flossen bisher nur in OpenOffice und BibTeX als eigener Datentyp in die Dateien mit wissenschaftlichen Arbeiten ein. In den meisten Fällen (wie z. B. Word-Dateien) wurden die bibliographischen Daten als formatierter Text in die Dateien integriert. Im OpenOffice gab es eine Trennung zwischen Literatursammlung (binäre Datenbank) und Datei (binäre Datei). Die bibliographischen Daten verbleiben bis heute auf dem ursprünglichen Rechner und können nur selten (und oft nur über „Umwege“) auf einen anderen übertragen werden. Die zitierten Werke werden heute in ODF als ein eigener Satz von Daten abgespeichert (eingebettet). Diese sehr

195 In der öffentlichen Diskussion wird dieser Begriff unterschiedlich gedeutet: Entweder als Prozess der völligen Ausrichtung der Bibliothek an seinen Benutzern oder als völlige Öffnung der Bibliothek durch die Online-Zusammenarbeit von Bibliothekar und Benutzer. Vgl. dazu: Herb, Ulrich: *Ohne Web 2.0 keine Bibliothek 2.0*. URN: urn:nbn:de:bsz:291-scidok-12917 URL: <http://scidok.sulb.uni-saarland.de/volltexte/2007/1291/> Stand: 15.12.2008

simple Datenstruktur für bibliographische Daten lässt sich (aus technischer Sicht betrachtet) gegen jeden anderen Typ austauschen.

Die Einbettung eines informationsreichen Formates (z. B. in die ODF-Dateien) hätte den Vorteil, dass der Autor tatsächlich von einem bestimmten Arbeitsplatz (PC) unabhängig wäre. Er hätte in einer solchen Datei neben seinem Werk seine ganze Literatursammlung (samt Zusammenfassungen oder Relationen zwischen den Werken) mit dabei – egal, wo er arbeitet. Aus XHTML-Internetseiten könnte man mit einem Klick alle Daten zu zitierten Werken in die eigene Literaturdatenbank importieren.

5.1.7. Identität, Integrität und Homogenität der Daten

Eines der Probleme der bereits besprochenen Systeme war die Homogenität der Daten: jede erfassende Person hätte in gewissen Fällen einige Metadaten woanders zuordnen oder anders schreiben können. Trotz großer Regelwerke kann man nicht garantieren, dass die Daten homogen d.h. qualitativ gleichwertig bleiben. Die Lösung dieses Problems liegt zum großen Teil in der Genauigkeit der Vorgaben – z. B. in der atomaren Datenstruktur. Aber auch dieser Ansatz offenbart seine Schwächen angesichts großer Datenmengen: muss der Benutzer für jedes Werk desselben Autors seine Vornamen, seinen Nachnamen und womöglich Titel oder Suffixe erfassen, schleichen sich schnell Schreib- oder Erfassungsfehler ein. Es ist somit hilfreicher die atomaren Datenstrukturen an einem einzigen Ort einzusetzen. Dieser Ort wird von allen anderen Stellen referenziert. Auch andere Schreibweisen oder Übersetzungen können diesen Ort als Bezugspunkt nutzen. Wird dabei darauf geachtet, dass alle referenzierten Orte auch existieren, wird auch die Datenintegrität sichergestellt.

Diese Forderung ist nicht unbedeutend. Alle gegenwärtig eingesetzten elektronischen Kataloge arbeiten im Bezug auf Referenzen systemimmanent d. h. das Wissen um die Referenz-Relation von „Benedikt XVI“ zu „Joseph Ratzinger“ bleibt immer im Katalog (z. B. der PND). Dem Benutzer werden zwar die Werke von Benedikt XVI. sowie die von Joseph Ratzinger aufgelistet, die Beziehung zwischen beiden Namen derselben Person wird

dem Client vom Server aber nicht explizit genannt (wird meist vom Menschen erschlossen). Diesem Verlust von Identität soll vorgebeugt werden, damit der Server wie der Client auf demselben „Wissenstand“ sind.

5.2. Ansätze zur Realisierung

Die genannten Anforderungen sind vielfältig. Wege und Techniken, die für die Umsetzung dieser notwendig sind, sind Gegenstand dieses Kapitels (5.2). Einige von ihnen beziehen sich auf den heutigen Stand der Technik und können in wenigen Jahren als überholt gelten. Andere stellen allgemeine und abstrakte Prinzipien dar. Dabei sollen die bewährten Lösungen berücksichtigt und weiterverwendet werden. Die bibliographische Auszeichnungssprache mit seiner Gliederung gilt dabei als Mittel, ein effizientes bibliographisches Informationssystem als Ziel zu schaffen. Das Datensystem, welches auch als Format bezeichnet wird, ist an diesem in der Einleitung skizzierten Ziel ausgerichtet. Die Effizienz wird der Vollständigkeit vorgezogen.

5.2.1. Drei-Ebenen-System

Die Einfachheit ist keine Bedingung aus sich heraus. Sie dient z. B. der besseren Umsetzung in Form von Computerprogrammen. Ihr wichtigstes Ziel ist das Ermöglichen einer genauen Erfassung durch möglichst viele Menschen. Ein Begriff wie „proceedings“ oder „Konferenzband“ existiert fast ausschließlich im Wortschatz der Bibliotheken und der Wissenschaft. Für einen Nichtbibliothekar und Nichtwissenschaftler sind einige wenige Typen präsent: Ein Artikel (bzw. Kapitel), ein Buch, eine Zeitschrift und eine Serie (z. B. ein mehrbändiges Lexikon). Dies sind aber auch die grundlegenden Typen von Veröffentlichungen. Ein Buch ist ein selbständig erschienenes Werk, dessen Erscheinung nicht von der Zeit oder einem periodischen Geschehen abhängt¹⁹⁶. Ein Buch kann „aus einem Guss“ sein oder aus mehreren Artikeln eines oder mehrerer Autoren bestehen. Ein Artikel ist

¹⁹⁶ Vgl. Definition des Buches nach UNESCO. Z.B. in: Funke, Fritz: *Buchkunde : Die historische Entwicklung des Buches von der Keilschrift bis zur Gegenwart*. VMA-Verlag, Wiesbaden 2006, ISBN 3-928127-95-0.

immer ein in sich geschlossenes Werk (ein gegliederter Text) und ein Teil eines anderen selbständig erschienenen Werkes. Die zweite – sehr verbreitete – Form eines selbständig erscheinenden Werkes ist eine Zeitschrift bzw. ein Periodikum. Diese besteht fast immer aus mehreren Artikeln. Die Erscheinungsweise hängt entweder direkt von der Zeit (Tag, Woche, Monat) oder von einem periodisch stattfindendem Geschehen ab. Mehrere selbständig erschienene Werke, die aufgrund des Titels, der gleichen physikalischen Gestaltung oder eines inhaltlichen Zusammenhangs als ein „großes Ganzes“ angesehen werden bezeichnet man als Serie. (In der Philosophie wird die Existenz solcher „Wesen“ wie „Serie“ durch die sog. Nominalisten bestritten. Für sie ist dieser Begriff leer, da mehrere einzelne „Wesen“ nicht von sich aus ein anderes ergeben, sondern das menschliche Gehirn ein solches „erzeugt“.) Bestreiten wir als Realisten die tatsächliche Existenz von Serien nicht, eröffnet sich die Möglichkeit, mehrere Artikel eines Autors oder mehrerer Autoren zum selben Thema als Artikel-Serie im Sinne eines Werkes (einer Komposition) zu betrachten. Solche Artikel-Serien werden manchmal in Zeitschriften realisiert, wenn der Platz für die Behandlung eines sehr umfassenden Themas nicht ausreicht.

Für die Abbildung der bibliographischen Werke reicht grundsätzlich eine Gliederung in drei Ebenen. Die unterste Ebene stellen die Artikel dar. Sie sind die elementarsten Teile eines wissenschaftlichen Buches, eines Lexikons oder einer Zeitschrift. Sie sind sozusagen das Fundament. Ihre Zahl ist höher als die der Bücher oder Zeitschriften. Deshalb stellen sie in der Abbildung 8 den breiten Sockel dar. Auf diesem Sockel bauen die Bücher und Periodika auf. Sie sind komplexer und abstrakter als die Artikel. Ihr Inhalt wird oft von vielen Autoren verfasst. Es werden manchmal sehr unterschiedliche Themen innerhalb eines Buches oder Zeitschrift behandelt. Ist das behandelte Thema in allen ihren Teilen gleich oder miteinander verwandt, ist ein Sonderheft oder Themen-Heft einer Zeitschrift bzw. eine Monographie als eine spezielle Form des Buches anzunehmen.

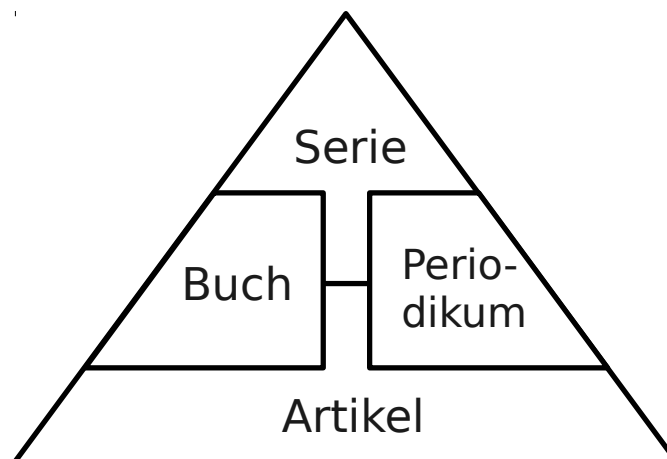


Abbildung 8: Die graphische Darstellung des Drei-Ebenen-Systems.

Eine Serie kann aus Objekten des Typs Buch, Periodikum oder Artikel bestehen. Ein Buch oder ein Periodikum kann Artikel enthalten. Artikel können einzeln existieren. Ein Periodikum kann kein Teil eines Buches sein und umgekehrt. (Die Überlappung deutet eine mögliche aber nicht notwendige Beinhaltung Objekte angrenzender Typen an. Die Eigenschaften werden von oben nach unten vererbt.)

Die Bücher und Zeitschriften sind gängige Begriffe, mit denen auch ein Nichtbibliothekar und Nichtwissenschaftler umgehen kann. Das bereits als Beispiel genannte Konferenzband ist nur wenigen bekannt. In seiner Erscheinungsform erinnert es sehr an ein Buch. Die zeitliche Abfolge von jährlich stattfindenden Konferenzen erinnert hingegen an eine Zeitschrift. Gleiches trifft auf den Titel bzw. den Untertitel der Form „Die XX. Konferenz der Gesellschaft für N.N.“ zu. Die wechselnde Zahl „XX“ ähnelt einer Zählung, wie sie bei Zeitschriften und Journalen üblich ist.

Diese Zuordnung von Tagungsbänden zu den Periodika ist sehr unorthodox. Die Kontinuität des begonnenen (und nicht in absehbarer Zeit abzuschließenden) Werkes in Form einer Serie spricht jedoch für dieses Vorgehen. Ebenso verhält es sich mit dem gleich bleibenden Herausgeber (meist die veranstaltende Gesellschaft) oder der Titel.

Bücher und Zeitschriften haben eine physikalische Erscheinung, die es ermöglichen, das Werk als ein „Etwas“ wahrzunehmen, obwohl sie nicht „aus einem Guss“ sind oder sein müssen. Die Serien sind hingegen sehr abstrakt¹⁹⁷. Im Bezug auf die bibliographischen Werke kann man den Titel eines Periodikums als eine solche wesentliche Eigenschaft aller dieser Periodika betrachten. Außer dem Titel sind viele weitere Eigenschaften mögliche Indikatoren für die Zugehörigkeit zu einer Serie: der Herausgeber, das Thema oder die Autoren. Man kann sie nicht alle aufzählen. Deshalb ist es wichtig, dass das Prinzip eines wesentlichen Zusammenhangs erhalten bleibt. Einen guten Beispiel liefert die Belletristik: mehrere nacheinander folgende Romane desselben Autors, die einen Zusammenhang wie Zeit und Ort, Personen oder Thema aufweisen, werden von den Lesern als Fortsetzung (Serie von Büchern) verstanden. Manchmal ist die grafische Gestaltung des Umschlags ein Hinweis auf eine Serie. Die Abstraktion ist ein weiches Unterscheidungskriterium. Es ist dennoch ausreichend.

Wegen der abstrakten Existenz einer Serie ist sie an der Spitze der Grafik angeordnet. Das Angrenzen an Buch, Periodikum und Artikel weist auf die Tatsache hin, dass alle drei Typen Serien bilden können. Die Artikel erscheinen als Serie, wenn sie z. B. auf mehrere Ausgaben einer Zeitschrift aufgeteilt werden. Solche Artikel-Serien werden von den meisten Informationssystemen nicht erfasst. Dabei ist es für den recherchierenden Wissenschaftler schon im Voraus interessant, ob er nur die eine Zeitschrift per Fernleihe bestellen muss, oder ob das Thema ausführlicher in mehreren Ausgaben behandelt worden ist. (In Dublin Core Terms kann dies durch das Element „isRequiredBy“ realisiert werden: „isRequiredBy“ soll die Kohärenz zwischen zwei Werken wahren, indem das Werk A das Werk B voraussetzt. Dies ist ein pragmatischer Weg ohne abstrakte Serien. Ein unerwünschter

¹⁹⁷ Abstraktion (aus dem Lateinischen „abstrahere“ = abziehen, trennen) ist ein Denkprozess, welches das Wesentliche vom Unwesentlichen trennt und das Vorkommen der wesentlichen Eigenschaften an mehreren Dingen zu einer „Wesensklasse“ zusammenfasst. Ein ähnliches Vorgehen wird bei der Generierung von Klassen einer Klassifikation verwendet. (vgl. Manecke, Hans-Jürgen: *Klassifikation, Klassieren*. In: KSS. S 125.)

Nebeneffekt wäre die fälschliche Annahme, Teil 2 einer Artikelserie ist nur dann zu verstehen, wenn man Teil 1 gelesen hat. Dies muss nicht unbedingt zutreffen.)

5.2.2. Unicode

Die internationale Gemeinschaft der Forscher und Wissenschaftler braucht Werkzeuge, die von den nationalen oder sprachlichen Grenzen nicht betroffen sind. Zu diesen zählen die ISO-Standards. Ein besonders wichtiges Werkzeug für die wissenschaftliche Tätigkeit einiger Disziplinen stellt der Zeichensatz dar. Deshalb entwickelte man verschiedene 8-Bit-Zeichensätze wie z. B. ISO-8859. Ihre Verwendung ist sehr begrenzt: man kann nicht gleichzeitig in einer Text-Datei zwei unterschiedliche ISO-8859-Zeichensätze verwenden. Um das Verfassen von Texten mit gemischten Zeichensätzen zu ermöglichen, bräuchte es einen Zeichensatz, der als international anerkannter Standard alle bisher verwendeten Zeichensätze in sich vereinigt¹⁹⁸. Ein solcher Standard ist das Unicode 4.0. Dieser enthält sogar die Zeichen exotischer oder ausgestorbener Sprachen und hält eine Million zu vergebender Code-Stellen bereit¹⁹⁹. Eine Art davon ist UCS-2 (ISO-10646-UCS-2). Die Nachteile dieses Standards sollen nicht verschwiegen werden: Es ist zu ASCII (oder ISO-8859-1) inkompatibel und die Verdopplung des Speicherbedarfs (bzw. die zu übertragene Datenmenge) bei gleichem Inhalt (bezogen auf Latin-1). UTF-16 nutzt das UCS-2 und Ersatzpaare und wird eingesetzt, um z. B. die Zeichen von Bengali oder Thai abzubilden²⁰⁰. Die höchste Akzeptanz genießt UTF-8, welches die Zeichen in einem, zwei oder vier Bytes kodiert. Es ist abwärts zu ASCII (Bytes 1-127) kompatibel und beansprucht bei der Nutzung von vorwiegend lateinischen Zeichen nicht wesentlich mehr Platz, wie eine ISO-8859 Kodierung. Deshalb soll UTF-8 als

198 *Unicode*. In: *KSS-Glossar*, S 123

199 „... aber niemand ist bereit, in der Öffentlichkeit zuzugeben, woher seiner Meinung nach die verbleibende Million Zeichen stammen werden“ aus: *XML in a nutshell*. S. 77

200 Die Möglichkeit, diese Zeichen zu erfassen, bedeutet aber nicht, dass diese auch dargestellt werden können. Fast alle Schriftarten definieren gegenwärtig nur einen kleinen Satz an Glyphen (meist Latin-1, -2 sowie griechische und kyrillische Buchstaben).

das Mindestmaß an Internationalität hinsichtlich der Technik angesehen werden.

5.2.3. XML

Für eine mögliche Umsetzung ist es nicht unerheblich, wie komplex die eingesetzten Standards sind. Im zweiten Kapitel dieser Arbeit wurde an einigen Beispielen ersichtlich, dass Ideen wie RIS oder BibTeX deshalb so oft eine vollständige Umsetzung erfahren, weil sie einfach zu realisieren sind und eine hohe Akzeptanz genießen. Obwohl das Z39.50-Protokoll als ein allgemeines Protokoll für Information Retrieval entworfen wurde, ist dies nur in der Literaturverwaltungssoftware umgesetzt worden. Die Austausch-Formate wie MAB2 oder verschiedene MARC-Dialekte wurden ebenfalls in kommerziellen Programmen implementiert. Diese Tatsache hängt nach Meinung des Autors mit der Kosten-Nutzen-Rechnung zusammen. Je mehr Standards (Protokolle, Formate etc) eine Firma umsetzen muss, desto zeitintensiver wird die Entwicklung und teurer das Produkt. Kann man hingegen kostengünstig auf die Arbeit anderer Entwickler aufbauen, spart man Zeit und Kosten.

Ein weiteres Hindernis für die freien Software-Entwickler stellen aus Sicht des Verfassers die sog. Software-Patente dar. Die Verwendung patentierter Funktionen setzt in den meisten Fällen eine kostenpflichtige Lizenz vom Patentinhaber voraus. Die Kosten einer solchen Lizenz können in Open-Source-Software mangels Einnahmen aus dem Verkauf nicht gedeckt werden. Aus diesem Grund soll nach Möglichkeit auf patentierte Lösungen verzichtet werden, zumal sie die sog. Digitale Kluft unnötig fördern²⁰¹ und nicht notwendigerweise einen deutlichen Fortschritt bedeuten (z. B. Trivialpatente). Das XML ist als solches patentfrei und somit jedermann kostenfrei zugänglich.

201 vgl. Torvalds, Linus: *Offener Brief von Linux-Entwicklern zum Thema "Software-Patente"*. E-Mail vom 21. September 2003 an den Präsidenten des Europäischen Parlaments, Herrn Pat Cox, und an die Mitglieder des Europäischen Parlaments. Archivierte Kopie zugänglich unter http://www.ffi.org/patentit/patents/_torvalds_cox.html

Seit einigen Jahren wird eXtensible Markup Language („erweiterbare Auszeichnungssprache“) als der Meilenstein auf dem Weg zur Computergerechten Aufarbeitung von Daten angesehen. „Verständlich“²⁰² sind XML-Dateien für den Computer nicht. Erst das auf XML aufbauende Resource Description Framework (RDF) bringt eine Syntax und Beschreibungslogik, die es dem Computer ermöglicht, Aussagen in Folge logischer Schlüsse zu generieren. RDF ist zusammen mit Dublin Core ein weltweit verbreiteter Standard für die Erfassung von Metadaten der im Web liegenden Ressourcen. Aufgrund der schwachen Ausprägung von DC ist diese Kombination für die Erfassung bibliographischer Werke nicht hinreichend. Die Stärke von RDF ist die gute semantische Beschreibung einer Ressource, Eigenschaft für Eigenschaft. Dieses Vorgehen ist für einen Menschen eher mühsam, so dass er auf spezielle Editoren bei der Erfassung angewiesen ist.

Das XML besitzt eine Grammatik, die aus nur wenigen Regeln besteht. Es ist vielmehr ein Grundgerüst, das in einer Realisierung den Rahmen bereitstellt, aber nichts über die darin enthaltenen Daten aussagt. Die Spezifikation nimmt ein XML-Schema vor. Es legt fest, welche Inhalte in welcher Reihenfolge und wie oft vorkommen dürfen. Ein Computer kann eine Datei auf ihre Gültigkeit überprüfen, ohne zu wissen, was die einzelnen Elemente bedeuten sollen. Solche Funktionen wie die Validitätsprüfung von XML wurden bereits in vielen sog. Funktionsbibliotheken realisiert. Jede moderne Programmiersprache beherrscht den Umgang mit XML-formatierten Daten. Mit sog. Frameworks wie Hibernate ist es sogar möglich, ohne großen Aufwand, die Daten eines Programms (die sog. Objekte) in Form von XML zu serialisieren d. h. die Objekte so persistent abzuspeichern, dass auch die

202 Man unterscheidet im Computer-Wesen zwischen "Verständlich" und "Lesbar". Gut lesbar ist ein Dokument, wenn sein Inhalt und somit die Struktur der darin beinhalteten Daten ohne größere Verluste im Speicher des verarbeitenden Programms abgelegt werden können. Gute Lesbarkeit ist die Voraussetzung für gute syntaktische Verarbeitung. "Verständlichkeit" von Daten kann mit einem einfach gehaltenem Turing-Test untersucht werden. Wenn die Antworten eines Rechners von den Antworten eines Menschen kaum zu unterscheiden sind, kann man vom "verstehen" der Inhalte sprechen.

kleinste mögliche Eigenschaft eines Objektes im XML abgebildet (und auf einem anderen Rechner wiederhergestellt) werden kann. Es sind vor allem die Eigenschaften (in XML „Attribute“ genannt), die XML so "universal" machen: bei mehreren Vornamen eines Buchautors könnte man sogar bestimmen, welcher davon der „erste“ oder der „gebräuchliche“ ist, bei Geldwerten kann man dem Element „Betrag“ die passende Währung „vorgeben“. Im Gegensatz zu RIS oder MARC kann XML die Daten (Bestandteile eines Datensatzes) sehr genau auszeichnen (z. B. „Vorname“, „Nachname“, „Suffix“ oder „Rufname“) und gleichzeitig weitere Aussagen darüber treffen (typisieren).

Das XML scheint der Realisierung der Ziele dieser Arbeit dienlich zu sein: es kann die Daten exakt bestimmten Typen zuordnen, mit Hilfe von XML-Schemata kann es auf Gültigkeit überprüft werden. Des Weiteren genießt das XML hohe Akzeptanz, ist patentfrei und in jeder Programmiersprache lesbar²⁰³.

5.2.3.1. XSLT-Stile

Um das XML entstanden weitere Standards und Funktionen. Eine für dieses Konzept besonders relevante Anwendung stellt das XSL dar. Das XSL-T beschreibt, wie die „Übersetzer“-Dateien formuliert sein muss, damit eine Übersetzung einer XML-Datei aus einem Dialekt in ein anderes funktioniert. Die „Übersetzung“ nimmt ein XSLT-Prozessor (ein Computer-Programm) vor. Auf diese Art und Weise kann man aus einer bestimmten XML-Datei eine XHTML-Datei generieren, die von jedem Web-Browser gelesen werden kann. Des Weiteren kann man Inhalte einer MODS-Datei in Elemente von

203 Auch das Vascoda-Projekt setzt auf Web Services, die den Beteiligten ermöglichen, Anfangen im XML zu senden (SOAP) und XML-Daten zu empfangen. Dabei wird für die Antwort-Datensätze ein eigenes XML-Format entworfen. Vgl. dazu: Helmes, Leni; Steidl, Nicole: *Webservices praktisch angewandt. FIZ Karlsruhe entwickelt Automatisierung der Informationsstruktur*. S. 421-428. In: Information. Wissenschaft und Praxis. 2003.

ODF (ISO-standardisierte Open-Office-Dateien) übersetzen – wie es im CiteProc-Projekt geschieht²⁰⁴.

XSLT ist sozusagen das Bindeglied zwischen den Inhalten und der Darstellung, analog zu CSS beim X-/HTML. Dieses Prinzip ist schon in BibTeX verwendet worden: die bibliographischen Daten als Datenbasis werden bei der Verarbeitung an gewünschter Stelle unter Verwendung von Vorlagen entsprechend formatiert eingefügt, ohne dass die BibTeX Dateien selbst verändert werden. Im Unterschied zur BibTeX-Macrosprache ist XSL nicht auf einen Bereich begrenzt, sondern kann bei allen XML-Dateien eingesetzt werden, was seine Popularität steigert. Dadurch erhöht sich die Anzahl derjenigen, die geeignete Vorlagen für die Verarbeitung von Daten der einfachen bibliographischen Auszeichnungssprache in z. B. ODF schreiben können. Dank einer atomaren Datenstruktur ist die Umsetzung in jede gewünschte Norm möglich. Wenn diese Umsetzung der Ausgangsdaten in eine Zitierform erst in der Phase der Darstellung geschieht und die ursprünglichen Daten unverändert bleiben, kann jeder Leser einer solchen Datei die Quellenangaben in einem eigenen (bevorzugten) Format sehen. Das kann eine bestimmte Formatierung sein, aber auch die Nutzung der unveränderten Ausgangsdaten (die bibliographische Datenbasis) ist denkbar. Der Autor stellt dem Leser seine Literatursammlung als Rohdaten zur Verfügung. Dieser kann dann mit Hilfe geeigneter Programme diese Datenbank extrahieren und zu seiner persönlichen Sammlung hinzufügen. Die Datenbasis ist somit multifunktional und kann effizienter genutzt werden.

Neben XSLT gibt es einen ähnlichen Standard namens XSL-FO, der XML-Daten in eine binäre Form übersetzen kann. Dies kann eine PDF-Datei, ein Bild oder eine beliebige Textdatei sein, wie z. B. Graphen-Beschreibung, die unter Einsatz anderer Computer-Programme eine baumartige Darstellung des Datenbestandes ermöglichen. Dieses Vorgehen ist bei der Visualisierung einer Literatursammlung denkbar, die in komplexen Strukturen Übersicht verschaffen kann²⁰⁵.

204 Vgl. <http://bibliographic.openoffice.org/citeproc/index.html> Stand: 01.06.2010

205 Vgl. Hobohm, Hans-Christoph: *Persönliche Literaturverwaltung im Umbruch. Vom*

Aus diesem Grund bewährt sich das XML (zusammen mit XSL-T/-FO) als Fundament für eine bibliographische Auszeichnungssprache.

5.2.3.2. Erweiterbar nach Bedarf

Eine einfache Auszeichnungssprache, die nur aus wenigen Elementen besteht, ist nicht besonders ausdrucksstark. Man kann mit ihr nur die nötigsten Angaben erfassen – wie es z. B. im Fall von Dublin Core sichtbar wurde. Die Einfachheit der Handhabung des DC war erwünscht und brachte ihm hohe Akzeptanz. Gleichzeitig wurde sie zur Hürde, wenn es um den professionellen Einsatz in den Bibliotheken oder Archiven ging. Deshalb sollte ein Datenformat für Metadaten flexibel (anpassbar bzw. erweiterbar) bleiben. Es muss erweiterbar sein und neuartige Publikationsformen aufnehmen können. Dies ist mit MARC oder MAB nicht realisierbar. Sobald alle 999 zur Verfügung stehende Felder mit Bedeutung belegt worden sind, ist die Aufnahmefähigkeit ausgeschöpft. Bei XML existieren derartige Begrenzungen nicht. Das XML benutzt sog. namespaces (Namensräume), die es ermöglichen, innerhalb einer Datei viele verschiedene Elemente und Attribute unterschiedlicher Herkunft parallel zu verwenden. Mit jedem Namensraum wird ein bestimmter Satz an Schema-Definitionen verknüpft. Diese Schemata liegen in Form von XMLS- oder RelaxNG-Dateien unter angegebener URL vor und können vom verarbeitendem Computerprogramm zwecks Überprüfung der Validität nachgeladen werden. Jedem verwendeten Schema wird ein Name zugeordnet. Diesen Namen schreibt man vor ein daraus verwendetes Attribut oder Element (z. B. „svg:layer“ oder „MODS:author“). Auf diesem Wege teilt man dem Computer mit, dass man Elemente aus einem bestimmten Definitionsbereich nutzt. Diese Methode kann dazu verwendet werden, einen begrenzten (ausdrucksarmen) Definitionsbereich um weitere Elemente zu ergänzen. Es wäre denkbar, dass ein Bibliothekar das Objekt „Buch“ um das Attribut „typ“ mit dem Wert „Monographie“ erweitert. Statt „<Buch><Titel>XML in a

Bibliographie-Management zum Social Bookmarking. Anmerkungen zu EndNote, Reference Manager, RefWorks und Connotea. S. 358-388. In: Information. Wissenschaft und Praxis. 7 / 2005

`nutshell</Titel></Buch>`“ könnte dieser Datensatz nach einer Korrektur durch den Bibliothekar beispielsweise so aussehen: „`<sbml:Buch bibliothekar:typ='Monographie'><sbml:Titel>XML in a nutshell</Titel></Buch>`“ wobei „*sbml*“ eine einfache bibliographische Auszeichnungssprache bedeutet und „*bibliothekar*“ ein erweiterndes Schema. Dies wäre mit MAB oder MARC nicht möglich.

5.2.3.3. Einbindung in andere Dateien

Man kann die bibliographischen Daten innerhalb eines dafür vorgesehenen Elementes einer anderen XML-Datei einbinden. Zum Beispiel: ein besonderes wissenschaftliches Format mit dem Namensraum „*XXML*“ besitzt ein Element namens „*zitiertVon*“ und die einfache bibliographische Auszeichnungssprache wird dem Namensraum „*sbml*“ zugeordnet. Um *sbml*-Daten innerhalb von *XXML* zu verwenden reicht die folgende Schreibweise: „`<XXML:zitiertVon><sbml:Buch><Titel>XML in a nutshell </Titel></Buch><zitiertVon>`“. Das verarbeitende Computerprogramm kann den Inhalt von *sbml:Buch* ignorieren (wenn es mit *sbml* nicht umgehen kann), es anders formatiert (mit XSLT-Vorlage) in der Fußnote darstellen oder diesen Datensatz in die lokale Literatursammlung übernehmen.

Eine einfache bibliographische Auszeichnungssprache ist auf diese Art und Weise kontextunabhängig und könnte innerhalb von XHTML- oder in den ISO-standardisierten ODF-Dokumenten verwendet werden. Des Weiteren kann die Einbettung der einfachen bibliographischen Auszeichnungssprache als XML in anderen XML-Dokumenten zur Lösung eines anderen Problems beitragen: Der Wechsel des Arbeitsplatzes bringt den Nachteil mit sich, dass die Literatursammlung genauso wie das Dokument auf jedem Rechner synchron sein müssen. Ist die eigene Literatursammlung in das aktuelle Dokument eingebettet, muss der Wissenschaftler nur das Dokument zwischen den benutzten Arbeitsplätzen synchronisieren (z. B. durch das Abspeichern der Datei auf einem externen Datenträger oder durch das Versenden der Datei per Email).

5.2.3.4. Hierarchische und relationale Struktur

Das XML kennt drei besondere Datentypen Namens ID, IDREF und IDREFS. Sie sollen die Identität garantieren bzw. das Verweisen ermöglichen. Somit muss ein Wert des Attribut-Typs ID in der ganzen XML-Datei einmalig sein. Das Attribut des Typs IDREF verweist immer auf ein existierendes Attribut des Typs ID. „IDREF-Attribute werden üblicherweise dazu eingesetzt, um Beziehungen zwischen Elementen zu realisieren, wenn einfache Enthalten-in-Beziehung nicht ausreicht.“²⁰⁶ IDREFS verweist auf mehrere Elemente mit dem Attribut „ID“, indem es die referenzierte IDs als eine durch Leerzeichen getrennte Liste aufnimmt.

Diese Datentypen machen es möglich, unterschiedliche Elemente miteinander zu verbinden, ohne sie mehrmals schreiben zu müssen. In einer großen Literatursammlung wird eine Person nur einmal präzise erfasst und kann von jedem anderen Element referenziert werden. Natürlich muss die Referenzierung einen Sinn ergeben: Elemente wie Autor und Herausgeber wären dazu geeignet. Ein nicht zu unterschätzender Vorteil dieser Lösung ist, dass im Falle eines Fehlers eine Korrektur nur an einer einzigen Stelle notwendig ist.

Hier wird ein weiteres Detail sichtbar: Will man diese Möglichkeiten nutzen, muss man Klassen schaffen, die alle Eigenschaften in Form von einzelnen XML-Elementen aufnehmen können. Alle Realisierungen dieser Klasse (in der Informatik "Objekte" genannt) besitzen ein ID und können mit Hilfe dieser referenziert werden. Dieses Vorgehen nennt man in der Informatik „objektrelational“. Das Element „*Buch*“ kann im XML-Element „*Autor*“ ein XML-Element namens „*Person*“ (mit allen seinen Elementen) beinhalten oder darauf lediglich verweisen. Durch das Verweisen senkt man die Redundanz der Daten. Das direkte Beschreiben im Elternelement entspricht nach Ansicht des Autors eher dem menschlichen Denken.

Um die Identität und Homogenität sicherzustellen, ist es notwendig, dass eine Person nur einmal erfasst werden darf. Jede weitere Stelle, die diese

²⁰⁶ XML in a nutshell. S. 47

Person als Autor oder Herausgeber nennt, muss sie (per ID) referenzieren. So bleibt die Datenqualität konstant und ein „einzelnes reales Etwas“ kommt nur einmal in der ganzen Datei vor.

Die Integrität der Daten wurden in MARC und MAB mit Hilfe von speziellen Datenfeldern sichergestellt. Die Quersummen sollten garantieren, dass ein Datensatz vollständig übertragen worden ist. Dies ist eine praktikable Lösung, wenn die Datensätze keinerlei Abhängigkeiten aufweisen. Ein Datensatz des MARC oder MAB beinhaltet alle Angaben zum Werk. Der Autor wird mehrfach (manchmal sogar unterschiedlich geschrieben) angegeben. Will man die Angaben zum Autor an einer einzigen Stelle vorhalten, wird man sich sehr wahrscheinlich der relationalen Methode bedienen. In diesem Falle muss überprüft werden, ob die referenzierte Stelle existiert, sonst sind die Daten inkonsistent. Bei XML kann dies ein validierender Parser feststellen: Aus dem Schema weiß er, welche Attribute als IDs und welche als IDREFs zu betrachten sind. Eine Überprüfung des XML kann falsch verwendete Elemente (wie z. B. „ISBN“ in Objekten der Klasse „Person“) aufzeigen und IDREFs auf die Existenz der verwendeten IDs untersuchen. Was ein validierender Parser (der nur die Grammatik aber keine Semantik kennt) nicht überprüfen kann, ist der Typ der ID. Deshalb liegt es nahe, für IDs spezielle Regeln zu entwerfen (sog. Konventionen). Ein Buch darf nur eine ID besitzen, die mit dem Buchstaben „b“ anfängt. Die ID einer Person könnte mit einem „p“ als Anfangsbuchstabe sichtbar gemacht werden. Die Einhaltung dieser Regeln sollten die Programme prüfen (und selber sicherstellen), die die einfache bibliographische Auszeichnungssprache umsetzen.

Die innere Struktur einer einfachen bibliographischen Auszeichnungssprache besteht aus verschiedenen Klassen. **Diese Klassen sind Gruppen von XML-Elementen und Attributen, die eine Entität mit allen seinen (und nur seinen) Facetten beschreibt.** Aus der Drei-Ebenen-Struktur werden die **Hauptklassen** „Serie“, „Buch“, „Periodikum“ und „Artikel“ abgeleitet. Eine Person ist keine Eigenschaft eines Buches (sondern nur der Autor). Solche

Klassen, die als Eigenschaften einer anderen Klasse vorkommen (z. B. Verlag im Buch) sollen innerhalb dieser Hauptklassen angegeben werden können. (Man bezeichnet dies in der Informatik auf Englisch als „**inline**“.) Damit diese inline-erfassten Klassen referenziert werden können, ist es notwendig, dass sie das Attribut „*id*“ besitzen. Werden diese Daten maschinell erfasst oder ausgegeben, werden solche **Hilfsklassen** wie „*Person*“, „*Organisation*“, „*Abstract*“ oder „*Verlag*“ aus Gründen einer einfacheren Handhabung erlaubterweise getrennt in Blöcken zusammengefasst.

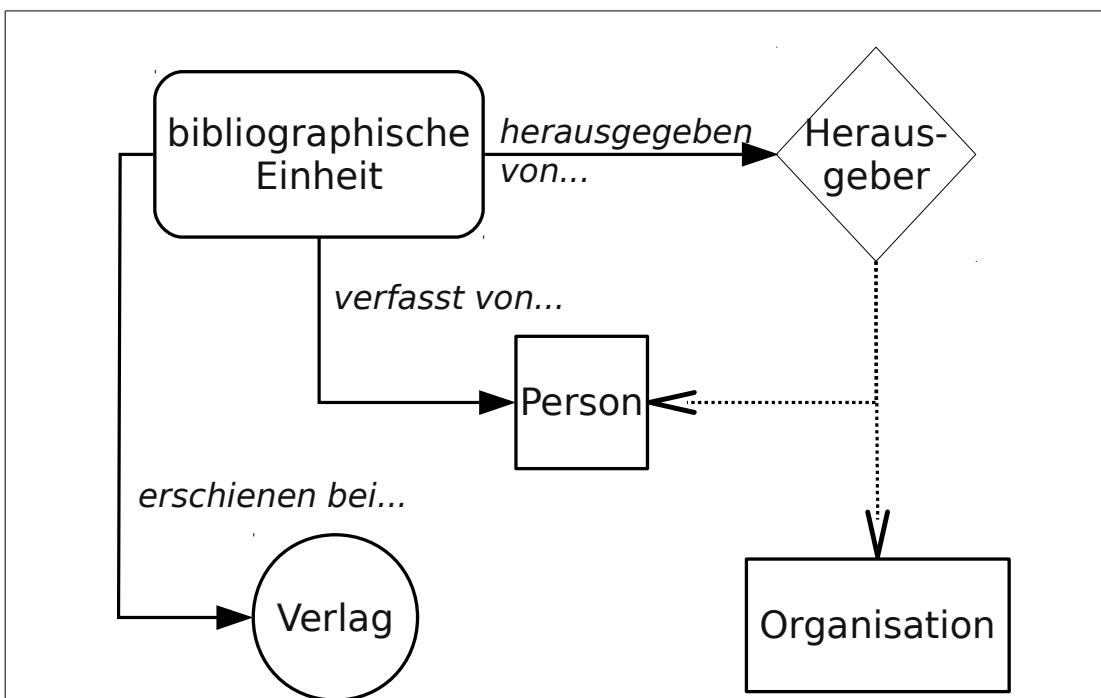


Abbildung 9: Graphische Darstellung des Klassenmodells.

Eine bibliographische Einheit wird unter Verwendung von Hilfsklassen beschrieben. (Dabei ist zu beachten, dass "Herausgeber" eine Person oder eine Organisation sein kann.) Jeder Klasse ist in dieser Darstellung eine andere geometrische Figur zugeordnet. Die Art der Beziehung wird in Kursivschrift dargestellt.

Zwischen den Klassen kann es unterschiedliche Beziehungen geben: **hierarchische Beziehungen** (vertikal), **Verweise** (horizontal) und **Abhängigkeit** (unterschiedliche Klassentypen). Auf der Abbildung 9 werden

gleich mehrere Abhängigkeitsbeziehungen dargestellt: Diese entstehen dann, wenn eine der Hauptklassen eine Hilfsklasse voraussetzt und sich ihrer bedient, um gewisse komplexe Eigenschaften (Eigenschaftsgruppen) zu erfassen.

5.2.3.5. Vertikale Beziehungen (bibliographische Hierarchie)

Vertikale Beziehungen kann es in diesem Modell nur in einer Hierarchie geben. Eine solche Hierarchie bilden die Hauptklassen. Ein Objekt der Klasse „*Serie*“ darf ein oder mehrere Objekte der Klasse „*Buch*“ oder „*Periodikum*“ enthalten. (Bei Artikel-Serien – einer abstrakten Schöpfung – darf die Serie zwei oder mehr Artikel beinhalten.) Natürlich ist eine Serie nur dann sinnvoll, wenn sie mehr als ein Element beinhaltet. Ihre sog. „Kinder“²⁰⁷ (die beinhalteten Elemente) dürfen selber weitere Elemente enthalten. Bei Zeitschriften wird diese Hierarchie sehr gut sichtbar: Jede Zeitschrift (Ausgabe) ist ein Teil einer Serie (gemeinsamer Titel, gleicher Verlag,

```
<Serie>
  <Titel>c't</Titel>
  <Verlag>
    <Name>Heise Zeitschriften Verlag</Name>
    <Ort>Hannover</Ort>
  </Verlag>
  <Periodikum>
    <Nr>26</Nr>
    <Jahr>2007</Jahr>
    <Artikel>
      <Titel>Alles unter einem Dach</Titel>
      <Seitenumfang>
        <Anfang>166</Anfang><Ende>168</Ende>
      </Seitenumfang>
      <Autor>
        <Vorname>Andreas</Vorname><Nachname>Beier</Nachname>
      </Autor>
    </Artikel>
  </Periodikum>
</Serie>
```

Text 16: Beispiel für eine vertikale Beziehung zwischen den Hauptklassen. Jede Eltern-Klasse beinhaltet die Kinder-Klasse. Die Form einer direkten Inklusion („inline“).

207 Im XML wird die Hierarchie als Parent-Child-Beziehung bezeichnet. Das beinhaltende Element wird „Parent“ (Eltern) und das beinhaltete Element „Child“ (Kind) genannt.

derselbe Herausgeber). Gleichzeitig ist es kein in sich geschlossenes Werk einer einzelnen Person zu einem Thema. Vielmehr sind es viele Teile (Artikel), die zusammen in einem Heft eine Zeitschrift ergeben. Bücher müssen keine Artikel besitzen, obwohl das bei wissenschaftlichen Werken der Fall ist.

Diese Art der Beziehungen ist nach Ansicht des Autors mit XML sehr einfach zu realisieren. Das Element „**Serie**“ besitzt mehrere Kinder-Elemente wie z. B. *Titel*, *Herausgeber*, *Verlag*. Hinzu kommen mehrere Elemente des Typs „*Buch*“ oder „*Periodikum*“. Diese dürfen neben den Elementen wie Erscheinungsdatum, Auflage oder Seitenzahl mehrere Kinder-Elemente des Typs „*Artikel*“ besitzen.

Eine für die Menschen etwas schlechter lesbare Lösung, die im XML ebenfalls realisierbar ist, besteht darin, die Kinder-Elemente an einem anderen Ort zu erfassen und mit Hilfe von ID und IDREF / IDREFS zu referenzieren. Dabei ist zu beachten, dass jedes Kinder-Element ein ID-Attribut besitzen muss und das Eltern-Element die alternative Schreibweise in Form von „*<Periodikum><Artikel enthaelt='artikel1 artikel2 artikel3'/></Periodikum>*“ unterstützen muss. Der Vorteil einer solchen Lösung liegt darin, dass Artikel, die in mehreren Werken abgedruckt worden sind, an einer einzigen Stelle erfasst werden. Es ist jedoch für den Menschen schlecht lesbar und nur für Computerprogramme gut geeignet. Daher soll die bereits erwähnte Regel gelten: Die Inklusion von Kinder-Elementen soll ebenso erlaubt sein, wie die Referenzierung.

5.2.3.6. Horizontale Beziehungen

Neben den Haupt- und Hilfsklassen gibt es einen dritten Datentyp, der von großem Nutzen sein könnte. Dieser ist die „Relation“, die nicht selbständig sondern immer in Verbindung mit zwei Klassen vorkommt. Sie kann in Form von Attributen des Typs IDREF oder IDREFS realisiert werden. In einigen Fällen sind einige Relationstypen unverzichtbar. Dies gilt für unterschiedliche Schreibweisen eines Personennamens oder Übersetzungen von bibliographischen Werken. Solche Relationen drücken die Gleichwertigkeit

zwischen zwei Objekten aus. „Benedikt XVI.“ ist bibliographisch gleichwertig zu „Joseph Ratzinger“, dennoch ist es aus Sicht der Theologen nicht unerheblich, wer als Autor genannt wird²⁰⁸. Ähnlich verhält es sich mit Übersetzungen: der Inhalt ist gleichwertig, wenn auch die Übersetzung nicht so hoch angesehen werden kann wie das Original²⁰⁹. Solche horizontalen Beziehungen existieren zwischen Objekten desselben Typs: „Eine Person“ ist „einem Artikel“ formal nicht gleichwertig.

Es ist ebenso wichtig zu beachten, dass die Relationen immer zu einem bestimmten Typ gehören. Die Äquivalenz und die Übersetzung wurden bereits genannt. Weiter denkbare Typen wären „vorhergehend“ und „fortführend“ (z. B. für Artikel, die in mehreren Teilen erscheinen) oder „gehört zu“ (bei Organisationen wie z. B. „NABD gehört zu DIN“). Eine bereits in der SWD genutzte Relation ist ein Schlagwort und seine Vorzugsbenennung: Ein Schlagwort verweist auf eine Vorzugsbenennung als normierten Ersatz, ohne darüber eine Aussage zu treffen, ob das Schlagwort ein Oberbegriff oder eine Äquivalenz ist.

Solche Relationen sind nötig, um die unterschiedlichsten Beziehungen zwischen den Werken (oder Personen) herzustellen. Es ist denkbar, dass jemand mit einem Schlagwort sucht (welches nur ein Aspekt eines größeren Zusammenhangs ist) ohne zu wissen, wie man weitere Aspekte richtig bezeichnen soll. An dieser Stelle wäre eine Auflösung des Schlagwortes über die Vorzugsbenennung auf alle anderen Schlagworte möglich.

5.3. Umsetzung

Für die Umsetzung einer einfachen bibliographischen Auszeichnungssprache scheint das XML aus den im vorherigen Kapitel genannten Gründen

208 Die Regeln des RAK verlangen vom Katalogisierer ebenfalls die Nennung des Autors in der angegebenen Form (wenn auch als Verweisung). Dies Vorgehen ist für das Wiederfinden des Werkes zweckmäßig.

209 Dies kann nach Ansicht des Verfassers vor allem in der Literaturwissenschaft oder bei historischen bzw. juristischen Quellen gelten.

hinreichend zu sein. Darauf baut das Klassenmodell auf²¹⁰. Es besteht aus einer Gruppe hierarchisch geordneter bibliographischer Klassen und einer Sammlung von Hilfsklassen, die einige zusammenhängende Merkmale einer bibliographischen Einheit beschreiben. Die Hilfsklassen dienen der Eindeutigkeit und sollen verhindern, dass z. B. eine Person (die als Autor, Dolmetscher oder Herausgeber in unterschiedlichen Werken mehrfach vorkommen kann) mehrfach erfasst werden muss. Diese Person (nicht bloß ihr Name) ist aus bibliographischer Sicht ein Merkmal eines oder mehrerer Werke. Somit hilft die Klasse Person, ein Objekt der Klasse Buch zu beschreiben, ist von diesem jedoch zu unterscheiden.

Diese Vorgehensweise stammt aus der Mengenlehre, in der eine Gruppe von gemeinsamen Eigenschaften mehrerer einzelnen Objekte „Klasse“ genannt wird. Dieses Prinzip wird sowohl in der Informationswissenschaft als auch in der Informatik (in der Objekt-Orientierten-Programmierung) angewendet. Hier wird der besseren Verständlichkeit wegen die IT-nahe Deutung dargestellt. Darin verfügt eine Klasse über Eigenschaften (wie „Farbe“, „Länge“, „Gewicht“) und Methoden (wie „es kann fahren“, „es läuft“, „es steht“). Alle Eigenschaften und Methoden einer Klasse sind allen Realisierungen dieser Klasse (sog. Objekten) gemeinsam. Eine Klasse ist sozusagen der kleinste gemeinsame Nenner einer Gruppe von Objekten desselben Typs. Sind die Objekte in weitere Klassen aufteilbar (Unter-Klassen), übernehmen sie alle Methoden und Eigenschaften ihrer Eltern-Klasse. Dies nennt man „Vererbung“. Am Beispiel eines Buches würde dies wie folgt aussehen:

Alle Bücher haben einen Umschlag, eine Anzahl an Seiten, einen Titel und einen Autor bzw. mehrere Autoren oder einen bzw. mehrere Herausgeber. Darunter können wir zwischen Belletristik und Sachbüchern unterscheiden. Allen Büchern aus der Gruppe Belletristik ist es eigen, dass sie keine

210 Theoretisch ist jede Notation denkbar. Die Darstellung ist von Inhalt – dem Klassenmodell – zu trennen. Aus den bereits genannten Gründen ist das XML für den Austausch und die Verarbeitung (nicht für die Aufbewahrung) optimal. Es ist sehr wahrscheinlich, dass dieses Datenmodell ebenfalls als interne Struktur der verarbeitenden Programme sein wird, wo es nicht als XML vorkommen wird.

wissenschaftliche Literatur sind (bzw. „eher der Unterhaltung dienen“).
Zugleich erben sie von der Klasse „Buch“ alle oben genannten
Eigenschaften.

Das Wissen um die Funktionsweise dieser Klassifikation ist für das
Verständnis der Hauptklassen unentbehrlich. An dieser Stelle soll noch kurz
auf einen besonderen Typ eingegangen werden: Die Meta-Klasse
Herausgeber ähnelt der Kontravalenz in der Logik. Es ist sozusagen eine
Pseudo-Klasse, die auf ein Objekt der Klasse Person oder der Klasse
Organisation verweist. Sie besitzt selber keine Eigenschaften. Dadurch wird
es möglich, eine Aussage zu treffen, die zwei bereits definierte Klassen
verwendet: „Das Buch X haben **herausgegeben**: Prof. A, Prof. B und die
Organisation C“.

Das XML verlangt die Existenz eines sog. „Wurzelementes“, das alle übrigen
Elemente umschließt und als oberstes „Elternelement“ fungiert. Der Name ist
frei wählbar. Dieser soll hier „Bibliographie“ heißen.

5.3.1. Hauptklassen

Die Hauptklassen **Serie**, **Buch**, **Periodikum** und **Artikel** sind hierarchisch
angeordnet und bilden drei Stufen. Die oberste und abstrakteste Klasse stellt
die Serie dar. Zu ihr können Periodika oder Bücher gehören. In besonderen
Fällen kann sie nur Artikel beinhalten. Die Ebene darunter bilden die Klassen
Buch und Periodikum. Ein Objekt der Klasse Buch kann mehrere (mehr als
ein) Objekte der Klasse Artikel enthalten. Ein Objekt der Klasse Periodikum
muss ein oder mehr Objekte der Klasse Artikel enthalten. Der Artikel und das
Buch bilden das Fundament dieser Hierarchie, weil sie die elementarsten
bibliographischen Werke darstellen. Gemäß der „Vererbung“ übernehmen sie
von allen höheren Instanzen deren Eigenschaften z. B. den Serientitel, den
Herausgeber (aus Serie oder Buch) oder den Verlag.

5.3.1.1. Serie

Eine Serie ist eine abstrakte "Schöpfung" und beinhaltet (sinngemäß) immer
zwei oder mehr Kinder-Elemente. Es kann selten vorkommen, dass eine

Serie nur ein Element besitzt. Dies betrifft vor allem die Fortsetzungswerke: solange kein zweites Element der Serie herausgegeben worden ist, besitzt diese nur ein Element. Eine Serie ohne Kinder-Elemente ergibt aus zwei Gründen keinen Sinn:

1. Eine Serie ist eine Abstraktion vom Einzelwerk und beinhaltet alle gemeinsamen Eigenschaften ihrer Elemente (wie Herausgeber, Verlag oder Titel bzw. Gesamttitel) im Sinne einer Konjunktion der Logik.
2. Eine Serie realisiert eine besondere Beziehung zwischen einzelnen bibliographischen Werken, die zu ihr gehören. Es kann eine Fortsetzung (mehrbändiges Fortsetzungswerk) sein oder einen Zusammenhang (einzelne Bände eines Lexikons).

In beiden Fällen sind zumindest zwei Werke notwendig, um eine Serie daraus zu abstrahieren. Will ein Benutzer schon im Voraus eine Serie anlegen, so muss er genau wissen, welche Eigenschaften allen nachfolgenden Werken gemeinsam sein wird. Eine zuverlässige Aussage diesbezüglich kann aber nur der Herausgeber treffen.

Eine solche Abstraktion bringt zwei Vorteile mit sich: die Vermeidung von Redundanz und die damit verbundene Reduktion der Datenmenge. Alle Werke, die zu einer Serie gehören übernehmen („erben“) von ihr alle darin enthaltene Daten wie Herausgeber, Serientitel oder Verlag. In einigen Fällen kann dieser Vorteil zu Schwachstellen führen. Ändert der Herausgeber seinen Namen, oder wird die Serie (z. B. aufgrund einer Firmenübernahme) von einem anderen Verlag herausgegeben, darf das betroffene Feld nicht einfach ersetzt werden. Ein solches Vorgehen würde zu falschen Daten führen. Die einzige Möglichkeit besteht darin, den Datensatz der Serie in eine neue Identität zu kopieren (neue ID) und unter Nutzung der Relation „fortführend“ auf den ursprünglichen Datensatz zu verweisen. Damit sind beide Serien-Elemente gleichwertig (z. B. bei der Suche) aber nicht beliebig austauschbar.

Datenfelder einer Serie sind: der (*Serien-*)*Titel*, der *Untertitel* (optional), der *Herausgeber* (optional), der *Verlag* (optional) und die vier optionalen, mehrfach anwendbaren Ordnungs- und Beschreibungsfelder: *Sprache*, *Klassifikation*, *Schlagwort* und *Abstract*. Bei den wissenschaftlichen Zeitschriften (sog. Journale) hat sich die Verwendung eines eindeutigen *Kürzels* an Stelle des Serientitels eingebürgert, deshalb soll diese Abkürzung angegeben werden können. Sind an der Entstehung oder der Pflege einer bibliographischen Serie Personen oder Organisationen beteiligt, die nicht als Herausgeber gelten, können sie im Element „*Mitarbeiter*“ oder „*Organisation*“ erfasst werden. Deren Beteiligung sollte im Attribut „*Funktion*“ näher beschrieben werden. Z. B: „in Zusammenarbeit mit“ oder „gegründet von“.

Die *ID* einer Serie sollte durch die Benutzung des Buchstabens „s“ am Anfang zusätzlich für Klarheit sorgen, dass es sich hierbei um den Datentyp „*Serie*“ handelt. Optional könnte man den Vermerk zur Lizenz und den Urheber-Rechten im Feld „*Lizenz*“ aufnehmen. Ebenso ist es nützlich, ein Feld namens „*Identifikator*“ mit einer URI, DOI, ISSN, ISSN-L oder Signatur (die auf eine Auflistung der Volltexte der Serienteile aufgelöst werden kann bzw. die Serie als Ganzes identifiziert) optional zu ergänzen. Die zulässigen optionalen Relationen sind: „*fortführend*“ (als Verweisung auf die frühere Identität einer Serie, die fortgesetzt wird) und „*Äquivalenz*“ sowie „*Original*“ (für Übersetzungen etc).

Die Kinder-Elemente einer Serie sind: *Buch*, *Periodikum* und *Artikel*.

5.3.1.2. Buch

Als Buch soll ein gebundenes Werk gelten. Es kann keine oder mehrere Artikel enthalten. Ein Buch wird grundsätzlich Zeit- und Ereignis-unabhängig herausgegeben und erscheint deshalb nicht periodisch. Dies ist ein wichtiges Unterscheidungsmerkmal zwischen den Periodika und den Büchern. Die „Definition“ eines Buches ist den meisten Menschen geläufig und trägt so zur Einfachheit bei. In der Überzeugung der meisten Menschen ist ein Buch eine „dickere“ Veröffentlichung mit Einband. Dieses Verständnis ist irreführend, da auch wissenschaftliche Zeitschriften für eine bessere Lagerung in den

Bibliotheken gebunden werden. Die Seitenzahl ist für eine richtige Zuordnung ebenso wenig entscheidend²¹¹. Das "Jahresbuch" des Statistischen Bundesamtes ist ebenfalls ein Sonderfall: sein Titel und Umfang bzw. physikalische Beschaffenheit legen ein "Buch" nahe. Seine periodische Erscheinungsweise deutet den Typ "Periodikum" an. Im Gegensatz zum Buch muss ein Periodikum ein oder mehr Artikel enthalten. Das Statistische Jahresbuch kann mangels Aufteilung in Artikel nach dieser Definition nur als Buch fungieren.

Es ist besonders wichtig auf die Wiederholbarkeit zu achten. Erscheint ein „*Buch*“ mehrmals mit gleichen Titel oder Untertitel (wie „Konferenzband der XX. Tagung der Gesellschaft N.N.“) und ist mit der Erscheinung ein bestimmtes Geschehen oder ein gewisser Zeitabschnitt in Zusammenhang zu bringen, so handelt es sich um ein Periodikum. Bei einem Buch können wir lediglich die Band-Nummer und die Auflage-Nummer erfassen. Bei den Periodika ist die Nummer des Jahrgangs zulässig. Die Nummerierung bei diesen drückt eine zeitliche Folge aus (die Relation „*fortführend*“) - dies ist die übliche Relation. Eine solche tritt bei Büchern nur sehr selten auf (z. B. wenn Prof. X nach dem Tode des Prof. Y dessen Werk „Lehrbuch / Handbuch des ...“ fortführt und aktualisiert, dann wird eine neue Auflagenummer verwendet, die manchmal um einen Hinweis in Form eines Textes erweitert wird). Die Bücher haben im Gegensatz zu den Periodika die Besonderheit, dass sie unter demselben Titel (oft mit gleichem Inhalt) mehrfach aufgelegt werden können. Zu erkennen ist dies am Fehlen einer Folgenummer sowie einer höheren Auflagenummer (wenn vorhanden).

Ein Buch besitzt zumindest einen *Titel*, einen *Untertitel* (optional), beliebig viele *Autoren*, beliebig viele *Herausgeber* (aber zumindest einen

211 UNESCO erwähnt zwar die Seitenzahl als Zuordnungsmerkmal, nennt aber viele weitere. („A book is a non-periodical printed publication of at least 49 pages, exclusive of the cover pages...“ in: *Recommendation concerning the International Standardization of Statistics Relating to Book Production and Periodicals*, II. 6. a, URL: http://portal.unesco.org/en/ev.php-URL_ID=13068&URL_DO=DO_TOPIC&URL_SECTION=201.html, Stand: 10.05.2008)

Herausgeber, wenn kein Autor genannt werden kann²¹²), einen *Verlag* (optional²¹³), eine *Bandbezeichnung* (optional), eine *Auflagenummer* (Zahl; optional), *Edition* (wie „gebunden“ oder „erweitert“; optional), das *Erscheinungsjahr* (vierstellige Zahl; optional), der *Seitenumfang* (Zahl) und einen *Hyperlink zum Volltext* (URI oder DOI, die dorthin verweist; optional, auch mehrfach) sowie beliebig viele *Identifikatoren* (DOI, URI, ISBN, ISSN, ISSN-L oder Signatur, optional). Die *Mitarbeiter* und beteiligte *Organisationen* werden beliebig oft zusammen mit ihrer jeweiligen Funktion angegeben. Ebenso wie bei einer Serie ist es beim Buch zulässig, die vier optionalen, mehrfach anwendbaren Ordnungs- und Beschreibungsfelder (*Sprache*, *Klassifikation*, *Schlagwort* und *Abstract*) sowie „*Lizenz*“ (optional) zu verwenden. Ein in der wissenschaftlichen Community als üblich geltendes *Kürzel* kann angegeben werden, um seine Verwendung als Zitierschlüssel zu ermöglichen. Die *ID* eines Buches sollte mit dem Buchstaben „b“ am Anfang versehen werden. Die zulässigen optionalen Relationen sind: „*gehört_zu*“ (an Stelle direkter Inklusion im Eltern-Objekt), „*fortführend*“ (als ordnendes Merkmal für Serien mit vorgegebener chronologischer bzw. thematischer Reihenfolge oder als Verweisung auf einen „Titel“ der z. B. nach dem Tod des Autors mit diesem Buch fortgesetzt werden soll) und „*Äquivalenz*“ sowie „*Original*“ (für Übersetzungen und andere Ausgaben dieses Werkes).

Die Diplom-, Magisterarbeiten, Promotions- oder Habilitationsschriften müssen aufgrund ihrer inhaltlichen Struktur in diesem System als Buch

212 Dieses Problem kann oft auftreten. Bei historischen Quellen, deren Autor unbekannt – dafür aber der Auftraggeber (der Herausgeber) bekannt ist – oder bei gemeinschaftlichen Werken auf „Community-Basis“ wie Wikipedia. In diesem Fall wäre die „Wikimedia Foundation Inc.“ als Verlag (gemäß ISO 639-2 Regeln) oder sogar als Herausgeber anzusehen.

213 In den RAK-Regeln wird bei Werken ohne Verlag von „Selbstverlag“ gesprochen. Dies sagt zwar aus, dass ein Exemplar des Werkes nur beim Autor zu erhalten ist, ist beim Zitieren jedoch unüblich. Damit ist diese Information nur für den Bibliothekar, der ein solches Werk besorgen soll, von Interesse. Online-Handbücher besitzen ebenfalls keinen Verlag. Hier reicht die URL aus. Die Erwähnung des „Selbstverlages“ wäre an dieser Stelle ebenfalls unüblich.

erfasst werden. Die Angaben zur Universität und der Fakultät (bzw. Fachrichtung) sollten deshalb in Elementen des Typs „*Organisation*“ Platz finden. Die Gutachter und sonstige beteiligte Personen können im Element „*Person*“ eingetragen werden. In beiden Fällen ist die Nennung der Funktion notwendig.

Für eine maschinelle Generierung von Quellenangaben wird zusätzlich das Attribut "*Typ*" benötigt. Es legt fest, welcher der Vorlage-Typen angewendet werden soll. Wurde das Attribut nicht explizit angegeben, gilt "*Monographie*" als Vorgabe. Die folgenden Typen stehen zur Verfügung: "*Monographie*", "*Dissertation*", "*Abschlussarbeit*" (als Sammelbegriff für Diplom-, Magister-, Master- und andere wissenschaftliche Arbeiten, die für die Erreichung eines akademischen Grades oder einen Abschluss notwendig sind), "*Report*", "*Handbuch*", "*Katalog*", "*Atlas*" und "*Lexikon*". Zusätzlich darf der Benutzer selber weitere Typen definieren, die beim Fehlen einer passenden Vorlage als Monographie-gleichwertig behandelt werden.

Ein Buch kann nur Elemente der Klasse *Artikel* als Kinder-Elemente besitzen. Es kann keine oder mehrere Artikel enthalten.

5.3.1.3. Periodikum

Das Periodikum weicht in vielerlei Hinsicht vom Buch ab: statt der Auflagenummer ist die Folgenummer und Jahrgang (und/oder Erscheinungsjahr) für die richtige zeitliche Zuordnung wichtig. Auch die Erscheinungsweise, die in den oben genannten bibliographischen Angaben in Verbindung steht, unterscheidet das Periodikum deutlich von einem Buch: Ein bestimmtes wiederkehrendes Ereignis oder eine Zeitperiode ist für die Herausgabe entscheidend. Die Inhalte ändern sich im Gegensatz zu den Büchern in jeder Auflage. Gute Beispiele für Periodika sind: Zeitschriften und Konferenz- oder Tagungsbände.

Im Zweifelsfall können: die Ähnlichkeit des Titels (mit variierenden Zahlen), das Fehlen des Titels, die Angabe einer Nummer oder eines Datums (nicht bloß Jahres) sowie das Fehlen einer Auflagenummer als Hinweise für die Zugehörigkeit zur Klasse „Periodikum“ angesehen werden.

Die Klasse Periodikum besitzt neben den vier optionalen, mehrfach anwendbaren Ordnungs- und Beschreibungsfeldern (*Sprache*, *Klassifikation*, *Schlagwort* und *Abstract*) die folgenden Felder: *Herausgeber* (wenn nicht in der übergeordneten Serie enthalten), *Titel*, *Untertitel* (beides optional und nur anzugeben, wenn ein Heft einen eigenen Titel besitzt), *Jahrgang* (optional), *Datum* (optional; formatiert gemäß XML-Schema-Datentyp „date“), *Nummer* (optional: wenn keine Nummer vorhanden, ist das Datum anzugeben), *Verlag* (optional: wenn nicht von der Serie „geerbt“ wird), einen optionalen Identifikator wie *ISSN*, *ISSN-L*, *ISBN* (für Tagungsbände), DOI, URI oder Signatur; *Hyperlink zum Volltext* (URI oder DOI zur Liste der Volltexte der Artikel; optional) und „*Lizenz*“ (optional). Mit „*Kürzel*“ soll dem Benutzer die Möglichkeit gegeben werden, das in der Fach-Community übliche Kürzel einzugeben. Da Periodika im wissenschaftlichen Umfeld eine fortlaufende Seitennummerierung besitzen können (damit sie gebunden besser zu handhaben sind), sollten die Angaben zum Umfang (als eine Zahl) bzw. der enthaltenen Seiten (als „*Anfang*“- und „*Ende*“-Felder) angegeben werden können. Ist das Periodikum gebunden und besitzt eine Bandnummer, wird diese unter „*Band*“ angegeben. Das *ID*-Attribut eines Periodikums beginnt mit einem „*p*“. Die zulässigen optionalen Relationen sind: „*gehört_zu*“ (an Stelle direkter Inklusion im Eltern-Objekt), „*fortführend*“ (als ordnendes Merkmal für Serien mit vorgegebener chronologischer bzw. thematischer Reihenfolge) und „*Äquivalenz*“ sowie „*Original*“.

In besonderen Fällen wird es nötig sein, die beteiligten Organisationen oder Personen zu nennen. Eine solche Angabe sollte von einer kurzen Bezeichnung der Art der Beteiligung begleitet sein.

Für eine maschinelle Verarbeitung zu Quellenangaben wird zusätzlich das Attribut "*Typ*" verwendet. Dieser steuert die Auswahl der richtigen Vorlage (Zitierstil). Implizit gilt "Journal" als Vorgabe. Folgende Typen können explizit angegeben werden: "*Journal*", "*Konferenz*", "*Bericht*" und "*Jahrbuch*". Zusätzlich darf der Benutzer weitere Typen definieren, die beim Fehlen einer passenden Vorlage ignoriert und als "Journal" behandelt werden.

Das einzige zulässige Kinder-Element gehört zur Klasse „*Artikel*“.

5.3.1.4. Artikel

Artikel sind bibliographisch gesehen die elementarsten Teile einer Veröffentlichung. Ihr Entstehen kann man relativ eindeutig einer Person oder mehreren Personen zuordnen (Autoren). Artikel besitzen keinen Herausgeber (können ihn nur von den übergeordneten Instanzen „erben“). Möglich ist eine Beteiligung von Organisationen²¹⁴ und Personen an der Entstehung von Artikeln. Wünschenswert (z.B. für eine bessere Transparenz) wäre nach Ansicht des Verfassers eine Angabe zum Typ ihrer Tätigkeit wie z.B.: „finanziell unterstützt von“, „im Auftrag von“, „der Illustrator“, „übersetzt von“ o. Ä.

Artikel sind fast immer ein Teil eines „größeren Ganzen“ wie ein Buch oder Periodikum und werden in einigen Fällen unterteilt. Manchmal werden sie unverändert oder übersetzt in anderen Werken abgedruckt. Dies macht den Umgang mit den Artikeln so schwierig.

Grundsätzlich gilt: Ein Artikel ist die kleinste Einheit, die immer zumindest einer Person zugeschrieben werden kann. Sie besitzt einen Titel und kann anhand der Seitenzahlen im übergeordneten Werk wiedergefunden werden. Ist ein Artikel in mehreren Teilen abgedruckt worden, kann man die Relation „*fortführend*“ für die Zuordnung der richtigen Reihenfolge verwenden. Wird ein Artikel in mehreren Werken unverändert abgedruckt, so muss man wegen der Unterschiede an Seitenzahlen ebenfalls eine Kopie anlegen und die Relation „*Übersetzung von*“ bzw. „*Äquivalenz*“ hinzufügen. Soll eine Reihe von Artikeln aus verschiedenen selbständig erschienenen Werken als eine „virtuelle Gesamtausgabe“ gesehen werden, kann man die Serie als Container benutzen. Dabei verweist die Serie auf die Artikel mithilfe ihrer ID und zwar unabhängig davon, wo sie physikalisch zuzuordnen sind.

(Das Kopieren von Objekten und Anpassen einzelner Felder löst zwar einige Probleme, führt jedoch zu Redundanzen. Denkbar wäre eine Lösung, in der auf eine ID verwiesen wird und nur die unterschiedlichen Datenfelder erfasst

214 Vgl. Feld „Organisation, bei der die Person beschäftigt ist“ im MAB-Format

würden. Auf diesem Wege würde man den Redundanzen vorbeugen, jedoch wäre eine solche Datei mit XSL schwer zu verarbeiten.)

Ein Artikel beinhaltet folgende Datenfelder: zumindest einen *Autor*, einen *Titel*, einen *Untertitel* (optional), eine *Seitenangabe* (als Aufzählung in „Seite“-Feldern oder als „Anfang“- und „Ende“-Felder), optional einen *Hyperlink zum Volltext* (URI, DOI), einen *Identifikator* (optional) sowie „Lizenz“ (optional) und die optionalen, mehrfach anwendbaren Ordnungs- und Beschreibungsfelder (*Sprache*, *Klassifikation*, *Schlagwort* und *Abstract*). Beteiligte Personen und Organisationen werden mit ihrer jeweiligen Funktion angegeben. Die *ID* von Artikeln sollten der Übersicht wegen mit einem „a“ anfangen. Die zulässigen optionalen Relationen sind: „gehört_zu“ (an Stelle direkter Inklusion im Eltern-Objekt), „fortführend“ (für virtuelle Serien mit vorgegebener chronologischer bzw. thematischer Reihenfolge), „Äquivalenz“ und „Original“. In Wissenschaften, deren Publikationen sehr stark Artikel-orientiert sind, kann es nützlich sein, ein *Kürzel* anzugeben, welches als Zitierschlüssel gelten kann.

Das Attribut "*Typ*" legt fest, welcher besondere Zitierstil angewendet wird. Die folgenden zwei Typen stehen zur Auswahl: "*Artikel*" und "*Loseblattausgabe*". Die implizite Zuordnung ist "*Artikel*". Zwischen selbständig und unselbständig erschienenen Artikeln wird nicht weiter unterschieden. Die Verarbeitung der Angaben des Elternelements als dem beinhaltendem Werk (meist hinter "In:") richtet sich nach dessen Typ. In einigen Fällen kann es sinnvoll sein, einen anderen Typ als Freitext anzugeben. Einige neue Veröffentlichungstypen wie "poster" oder "research paper" werden je nach Zitiernorm besonders stark oder gar nicht berücksichtigt. Andere wie "Loseblattausgabe" oder Stadtpläne können schon dem Umfang nach weder Büchern noch Periodika zugeordnet werden. Damit sie entsprechend der verwendeten Norm zitierbar bleiben, sollen sie einen eigenen Typ bekommen.

Ein Objekt des Typs „Artikel“ als das unterste Element der bibliographischen Hierarchie kann keine Kinder-Elemente enthalten.

5.3.2. Hilfsklassen

Hilfsklassen sind Gruppen von Attributen, die meistens mehr als einem Werk eigen sind und als solche kein bibliographisches Werk darstellen aber zur Beschreibung solcher notwendig sind. Typisches Beispiel für ein Objekt einer Hilfsklasse ist eine Person. Sie kann in mehreren Büchern als Autor, Herausgeber oder Übersetzer vorkommen. Die Wiederverwendbarkeit dieser Objekte beugt Redundanzen vor und macht die Daten homogener. Zu den Hilfsklassen gehören: Person, Verlag, Organisation und Abstract.

5.3.2.1. Person

Die Klasse Person ist eine der häufigsten Hilfsklassen²¹⁵. Darin können Personennamen genau erfasst werden. Mithilfe einer *ID*, die einer besseren Lesbarkeit wegen mit einem „*m*“ eingeleitet werden sollte, wird zunächst eine Schreibweise eines Personennamens und nicht die Person als solche eindeutig identifiziert. Besitzt eine Person mehr als eine Schreibweise Ihres Namens (wie „Friedensreich Hundertwasser“ und „Friedensreich Stowasser“), werden diese Schreibweisen bei einer Suche als gleichwertig betrachtet. Um nicht alle möglichen Schreibweisen bei der Verweisung berücksichtigen zu müssen, empfiehlt es sich immer, zur ersten originalen Schreibweise (d. h. Geburtsname in Muttersprache) hin zu verlinken. Mit einer Abfrage nach allen Personennamen, die auf diese Schreibweise verweisen, kann eine Rückwärtsauflösung erreicht werden²¹⁶.

Mit Hilfe des XML-eigenen „*lang*“-Attributes soll die Sprache angegeben werden, in welcher diese Schreibweise verwendet wird (sofern nicht mit der Muttersprache der angegebenen Person identisch). Diese Lösung ermöglicht

215 Wie bedeutend ein geeignetes Format für die Erfassung von personenbezogenen Daten zeigt unter anderem die Umfrage der KIM-AG: Personen gehören zu den 3 am meisten beschriebenen „Objekte“ und stehen hinsichtlich der Anzahl verwendeter Formate an zweiter Stelle, hinter Büchern. (Vgl: Kompetenzzentrum Interoperable metadaten (Hrsg.): *Umfragerport zur Nutzung von Metadaten*. URL: <http://www.kim-forum.org/material/pdf/KIM-Umfragerport.pdf> Stand: 28.06.2008)

216 Dieses Prinzip verwenden die RAK-WB (Ansetzungsform und Verweisungsformen nach §§ 311 und 314) und die PND.

die Nutzung von Übersetzungen (bzw. in einer Sprache gebräuchliche Schreibweise: Translationen oder Transliterationen) von Personennamen als gleichwertig zum Original²¹⁷.

Erfasst und verlinkt wird die im Werk angegebene Schreibweise. Auf diese Art und Weise kann jeder Erfasser schnell und unproblematisch die Angaben zum Autor oder Herausgeber übernehmen. Will man in einer größeren Wissenschaftler-Gemeinschaft die Identität trotz vieler Schreibweisen bewahren, sollte man den Geburtsnamen in der Muttersprache des Autors („das Original“) herausfinden und einmal erfassen (analog zur Ansetzungsform in RAK-WB).

Das Problem der Namens-Homonymie wurde in RAK-WB sehr elegant gelöst: Hinter den Personennamen werden die Lebensdaten (Jahreszahlen) angehängt. Reicht dies nicht aus, wird von PND eine Angabe zur Tätigkeit (weiter: Werktitel, Wirkungsjahre, Wirkungsort, Geburtsort, Sterbeort, Adelstitel etc) erwartet²¹⁸. Da solche Polysemien in kleinen Arbeitsgruppen nur selten auftreten, soll das Feld „Anmerkungen“ einen Platz für derartige Orientierungshilfen bieten. Ein optionaler *Identifikator* zu anderen Datenbanken (die im Attribut "Typ" frei definiert werden dürfen) kann für Verlinkung und bessere Einbettung in fremde Datenbestände sorgen.

Ein weiteres Problemfeld sind historische, Ruf- oder Amtsnamen. Diese bestehen nicht aus Vor- und Zunamen, sondern aus Vornamen, Zahlen oder Ortsangaben in wechselnder Reihenfolge. Auch lateinische Eingeschaftsbeschreibungen (wie „Gallus Anonymus“) oder Zusätze wie „Heilig“ sind nicht selten. Dieses Problem stellt die Linguistik vor eine große Herausforderung. In einigen Fällen ist die Reihenfolge der Angaben entscheidend („Alexander der Große“ und „der große Alexander“). Daher dürfen diese Namen nicht

217 Zum Problem der multilingualen Recherche nach Eigennamen Vgl. Womser-Hacker, Christa : *Zur Rolle von Eigennamen im Cross-Language Information Retrieval*. In: *Festschrift Zimmermann*. S. 123

218 Vgl.: Arbeitsstelle für Standardisierung der Deutschen Nationalbibliothek (Hrsg.) : *Individualisierungsrichtlinie für die Personennamendatei*. S. 4 ff. URN: urn:nbn:de:1111-20040721161

umformatiert werden²¹⁹. „Benedikt XVI.“ ist ein feststehender Begriff und zugleich alternative Schreibweise zu „Joseph Ratzinger“ oder dem lateinischen „Benedict XVI.“. Solche Bezeichnungen sollen im Element „*Rufname*“ eingegeben werden können.

Alle übrigen Namen, die aus Vor- und Nachnamen bestehen, sollen Teil für Teil angegeben werden. Das Feld „*Nachname*“ muss den ganzen Nachnamen beinhalten. Dies betrifft Doppelnamen sowie die dazu gehörigen Präpositionen. Namenszusätze wie „der Ältere“ oder „Senior“ sind keine Teile des Nachnamens und sollen wegen ihrer identifizierenden Funktion (Zusatzmerkmal) im Feld „*Suffix*“ optional angegeben werden können. Die Vornamen, deren Anzahl stark variieren kann, sind in einer bestimmten Reihenfolge angeordnet, die beachtet werden muss. Deshalb soll das Feld „*Vorname*“ mehrfach vorkommen dürfen, jedoch in der richtigen Reihenfolge.

Zu beachten ist, dass die scheinbaren Vornamen – wie sie z.B. im Russischen gebräuchlich sind – „Iljitsch“ oder „Konstantinowna“ eine Ableitung des Vornamens des Vaters ist und nicht der Vorname der Person selbst. In einigen asiatischen Ländern fließt der Vorname des Vaters in unveränderter Form sowie ein ehrenvoller Beiname, Herkunftsort oder ein Spitzname in den vollen Namen mit ein.²²⁰ Im Falle des Kunstprofessors Daoud Salman Anad kann kaum ein Europäer diese Namensbestandteile richtig einordnen. Dies macht nach Ansicht des Verfassers die Handhabung solcher Namen etwas schwieriger. Zudem bedarf es einer grundsätzlichen Diskussion, wie und ob solche „fremden“ Namensbestandteile in Quellenangaben berücksichtigt werden sollen. (Die pragmatische Lösung der Aufteilung - ähnlich einem westeuropäischen Namen - wirkt auf die muttersprachigen Nutzer befremdend, ist jedoch Mangels einer praktikablen

219 BibTeX formatiert „Johannes Paul II.“ in „li, Johannes P.“ um. Die Umkehrung der Reihenfolge wird ebenso wenig von Erfolg gekrönt sein: „Paul II. Johannes“ wird zu „Johannes, Paul II.“ Nur die Auszeichnung als nicht-formatierbares Feld führt zum gewünschten Resultat.

220 Vgl. Wikimedia Foundation Inc. (Hrsg.): *Arabischer Name*. URL: http://de.wikipedia.org/w/index.php?title=Arabischer_Name&oldid=47726855

Alternative De-facto-Standard.) Viele Namens-Bestandteile, die zwischen den Vor- und Nachnamen angegeben werden²²¹, sind ein sehr gutes Identifizierungsmerkmal und sollen daher optional im „*Infix*“ eingetragen werden. Für Titulaturen und akademische Grade, die weder nach ISO-690 noch nach DIN 1505-2 in den Quellenangaben anzugeben sind, aber dennoch eine zusätzliche Aussage über die Person treffen (und in CIP bzw. Titelblatt angegeben werden), kann man optional das Feld „*Prefix*“ verwenden. Dieses Feld hat bei der Generierung von Quellenangaben keine Verwendung. Die *Relation "Original"* erlaubt die Verlinkung zum ursprünglichen Namen hin. Eine Mehrfachangabe von Person in einem Werk ist zulässig.

Die im Text 17 angeführten Beispiele zeigen den Umgang mit außergewöhnlichen Personennamen – die Vorteile einer solchen Einordnung.

Würde man die Daten des Sir George Tryon auf englisch erfassen, wäre das Wort „Senior“ klein geschrieben. In der Familie Reuß, in welcher alle Söhne den Vornamen „Heinrich“ tragen, ist es unerlässlich die Nummerierung beizubehalten. Diese Angabe alleine ist hier nicht ausreichend. In jedem Jahrhundert beginnt die Zählung bei 1, so dass nur die Lebensdaten einen eindeutigen Hinweis geben können. Wie es in vielen Adelsgeschlechtern üblich ist, ist auch hier die Angabe der Linie in der Anmerkung für den Fachmann sehr hilfreich. In Quellenangaben sind keine Lebensdaten vorgesehen, lediglich "Verwandschaftsbezeichnungen"²²². Die genauen Angaben zu Linie und Geburtsjahr sind lediglich innerhalb der Datensammlung von Bedeutung und dienen den Menschen zur Orientierung.

221 Beispiele: "maddog" in John 'maddog' Hall, "Andrejewitsch " in Nikolai Andrejewitsch Rimski-Korsakow

222 An dieser Stelle ist eine freie Deutung des DIN 1505-2 unerlässlich. "Familiename mit sämtlichen zu ihm gehörenden Bestandteilen einschließlich Präfixe und Verwandschaftsbezeichnungen..." lässt offen, welche Bezeichnungen damit gemeint sind. Es ist anzunehmen, dass die Linie oder die Angaben zu Eltern als solche anzusehen sind.

```

<Person ID='m0001' xml:lang='de'>
  <Prefix>Sir</Prefix>
  <Vorname>George</Vorname>
  <Nachname>Tryon</Nachname>
  <Suffix>Senior</Suffix>
</Person>
<Person ID='m0002'>
  <Prefix>Fürst</Prefix>
  <Vorname>Heinrich</Vorname>
  <Nachname>Reuß</Nachname>
  <Suffix>LXII.</Suffix>
  <Anmerkungen>1785-1854, jüngere Linie</Anmerkungen>
</Person>
<Person ID='m0003'>
  <Prefix>Freiherr</Prefix>
  <Vorname>Clemens</Vorname>
  <Vorname>August</Vorname>
  <Nachname>Heereman von Zuydwyck</Nachname>
</Person>
<Person ID='m0004' xml:lang='de'>
  <Prefix>Professor</Prefix>
  <Vorname>Daoud</Vorname>
  <Vorname>Salman</Vorname>
  <Nachname>Anad</Nachname>
</Person>
<Person ID='m0005'>
  <Vorname>Joseph</Vorname>
  <Nachname>Ratzinger</Nachname>
</Person>
<Person ID='m0006' xml:lang='de'>
  <Relation Typ='Original'>m0005</Relation>
  <Rufname>Benedikt XVI.</Rufname>
</Person>
<Person ID='m0006' xml:lang='lat'>
  <Relation Typ='Original'>m0005</Relation>
  <Rufname>Benedict XVI.</Rufname>
</Person>
<Person ID="m0007">
  <Vorname>John</Vorname>
  <Infix>'maddog'</Infix>
  <Nachname>Hall</Nachname>
</Person>
Text 17: Beispiele für mehrgliedrige Namen erfasst als Objekte der Klasse Person

```

Im dritten Beispiel ist das Wort „Heerman“ keine Entsprechung des deutschen Vornamens „Hermann“ sondern ein Bestandteil des Nachnamens. Dies wird auch im vollständigen Namen „Clemens August Freiherr Heerman

von Zuydwyck“ sichtbar: der Adelstitel steht zwischen den Vornamen „Clemens August“ und „Heermann von Zuydwyck“. Im vierten Beispiel kann man auf die übliche Erfassungsweise zurückgreifen und „Daoud“ (zu deutsch „David“) sowie „Salman“ (de. „segnen“ / „Friede“) in der deutschen Erfassungsweise (*lang*="de") als Vornamen betrachten, da sie so benutzt werden. Im arabischen würde die Zuordnung anders gelöst werden. Durch eine Verweisung auf das Original kann dieses Problem umgangen werden. Bei päpstlichen Schreiben wie der Enzyklika „Deus caritas est“ wird immer der Name des Papstes genannt (Angaben zum Autor im Buch). Da sein bürgerlicher Name bekannt ist, kann der *Rufname* in Verbindung zu diesem gesetzt werden, indem man auf das „*Original*“ verweist.

5.3.2.2. Organisation

Die Klasse „Organisation“ ist nach Ansicht des Autors ein oft unterschätzter Bestandteil einer bibliographischen Beschreibung. In den meisten Fällen wird sie nur erwähnt, wenn sie als Herausgeber fungiert. Fehlt der Autor – z. B. weil er unbekannt ist oder eine nicht näher beschreibbare Personengruppe als Urheber gilt (Communities wie Wikipedia oder technische Standards wie HTML von W3C) – wird meist die herausgebende oder veranlassende Körperschaft genannt. (Vgl. DIN 1505-2 § 3.2)

Da der Begriff „Körperschaft“ im Deutschen nur in zwei Bereichen üblich ist (im juristischen Bereich und in den Bibliotheken), soll bewusst das Wort „*Organisation*“ verwendet werden. Dies hätte den Vorteil, dass es aufgrund seiner allgemeinen Ausprägung wohl eher als Sammelbegriff für Angaben zu Firmen, Behörden, Vereinen oder Hochschulen verstanden wird. Gleichzeitig soll es dem Sinn des Wortes „Körperschaft“ im bibliographischen Bereich entsprechen und ihn erweitern.

Die Organisation muss nicht unbedingt als Herausgeber tätig sein, um einen gewissen Einfluss auf das Werk zu haben bzw. daran indirekt beteiligt zu sein. Diese Situation wurde bereits in der Entwicklung von MAB berücksichtigt: das Feld 103 in MAB („Körperschaft, bei der die 1. Person beschäftigt“) bzw. das Feld 817 in MAB2 („Körperschaft, zu der eine Person

in Beziehung steht“) können solche Zusatzinformationen aufnehmen. In Zeiten zunehmender Lobbyarbeit kann es nach Ansicht des Verfassers für einen Entscheidungsträger von Bedeutung sein, wer eine Veröffentlichung veranlasst, durchgeführt oder bezahlt hat.

Die Klasse „*Organisation*“ kommt im wissenschaftlichen Bereich meist nur als Herausgeber oder als Hochschule, an der eine Diplomarbeit, Magisterarbeit, Dissertation oder Habilitation eingereicht wurde, vor. Ist eine Dissertation oder Diplomarbeit nicht in Form eines Buches erschienen, ist die Angabe der Hochschulbezeichnung und des Hochschulortes laut DIN 1505-2 obligatorisch, weil zum Auffinden eines Exemplars einer solchen Arbeit zweckmäßig.²²³

Eine Organisation zeichnen vor allem zwei Eigenschaften aus: der Name und der Ort. Bei besonders großen Körperschaften ist die Angabe des Namens des Bereiches (wie „Fachbereich X“, „Institut für ABC“, „Arbeitsgruppe: XY“ oder „Abteilung: ABC“) nützlich. Daher soll „*Name*“ als obligatorisches Feld und „*Abteilung*“ sowie „*Ort*“ als optional gelten. Mit Hilfe der Relation „*fortführend*“ soll eine Verweisung zwischen Identitäten möglich sein, jedoch nur dann, wenn es sich um Namensänderung handelt. Die Relation „*gehört_zu*“ ermöglicht eine Zuordnung mehrerer untergeordneter Instanzen zur hierarchisch nächst-höheren Instanz. Die *ID* (optionales XML-Attribut) einer Organisation soll der besseren Nachvollziehbarkeit wegen mit dem Buchstaben „o“ beginnen. Ein *Identifikator* kann optional unter Erwähnung des Namens des externen Datenanbieters im Attribut "*Typ*" angegeben werden können. Im Attribut "*Funktion*" wird optional die Funktion der Organisation hinsichtlich der Entstehung des Werkes angegeben. Eine Mehrfachangabe von Organisation in einem Werk ist zulässig.

5.3.2.3. Verlag

Die Angaben zum Verlag sind aus der Sicht des Bibliotheksbenutzers nahezu irrelevant: Titel, Autor und ISBN bzw. ISSN reichen für ein Retrieval im OPAC der Universitätsbibliothek meistens aus. Will man ein solches Werk

²²³ Vgl. DIN 1505-2 §1 (*DIN-Taschenbuch-154*, S. 93) sowie § 6.2 (ebd. S. 99)

erwerben, ist diese Angabe sehr nützlich²²⁴ oder notwendig. Ein Verlag wird anhand des Namens und des Ortes identifiziert. Große internationale Verlagskonzerne geben in ihren Büchern eine ganze Reihe von Ortsnamen an. Dies macht wenig Sinn, da sich nur englischsprachige Werke in vielen Ländern dieser Welt gleichzeitig verkaufen lassen. In den meisten Fällen sind die Veröffentlichungen nur für einen begrenzten nationalen Markt vorgesehen. In den Quellenangaben wird die Angabe nur eines Ortes erwartet. Aus diesem Grund soll nur ein Ort dem Verlag zugeschrieben werden. Werden mehrere Orte angegeben, so wird der zur Sprache der Veröffentlichung passende Ort (Sitz des Verlag-Unternehmens) genannt bzw. im Zweifelsfall der erste oder der älteste. Aufgrund der begrenzten Bedeutung einer solchen Angabe, wird auf eine Verlinkung zwischen den unterschiedlichen Instanzen eines Verlages verzichtet. Die Klasse *Verlag* umfasst zwei XML-Elemente (*Verlagsname* und *Ort*) sowie das Attribut *ID* (dessen Wert mit einem „v“ beginnen sollte), welches die Referenzierung ermöglicht. Eine Mehrfachangabe von Verlag in einem Werk ist nicht zulässig.

5.3.2.4. Abstract

Der englische Begriff „Abstract“ hat sich im „wissenschaftlichen Deutsch“ als Oberbegriff für viele unterschiedliche beschreibende Textarten „eingebürgert“. Die deutsche Entsprechung „das Kurzreferat“, welches in der DIN-Norm 1426²²⁵ von 1988 näher spezifiziert wurde, ist in der Praxis nicht gebräuchlich. Aus diesem Grund wird hier der englische Begriff synonym zu „Kurzreferat“ verwendet.

Die DIN-Norm sieht eine kurze und klare Inhaltswiedergabe²²⁶ als ein Kurzreferat an. Sie unterscheidet acht Typen, unter welchen vor allem die Zusammenfassung, die Annotation, das eigentliche Kurzreferat und eine

224 DIN 1505-2, § 3.8 (*DIN-Taschenbuch-154*, S. 95)

225 Norm DIN 1426 (Inhaltsangaben von Dokumenten. Kurzreferate, Literaturberichte. Deutsche Norm. 1988. Berlin: Beuth Verlag)

226 Nach: Kuhlen, Reiner: *Informationsaufbereitung III: Referieren (Abstracts – Abstracting – Grundlagen)*. In: *KSS*, S. 189

Rezension zu finden sind. Die Klasse „*Abstract*“ soll alle diese Typen erlauben, ohne sie genau zu spezifizieren. Diese Texte können für eine maschinelle Auswertung anhand Algorithmen oder intellektuelle Einordnung der Relevanz eines Werkes verwendet werden.

In den Graduierungsschriften werden Abstracts vom Autor selbst geschrieben. In anderen Fällen werden sie vom Bibliothekar oder Dokumentar im Rahmen der formalen Erfassung bzw. der inhaltlichen Erschließung angereichert. In einer pluralistischen Welt des User-Generated-Content wäre eine solchen anonyme Beschreibung nicht sehr dienlich – zumal es sich in den meisten Fällen um eine Interpretation handelt²²⁷. Es wäre angebracht, neben dem (an Länge unbegrenzten) Text des Abstracts auch den Namen des Verfassers zu verzeichnen. Da die Wissenschaftler sowohl Konsumenten als auch Produzenten von Veröffentlichungen sind, liegt es nahe, die Objekte der Klasse „*Person*“ als Autoren zu verwenden.

Die Klasse „*Abstract*“ besteht somit aus zwei Feldern: „*Text*“ und „*Autor*“ (optional), wobei „*Autor*“ eine Instanz der Klasse „*Person*“ oder eine Referenz auf diese sein darf. Die Relation zwischen dem referenzierten Werk (Serie, Buch, Periodikum, Artikel) und dem Abstract wird – sofern Abstract nicht direkt im zu beschreibenden Objekt inkludiert – mit Hilfe der Relation „*betrifft*“ unter Angabe der ID des bibliographischen Werkes realisiert. Die Sprache des Kurzreferats kann optional durch die Verwendung des XML-eigenen Attributes „*lang*“ angegeben werden. Mehrfachangabe von Abstracts zu oder in einem Werk ist zulässig.

5.3.3. Besondere Datentypen

Einige Eigenschaften, die einer besseren Beschreibung oder Einordnung dienen, sind so vielseitig verwendbar und zugleich so besonders, dass sie genauer beschrieben werden müssen. Um diese besonderen Datentypen nicht der freien Interpretation des Anwenders zu überlassen, soll die Funktionsweise bzw. der zulässigen Inhalte angegeben werden.

227 Ebd.

5.3.3.1. Relation

Die Relation wurde bereits mehrfach verwendet. Grundsätzlich werden zwei Typen von Relationen unterschieden: die vertikalen (hierarchisch) und die horizontalen.

Die vertikalen Beziehungen können in Form von Inklusionen realisiert werden, indem das hierarchisch übergeordnete bibliographische Werk die untergeordneten Werke in sich „aufnimmt“²²⁸. Diese Vorgehensweise erfordert einen zusätzlichen Aufwand, da die Artikel genau dort erfasst werden müssen, wo das beinhaltende Buch ausgegeben wird. Das XML sieht bestimmte Datentypen vor, die eine ordnende Funktion haben und mehrere getrennt erfasste Elemente zusammenbringen können. Es sind *ID*, *IDREF* und *IDREFS*²²⁹. Wurden die einzelnen Objekte (ob Hilfs- oder Hauptklassen) mit dem Attribut „*ID*“ versehen, können sie von anderen Objekten unter Angabe des Wertes dieses Attributes „angesprochen“ werden. Die Frage, welche hierarchische Stufe auf welche verweisen soll (aufsteigend oder absteigend), kann hier mit „beliebig“ beantwortet werden. Wichtig ist, dass dies konsequent verwendet wird. Die Verweisung „von Unten nach Oben“ trägt jedoch eher der Tatsache Rechnung, dass die Zuordnung neuer Werke zu bestehenden Werken den Erfassungsvorgang effizienter macht, als das Erweitern der bereits erfassten Werke. Diese Relation wird in dieser Arbeit als „gehört zu“ bezeichnet. Da XML keine Leerzeichen in Attribut- und Element-Namen erlaubt²³⁰, soll es „gehört_zu“ geschrieben werden. „gehört_zu“ kann sowohl als XML-Attribut als auch als XML-Element realisiert werden (Beispiele: `<gehört_zu>s_123</gehört_zu>`, `<Buch gehört_zu='s_123'>...</Buch>`).

Zu den horizontalen Relationen gehören die Verweisungen zwischen dem Original und der Übersetzung, dem Vorgänger und dem Nachfolger oder zwischen Beschreibung und dem zu beschreibenden Objekt sowie die Äquivalenz.

228 Vgl. Beispiel im Text 16 auf Seite 165

229 Vgl. XML in a Nutshell. S. 46-48

230 Ebd. S. 18

Die „Äquivalenz“ drückt die Identität zweier Objekte mit der Einschränkung aus, dass die Seitenangaben und Relationen zu übergeordneten bibliographischen Einheiten nicht zu berücksichtigen sind. In der FRBR-Ontologie ausgedrückt, wäre dies eine Verbindung zwischen zwei Manifestationen desselben Werkes (z. B. veränderte Auflagen). Dieser Zustand trifft oft ein, wenn ein Artikel mit demselben Inhalt in zwei unterschiedlichen selbständig erschienenen Einheiten abgedruckt wird. Dies kann hilfreich bei der Suche nach einem Exemplar eines solchen Werkes sein: es zeigt die Alternative auf, wenn die angegebene Manifestation selbst nicht erreichbar ist. „Äquivalenz“ kann als XML-Attribut (IDREF oder IDREFS) oder als XML-Element (IDREF) realisiert werden.

Ähnlich verhält es sich im Falle von Übersetzungen. Hier wird jedoch nicht „bidirektional“ zwischen Manifestationen sondern immer zwischen den „Ausdrücken“ (FRBR: "expressions") unidirektional in die Richtung des Original verwiesen. Dies gleicht der Verlinkung zwischen Schreibweisen der Personennamen. Existiert mehr als eine Übersetzung eines Werkes, so ist es notwendig, das Original zu erfassen und dadurch eine Relation (durch Rückwärtsauflösung) zwischen den Übersetzungen zu ermöglichen. Da immer mehr Menschen zu Zwei- oder sogar Dreisprachigkeit erzogen und ausgebildet werden, kann eine „fremdsprachige Alternative“ durchaus interessant erscheinen. Mit der Relation „*Original*“ kann im einfachsten Falle auf das (nicht übersetzte) Original verwiesen werden. Diese Relation kann man in Form eines XML-Attributs oder eines XML-Elements umsetzen.

Eine weitere Relation ergibt sich aus der Zeitabfolge: Organisationen oder bibliographische Serien ändern ihren Namen, behalten jedoch die gleiche Zielsetzung. Verlässt man sich nur auf den Titel als das identifizierende Merkmal, kommt es unweigerlich zu Brüchen: Vor allem sehr einfache Formate wie RIS scheinen an dieser Stelle solchen Entwicklungen nichts entgegen setzen zu können. Ein weiterer Grund für die Einführung eines Fortsetzungsmerkmals ist die Aufteilung eines Werkes: Während bei Büchern mit demselben Titel und Bandangabe die Vermutung nahe liegt (für

Computer mit entsprechenden Algorithmen sogar eine "hohe Wahrscheinlichkeit"), es handle sich um eine Fortsetzung, ist die Situation bei Artikeln mit Angaben zum Teil im Titel oder Untertitel schon schwieriger. Es gibt Teil-Artikel, die in sich eine geschlossene Einheit ergeben. Andere sind nur mit dem Wissen aus dem vorherigen Teil verständlich. Die ersten können eine „virtuelle Serie“ bilden, weil hier die Reihenfolge keine Rolle spielt, die anderen setzen den zuvor erschienenen Teil voraus.

Diese beiden Aufgaben sollen mit Hilfe der Relation „*fortführend*“ realisiert werden können. Sie verweist auf das vorhergehende Objekt, welches fortgeführt wird. Es kann als XML-Attribut oder XML-Element realisiert werden. Wichtig ist zu beachten, dass diese Relation grundsätzlich nur zwischen zwei Objekten derselben Klasse möglich sind.

Die letzte Beziehung ist sehr spezifisch und kommt nur zwischen einem Abstract und einem bibliographischen Werk vor. Es drückt die Bezugnahme aus. „*betrifft*“ kann als XML-Attribut oder als XML-Element realisiert werden und bezieht sich immer auf eine bestimmte Manifestation eines Werkes.

"Äquivalenz" und "*betrifft*" können sich gleichzeitig auf mehrere Objekte beziehen und sind somit mehrfach wiederholbar. "*Original*", "*gehört_zu*" und "*fortführend*" verweisen stets auf nur ein Objekt und sind nicht wiederholbar.

5.3.3.2. Klassifikation, Schlagwort, Sprache

„Eine Klassifikation ist ein Ordnungsmittel, das eine Einteilung eines Gegenstandsbereiches mit Hilfe von Klasseneinteilungen leistet, wobei die – insbesondere hierarchische – Relation zwischen den Klassen ausgewiesen wird.“²³¹

Die Klassifikation ist der Definition nach ein Ordnungsmittel. Sie hilft bei der Einordnung von bibliographischen Werken zu Gegenstandsbereichen. Das Prinzip sorgt für ein relativ hohes Recall-Precision-Verhältnis bei relativ niedrigem Aufwand. Eine Klassifikation kann nur von Menschen

²³¹ *Klassifikation*. In: *KSS-Glossar*, S. 69

vorgenommen werden, die das Klassensystem und den Inhalt der zu beschreibenden Einheit gut kennen. Zu den wichtigsten Klassensystemen im bibliographischen Bereich gehören die DDC (Dewey Decimal Classification) und die Basisklassifikation. Im Patentwesen existiert eine eigene Internationale Patentklassifikation (IPC), in der Medizin sogar mehrere (ICD, ICPM), die in ihrem spezifischen Bereich bei der Einordnung der Literatur hilfreich sein können.

Die Klassifikation als ordnendes Zusatzmerkmal zu einem bibliographischen Werk kann als XML-Element realisiert werden. Die Angabe ist optional und muss mit Hilfe vom XML-Attribut „Typ“ um den üblichen Namen bzw. die übliche Abkürzung des Namens des Klassensystems erweitert werden. Ohne diese Typisierung des Klassensystems wäre dieses Merkmal wenig aussagekräftig. Ob Namen oder Abkürzungen und in welcher Sprache verwendet werden sollen, ist der Gruppe der Anwender überlassen. Englische Namen und Kürzel scheinen wegen der großen Verbreitung jedoch empfehlenswert.

„Ein Schlagwort ist eine Benennung, die einer dokumentarischen Bezugseinheit bei der Inhaltskennzeichnung zugeordnet wird. Dabei unterscheidet man das gebundene Schlagwort, das einer verbindlichen Liste entstammt, von einem freien Schlagwort. Unter einer Vorzugsbenennung in einem Thesaurus versteht man eine Benennung, die aus einer Äquivalenzklasse von Benennungen diejenige ist, deren Verwendung empfohlen oder vorgeschrieben wird.“²³²

Diese Definition beschreibt den Begriff „Schlagwort“ sehr allgemein und weist auf die in der Deutschen Nationalbibliothek übliche Praxis der Verweisung von Schlagwörtern auf Stichwörter hin. Im Alltag werden die Schlagwörter jedoch frei vergeben und zeigen die Assoziationen auf, die einem Benutzer eigen sind: aufgrund anderer Denkmuster, eines anderen Dialektes oder

232 Schlagwort. In: KSS-Glossar, S. 109

wegen eines anderen Forschungsbereiches. Während Badener das Wort „Münster“ verwenden, ziehen die Norddeutschen das Wort „Dom“ vor. Ein Theologe, der die Bischofskirche einer Diözese erwähnt, wird eher von der „Kathedrale“ sprechen. Diese drei Synonyme können während der Erfassung mit Hilfe einer Schlagwortnormdatei auf ein eindeutiges Schlagwort aufgelöst werden. Umgekehrt müsste dasselbe System die Anfragen (Suchbegriffe) in alle der Vorzugsbenennung entsprechenden Schlagworte umsetzen, bevor diese ausgeführt wird und bei Mehrdeutigkeit nachfragen, welcher Ober- oder Unterbegriff mit dem gesuchten Wort in Verbindung steht. Diese Vorgehensweise führt zum guten Recall-Precision-Verhältnis. Sie verlangt jedoch die Pflege eines Thesaurus. Es ist ein nicht zu unterschätzender Aufwand, der nur von Institutionen betrieben wird. Mit sog. „Tag-Wolken“, die im Internet immer öfter verwendet werden, kann sowohl ein Mensch als auch eine Maschine auf andere, verwandte²³³ Begriffe hingewiesen werden – darunter Synonyme²³⁴. Diese Technik geht so weit, dass sie mit statistischen Methoden freie Schlagworte gruppieren und sogar Relationen auszuwerten versucht.

Aus diesem Grund soll es nur das XML-Element „*Schlagwort*“ im Sinne eines freien Schlagwortes geben, welches mit einem XML-Attribut „*lang*“ („*xml:lang*“) optional einer Sprache zugeordnet werden kann. Die Angabe der Sprache soll der Tatsache Rechnung tragen, dass die englischen Begriffe in einigen Wissenschaften gleichzeitig mit anderen Sprachen verwendet werden. Das Fehlen einer solchen Auszeichnung kann bei umfangreichen Sammlungen zur Verwirrung führen: z. B. die englische „email“ (de. „E-Mail“) und das deutsche „Email“ (Schmelzüberzug).

233 Von „Verwandtschaft“ kann man hier nur unter Vorbehalt sprechen. Es handelt sich hier um einen losen Zusammenhang, der wegen bestimmter Quantität (des gemeinsamen Vorkommens) statistisch als wahrscheinliche inhaltliche Nähe gilt. Es basiert auf der Annahme, dass zwei Wörter umso „verwandter“ sind, je öfter sie gemeinsam vorkommen, welche jedoch nicht zwingend stimmen muss.

234 An dieser Stelle wird die auf der Seite 167 vorgestellte Idee der Persistenz und der Übertragung der Schlagwortnormdatei-Daten bewusst aus den genannten Gründen nur Ansatzweise realisiert.

Durch Vergabe von *Identifikatoren (ID)* und Nutzung von Relationen wie *"original"* (für die Verweisung auf die richtige bzw. die Original-Schreibweise), *"gehört_zu"* (für Verweisungen auf Oberbegriffe), *"äquivalent"* (für andere, gleichberechtigte Begriffe) oder *"fortführend"* (für den Ersatz einer nicht mehr gebräuchlichen Bezeichnung) können sogar kleine Thesauri gepflegt werden, indem die Schlagworte einander referenzieren (Attribut *"Referenz"*). Die bloße Angabe des Attributes *"Referenz"* ohne *"ID"* und *"Typ"* meint hingegen die Zuweisung des referenzierten Schlagwortes.

Die Unterscheidung der Sprache kann nicht nur bei Schlagwörtern sehr hilfreich sein. Bereits erwähnt wurden Übersetzungen von Werken oder Personennamen. Es ist ein sehr allgemeines Problem, weshalb der XML-Standard ein spezielles XML-Attribut definiert hat. Das „xml:lang“-Attribut sagt über den Inhalt des Elementes aus, in welcher Sprache dieser verfasst wurde. Als Wert dieses Attributes sind zugelassen²³⁵:

- Sprachcodes des ISO-639²³⁶
- Kombination aus ISO-639 Sprachcodes (klein geschrieben) und Subcodes der Länder aus ISO-3166 (groß geschrieben) verbunden durch ein Minus-Zeichen
- Sprachcodes für Kunstsprachen, die bei IANA²³⁷ registriert sind
- Demnächst wohl auch Kombination aus den oben genannten Codes und Varianten²³⁸

235 Angaben nach: XML in a Nutshell, S. 86f

236 Dieser kann unter <http://www.loc.gov/standards/iso639-2/php/langcodes-search.php> (Stand: 28.08.2008) gefunden werden. Es werden die ISO-639-1 (2 Buchstaben) sowie ISO-639-2 (3 Buchstaben) gezeigt. Die Suche nach „german“ ergibt einige besonders interessante Treffer wie: „Mittelhochdeutsch“, Niedersächsisch oder Schweizerdeutsch.

237 IANA (Internet Assigned Numbers Authority) ist die zentrale Stelle für die Registrierung von WWW-Weit eindeutigen Codes. Man kann auf Anfrage sogar fiktive sprachen registrieren. Zum Beispiel: <http://www.iana.org/assignments/lang-tags/i-klington> (Stand: 26.08.2008) für Klingonisch (die Sprache einiger Gestalten aus der Filmserie „Star-Treck“)

238 Der Standard „Tags for Identifying Languages“ befindet sich noch im Entwurfsstadium

- „x-“ oder „X-“ begleitet von eigenem Sprachcode, der noch nicht registriert ist (und frei vergeben werden kann)

Die Angaben zur Sprache sind optional. Betrifft eine Sammlung nur Werke in einer Sprache (des Benutzers), ist sie überflüssig. Ist dies nicht der Fall, kann diese Information für ihn und andere sehr nützlich sein.

5.3.3.3. Autor

In den bibliographischen Klassen wurde öfters der Autor erwähnt. Dieser kann nur eine Person (ein Mensch) sein. Mit „Autor“ erwähnt man eine ganz bestimmte Funktion einer Person, die an der Entstehung eines Werkes beteiligt war. In diesem Entstehungsprozess gibt es mehrere Funktionen, doch nur wenige sind so maßgebend, dass sie auf jeden Fall in den Quellenangaben erwähnt werden sollen: Der Autor, der Herausgeber, der Verlag und manchmal die Organisation (wie z. B. die Hochschule bei Diplom- oder Magisterarbeiten, Dissertationen und Habilitationen). Bis auf Sprach- (wie die Exegese = die Bibel-Theologie) und die Literaturwissenschaften ist der Übersetzer nicht von Bedeutung, weshalb er in diesem Entwurf kein eigenes Element erhält (kann jedoch mit seiner Funktion als Mitarbeiter genannt werden).

„*Autor*“ kann als XML-Element realisiert werden, welches ein gültiges XML-Attribut „*Referenz*“ (IDREF oder IDREFS) verlangt. Dieses Attribut ist eine Referenz auf die ID eines oder mehrerer Objekte der Klasse *Person*. Alternativ kann es ein Objekt der Klasse *Person* (inline) beinhalten. Objekte der Klasse *Autor* können in einem Werk mehrfach hintereinander erfasst werden, wobei die Reihenfolge der Angabe der Reihenfolge des Vorkommens in den Quellenangaben entspricht (1., 2. und 3. Autor müssen nicht dem 1., 2. und 3. Autor der alphabetischen Reihenfolge entsprechen, sondern können die Beteiligung widerspiegeln).

beim IETF und umfasst Codes, Subcodes und Varianten. Darunter drei Varianten der phonetischen Umschrift, darunter „fonipa“ (International Phonetic Alphabet). Dadurch wird eine phonetische Wiedergabe eines Namen möglich. Vgl. <http://www.iana.org/assignments/lang-subtags-templates/fonipa> (Stand: 26.08.2008) und <http://www.iana.org/assignments/language-subtag-registry> (Stand: 26.08.2008)

5.3.3.4. Herausgeber

Als Herausgeber kommen nur eine Person oder eine Organisation in Frage. Da diese Klassen bereits definiert wurden, muss der Datentyp nur auf Objekte dieser Klasse verweisen. „*Herausgeber*“ lässt sich als XML-Element realisieren, der das XML-Attribut „*Referenz*“ besitzen muss, welches eine Referenz zur ID eines Objekts vom Typ Person oder Organisation darstellt (IDREF). Alternativ ist es möglich, ein Objekt der Klasse „Organisation“ oder der Klasse „Person“ (inline) einzubinden. „*Herausgeber*“ darf in allen Werken verwendet werden, die einen solchen ausweisen. Mehrfachvorkommen ist erlaubt, die angegebene Reihenfolge muss – ähnlich dem Autor – eingehalten werden.

Autor und Herausgeber gehören zu den „beteiligten Personen“, und werden im DC unter dem Begriff „Contributor“ eingetragen. Ein Datentyp mit diesem Namen, der die beteiligten Personen zusammenfasst, hätte eine rein „ordnende“ Funktion und ist daher nicht notwendig. Interessant wäre sicher die Möglichkeit, weitere beteiligte Personen mit ihrer jeweiligen Funktion zu erfassen.

5.3.3.5. Mitarbeiter

Mit der Metaklasse „*Mitarbeiter*“ sollen Personen erfasst werden können, die ein Werk weder herausgeben noch zu den Autoren zählen, aber an der Entstehung beteiligt waren. Mit dem Attribut „*Referenz*“ wird auf die ID eines Objektes der Klasse Person verwiesen. Das „*Typ*“-Attribut kann eine in einer Sprache übliche Funktionsbezeichnung haben. Damit diese Angabe international verständlich bleibt, muss man mit dem XML-eigenen „*lang*“-Attribut entsprechende Angabe zur verwendeten Sprache machen. Eine Referenzliste kann für eine Übersetzung auf Computer-Ebene sorgen. Diese Klasse wird aus Gründen der DIN-1505-2-Kompatibilität geführt, worin die Angabe aller beteiligten Personen (mit ihren Funktionen) vorgeschrieben ist.

Die Metaklasse „*Mitarbeiter*“ ist mehrfach wiederholbar und kann alternativ ein Objekt der Klasse Person direkt (inline) erfassen. Die Reihenfolge der Erfassung entspricht der Reihenfolge der Nennung in den Quellenangaben.

5.4. Maschinelle Verarbeitung

Einer der Ausgangspunkte dieser Arbeit war die vielfältige Verwendungsmöglichkeit der bibliographischen Metadaten. Fast überall, wo sie verarbeitet werden, kommt ein Computer zum Einsatz. Dieser kann Zusammenhänge „erahnen“ oder Alternativen anbieten, wenn man ein bestimmtes Werk sucht. Studierende und Wissenschaftler verlassen sich ebenso auf die „Urteilkraft“ einer Maschine, wenn es darum geht, Quellen anzugeben (bei Zitaten) oder sie aufzulisten (Bibliographien). Dabei kann diese nur mit Hilfe einiger Algorithmen Daten einordnen und verändern, die als zugehörig zu einem bestimmten Typ ausgezeichnet sind.

Der Autor ist der Überzeugung, dass XML diesem Zweck sehr gut geeignet ist. Es stellt eine einfache und dennoch universelle Syntax zur Verfügung. Die Aussagen, die mit XML getroffen werden (die sog. XML-Anwendungen), machen sich diese Regeln zunutze und verwenden einen eigenen Wortschatz (sog. Dialekt), der die Inhalte beschreibt und typisiert. Die Vollständigkeit sowie die Granularität (die Genauigkeit und die Anzahl) der XML-Elemente und XML-Attribute sind für eine gute Verarbeitung zu Quellenangaben ebenso wie für ein gutes Retrieval im Datenbestand nach Ansicht des Verfassers entscheidend. Da es sich bei diesem Entwurf um eine völlige Neuentwicklung handelt, kann die Datenstruktur frei modelliert werden: Vornamen – wie viele es auch sein mögen – sind klar von Nachnamen (wie „Adam“) zu unterscheiden. Artikel werden den Büchern „gleichberechtigt“ und dürfen deshalb viele Metadaten enthalten. Schlagworte oder andere Schreibweisen eines Autorennamens werden mit explizit genannt, statt auf dem Server (für den Benutzer „unsichtbar“) zu verbleiben. Beim Austausch gehen die erfassten Daten und Zusammenhänge nicht verloren.

Die maschinelle Verarbeitung gestaltet sich nach den im XML-Standard definierten Regeln. Ein Programm, welches eine Datei in dem hier erarbeiteten Format öffnen will, wird sie zunächst auf XML-Syntax-Konformität überprüft. Ist das Dokument nicht wohlgeformt, weil ein

Anführungszeichen bei einem Attribut fehlt oder ein XML-Element nicht geschlossen wurde, wird der Benutzer auf die vermutlich fehlerhafte Stelle im Dokument hingewiesen und kann mit einem einfachen Texteditor Korrekturen vornehmen. Während der Validierung wird überprüft, ob alle Elemente und Attribute der Anzahl, Reihenfolge und Verwendungsort nach der Definition entsprechen. Dies setzt eine DTD oder ein XML-Schema unter der im XML-Datei-Header bzw. Namensraum angegebenen URL-Adresse voraus. Dabei werden Abhängigkeiten aufgelöst. Im dritten Schritt werden die darin enthaltenen Daten auf eine innere Darstellung abgebildet, wobei diese wohl sehr ähnlich der XML-Darstellung sein wird.

5.4.1. Auflösung von Abhängigkeiten

Durch die Verwendung von internen Identifikatoren ist es möglich, ein Objekt an mehreren Stellen zu referenzieren. So können Artikel an einem Ort in der Datei erfasst werden, während die verweisenden Bücher an einem anderen Ort gesammelt werden können. Will der Benutzer von Literaturverwaltungssoftware die Daten aus einer solchen Datei importieren, so muss das Programm die Zusammenhänge auswerten und auf die intern verwendete Struktur abbilden. Dies ist der einfachste Fall der Abhängigkeitsbehandlung (z. B: Artikel „a_1“ setzt Buch „b_1“ voraus). Während der Arbeit mit diesem Programm darf der Fall nicht eintreten, dass das Buch „b_1“ gelöscht wird und der darin enthaltene Aufsatz „a_1“ erhalten bleibt, da sonst für das Wiederfinden wichtige Angaben verloren gehen würden. Viele Angaben, die vom Buch "geerbt" werden, wären nicht auffindbar. Eine ins Leere laufende Referenz (wie hier vom Artikel auf das Buch) widerspricht der Validität einer solchen XML-Datei. Dieser inkonsistente Zustand der Daten muss vom verarbeitendem Programm verhindert werden.

Schwieriger zu behandeln sind die Abhängigkeiten, die durch die Relationen „fortführend“, „Äquivalenz“ oder „Original“ realisiert werden. Sie verweisen auf andere Objekte, die nicht notwendiger Weise für den Benutzer von Interesse sein müssen. „Benedikt XVI.“ ist als Name des Papstes und des

Autors der Enzyklika „Deus caritas est“ für einen deutschen Forscher völlig ausreichend, obwohl es auf das Original „Benedict XVI“ verweist. Es sind auch nicht alle Äquivalenzen zum zitierten Werk von Belang. Deshalb ist es angebracht, beim Benutzer nachzufragen, ob diese Relationen beachtet werden sollen. Da die referenzierten Werke ebenfalls referenzieren dürfen (ad infinitum), ist die Frage nach der sog. Tiefe zu stellen d. h.: ab welcher „Entfernung“ von dem zu exportierendem Objekt seine Relationen nicht mehr beachtet werden sollen. Bei Äquivalenzen kann ebenfalls die Zahl der Alternativen begrenzt werden. Solche Fragen sind im Prozess des Exports bibliographischer Daten in dieses Format bzw. während des Imports dieser Daten dem Benutzer zu stellen.

Das Exportieren von Daten aus thesauriartigen Datenbanken wie SWD oder PND als gerichtete Relation von Objekten (auf das „Original“ hin) hat den Vorteil, dass der Datenempfänger sowohl den verwendeten Suchbegriff (Personenname, Schlagwort) als auch den von der Datenquelle zugeordneten Ober- bzw. Unter-Begriff oder Personennamen erfährt. Der Benutzer weiß also, warum die Suche nach „Benedikt XVI.“ zu Ergebnissen mit „Joseph Ratzinger“ als Autor geführt hat. **Der Server und der Client sind diesbezüglich auf dem gleichen Wissensstand. Keine relevante Information ist während der Übertragung verloren gegangen.**

Würde man die vorgestellte Idee der persistenten Übertragung von Thesauri umsetzen wollen, müsste bei der Auflösung von Abhängigkeiten auf der Seite des Benutzers dafür gesorgt werden, dass Begriffe einer Ontologie (oder eines Thesaurus) nicht mit den Daten einer anderen Datenquelle vermischt werden²³⁹, da die meisten Ontologien fachspezifisch sind und dieselben Begriffe in unterschiedlicher Ausprägung benutzen können²⁴⁰. Nur

239 Zum Problem der Heterogenität von Ontologien und möglichen Lösungsansätzen vgl. Krause, Jürgen : *Shell Model, Semantic Web ans Web Information Retrieval*. In: *Festschrift Zimmermann*. S. 99f

240 vgl. Panyr, Jiri : *Thesauri, Semantische Netze, Frames, Topic Maps, Taxonomien, Ontologien - begriffliche Verwirrung oder konzeptionelle Vielfalt?* In: *Festschrift Zimmermann*. S. 147. (Es gibt keine allgemein gültige Ontologie. Vielmehr sind es

der menschliche Intellekt ist gegenwärtig in der Lage, solche Zweige aus einem Thesaurus in ein anderes Thesaurus konsistent einzupflegen.

Beim Datenimport soll ebenfalls verhindert werden, dass bereits existierende Objekte mit den zu importierten Objekten (derselben ID) automatisch überschrieben werden, da sonst die Auflösung von Abhängigkeiten nur scheinbar gelingt. In Wirklichkeit werden die internen Identifikatoren richtig verarbeitet aber mit falschen Daten verknüpft (konfuses Ergebnis). Daher ist es empfehlenswert, die Regeln für die Vergabe von Identifikatoren zu erweitern, so dass jedes Mitglied einer Gruppe ein eigenes Merkmal (als Daten-Anbieter-Merkmal oder Namensraum) in der ID der Objekte mitführt. Diese Vorgehensweise beugt der Gefahr der Überschneidung von Identitätsmerkmalen vor. Gleichzeitig ist es möglich, die fremden Daten (nach dem Export und Import) lokal zu korrigieren oder zu ergänzen und weiter zu verwenden. Es sollte jedoch möglich sein, (auf Wunsch) die „eigenen“ Objekte mit solchen korrigierten Daten zu überschreiben (sog. Reimport). Auf diese Art und Weise kann der ursprüngliche Besitzer (primärer Daten-Anbieter) in den Genuss der kollektiven Intelligenz kommen.

5.4.2. Normgerechte Einordnung

Die deutsche Norm DIN 1505 sieht in Teilen 2 und 3 besondere Regeln für die Einordnung von bibliographischen Werken vor. So verlangt DIN 1505-3 die alphanumerische Einordnung von Literatur anhand der Einordnungsformel. Eine solche Einordnungsformel besteht aus dem / den Autorennamen (wenn nötig: Ordnungs- oder Jahreszahl) bzw. einem markantem Wort des Titels und dem Erscheinungsjahr. Sie entspricht im großen und ganzem dem englischen Begriff „cite-key“²⁴¹. Die Einordnungsformel wird ebenso wie ein Zitierschlüssel anstelle der

einzelne Domänenontologien, die in einer Ontologie-Bibliothek dank einer Konkordanz koexistieren können.)

241 Der englischsprachige Begriff "cite-key" oder ausgeschrieben "citation key" kann ins Deutsche mit "Zitierschlüssel" übersetzt werden. In DIN 1505-3 wird er "Einordnungsformel" genannt. Bei wissenschaftlichen Journalen wird das in der Wissenschaftlergemeinschaft übliche Journal-Kürzel als Zitierschlüssel verwendet.

vollständigen Angabe zum Werk (z. B. bei weiteren Vorkommen des Werkes) verwendet. Das Erstellen einer solchen Formel ist bei Verwendung der genauen Angaben zum Namen des Autors oder des Herausgebers (enthalten in „Artikel“ bzw. „Periodikum“, „Buch“) sowie des Jahrgangs / Erscheinungsjahres („Periodikum“ bzw. „Buch“) sehr einfach und kann jedem beliebigen Format (ein oder mehrere Autoren, Nachnamen oder Vor- und Zunamen) angepasst werden. Die alphanumerische Einordnung kann sogar automatisch mit Hilfe von XSL-T für die in RFC-1766 erfassten Sprachen vorgenommen werden²⁴².

DIN 1505-2 sieht entgegen der üblichen Praxis in der (englischsprachigen) Wissenschaftswelt sowie den RAK-WB-Regeln vor, dass alle Autoren genannt werden. Da alle Autoren erfasst werden können, werden auch alle zitiert. Zitiert man nicht nach den DIN-Regeln, so kann die gewünschte Anzahl an Autoren nach ihrem Vorkommen in der Datei angegeben werden. Werden sehr viele Autoren genannt, kann das Programm eingreifen und (entsprechend einer vom Benutzer anzugebenden Zahl) die überzähligen Autoren (zusammen mit den Mitarbeitern) an der Position für „weitere beteiligten Personen“ einfügen, um die Quellenangaben übersichtlich zu gestalten.

5.4.3. Normgerechte Formatierung

Der Zusammenhang zwischen der Granularität der Daten (Genauigkeit der Unterteilung, Anzahl der Datenfelder und -Typen) und der Eignung als Datenbasis für Quellenangaben wurde bereits dargestellt. Je nach verwendeter Norm (DIN, ISO oder eine für ein bestimmtes Journal typische) kann die Formatierung sehr hohe Granularität der Datenbasis voraussetzen. Als eine solche ist die DIN 1505-2 anzusehen: Personen werden mit dem vollständigen Nachnamen (einschließlich Präfixe und Verwandtschaftsbezeichnungen) begleitet von einem Komma, dem ausgeschriebenen ersten Vornamen sowie abgekürzten weiteren Vornamen

²⁴² XML in a Nutshell, S. 497

aber ohne Adelstitel und Titulaturen angegeben²⁴³. Besteht ein Name aus einem oder mehreren Teilen, die nicht als Vor- und Zunamen gelten oder als solche nicht erkannt werden können, werden sie ohne Formatierung übernommen²⁴⁴.

Damit diese Teilnorm umgesetzt werden kann, ist es notwendig, dass ein Programm weiß, ob eine Zeichenkette einen nicht trennbaren Namen bzw. welchen Namensbestandteil (Typ und Reihenfolge) darstellt. Nur wenige Formate können diese Forderung erfüllen: DocBook und MODS (beide geben die Reihenfolge der Namen nicht an). In dieser Arbeit war es sehr wichtig, solche Probleme aufzuzeigen und hinreichend zu berücksichtigen. Aus diesem Grund werden die Personennamen mit Blick auf DIN 1505 entweder als untrennbarer "Rufname" ausgezeichnet oder Teil für Teil, entsprechend Ihrem Typ angegeben. Vornamen werden ihrem Vorkommen nach eingeordnet. (Dies entspricht der XML-Definition.)

Mit Hilfe von Stil-Vorlagen kann ein Programm auf Wunsch eine beliebige Norm sehr genau umsetzen. Der Vorteil von XML ist, dass eine solche Vorlage in Form von XSL-T-Datei vorliegen kann und deshalb sogar von aktuellen Internet-Browsern (die den sog. XSL-Prozessor besitzen) verarbeitet werden kann. Die XML-Daten können auf diese Art und Weise wie eine Bibliographie dargestellt werden. (Trennung von Datenbasis und deren Darstellung.)

Durch die Verwendung von Relationen sind einige Normen umsetzbar, die von DIN 1505-2 vorgeschrieben werden, aber bisher nicht realisiert werden konnten. Dazu gehört: Die Angabe des Serientitels bzw. Gesamtitelausgabe, die hierarchische Einordnung von Abteilungen eines Unternehmens (als Herausgeber) oder die vollen Angaben zum Fachbereich der Hochschule als Graduierungsort. „gehört_zu“ als Relation erlaubt die Bildung von Ketten mit hierarchisch aufsteigender Wertigkeit, wobei jedes

243 DIN 1505-2, § 3.1

244 ebd.

Glied in diesem ebenso wie ohne diesen Kontext an mehreren Orten der Bibliographie verwendet werden kann.

Mit „*Kürzel*“ steht ein Feld zur Verfügung, welches vorwiegend in den (englischsprachigen) wissenschaftlichen Journalen – aber nicht in der DIN-Norm – verwendet werden. Diese Praxis ist aus diesem Umfeld nicht wegzudenken. Diese Angabe wird nur selten von elektronischen Katalogen der Universitätsbibliotheken ausgeliefert. Beim PubMed wird diese Angabe bei den meisten Fundstellen genannt (bei fast allen Zeitschriftenartikeln).

Diese Arbeit orientiert sich insgesamt am Verwendungskontext der automatischen Generierung von Quellenangaben und bringt sehr viele passende Datenfelder mit (z. B. „Auflage“ als Zahl und „Edition“ als kurze Beschreibung). Angaben, die für das Zitieren keine Relevanz haben (wie "Monographie", "Lehrbuch" oder "Technischer Report") werden nicht als eigene Datenfelder berücksichtigt, können aber in Form von Beschreibung im Abstract aufgenommen werden. Gesetzestexte oder Bibelstellen kommen zwar als Quellen in einigen Forschungsbereichen vor, werden jedoch aufgrund ihrer inneren Textstruktur oft so behandelt, als hätten sie keine Ausgabe (im Sinne eines Exemplars). Die eigenen Regeln zur Bildung von Kürzeln und zu den Angaben von Stellen gehen über die Ansätze dieser Arbeit hinaus.

6. Abschluss

Diese Arbeit hat sich die effizientere Nutzung bibliographischer Metadaten zum Ziel gesetzt. Das dabei entstandene Format gilt dabei als Nachweis für die Machbarkeit in der Praxis. Der vorliegende Entwurf zeigt auf, wie die unterschiedlichen Verwendungskontexte der bibliographischen Metadaten aufeinander abgestimmt werden können. Im Gegensatz zu vielen neuen XML-basierenden Formaten handelt es sich hier um eine Neuentwicklung, die nicht einem einzigen Zweck dienen soll (Zitieren oder Titelmeldung). Die zunehmende Menge an wissenschaftlicher Literatur im Web sollte gleichberechtigt behandelt und ihre Eigenart berücksichtigt werden. Deshalb darf und soll eine solche Neuentwicklung die bestehenden Normen, Standards und Formate kritisch hinterfragen, um jeder der Publikationsformen und jedem Verwendungskontext gerecht zu werden. Die Einsatzgebiete sind: Recherche und Austausch, persönliche Literaturverwaltung sowie die möglichst automatisierte Generierung von Quellenangaben. Die Datenfelder wurden in Form eines XML-Formats beschrieben, können jedoch ebenfalls in Form einer Datenbank o.Ä. realisiert werden.

Zurecht darf man fragen, wie gut oder schlecht dieses Datenraster (weniger das XML-Format selbst) diese Ziele zu realisieren vermag. Dieser kritischen Untersuchung widmet sich der folgende Teil dieser Arbeit.

6.1. Vorteile des entwickelten Formats

Der Entwurf erfasst grundsätzlich nur die Grunddaten einer bibliographischen Beschreibung. Im Gegensatz zu vielen anderen XML-basierten Auszeichnungssprachen stellt sie eine Neuentwicklung dar und ist darauf bedacht, die zu übermittelnden Daten ihrem Typ und ihrer Beschaffung nach möglichst genau zu erfassen. (Dies geschieht vorwiegend auf syntaktischer Ebene.) Die Personennamen werden nicht vorformatiert aufgenommen (wie es in den meisten Bibliothekssystemen und bibliothekarischen Austauschformaten der Fall ist) sondern in einzelne Bestandteile zerlegt. Damit eröffnet sich die Möglichkeit einer präzisen Suche nach Vornamen, Nachnamen, alternativen Schreibweisen von Namen, Titeln der Zeitschriftenhefte oder der neu eingeführten ISSN-L. Gegenüber der bloßen Trunkierung (die zusätzlich benutzt werden kann) ist es dem Benutzer möglich, genauer zu kontrollieren, welche Felder durchsucht werden (ist „Adam“ der Vor- oder Nachname?). Damit **sinkt zwar der Recall, gleichzeitig steigt die Präzision**. Obwohl die Suche als solche nur marginal ein Übertragungsformat betrifft, ist sie neben dem "Browsen" eine der wichtigsten Methoden, die zum Abruf von Metadaten eines Werkes führen. Deshalb muss diese Art und Weise der Verwendung der bibliographischen Daten auch hier berücksichtigt werden. Eine Umsetzung in Form von XML-Format hat den Vorteil, dass die Datei lokal (mit Hilfe von XPath) durchsucht werden kann. XPath gehört zu den weit verbreiteten XML-Technologien, die universell einsetzbar sind. Die Suche mit XPath ist zwar nicht so performant wie die in binären Datenbanken, steht jedoch auf den meisten PCs als Funktionsbibliothek (oder als Funktion im Webbrowser) zur Verfügung. Ein Search-Scan-Verfahren (wie z. B. Z39.50) kann aufgetrennt werden, indem alle Fundstellen zum Benutzer übertragen werden, wo weitere Suchvorgänge ausgeführt werden können.

An vielen Stellen in dieser Arbeit wurde die "Maschinenlesbarkeit" beachtet: Der Computer soll „verstehen“ können, was eine Eingabe bedeutet d. h. wo und wie sie verwendet werden kann. Dies ist nach Ansicht des Verfassers für die richtige Zuordnung und Verwendung in Quellenangaben erforderlich. Vorformatierte Personennamen sowie „Titel-Doppelpunkt-Untertitel“-Angaben sind missverständlich und werden in einigen Programmen (wie bereits dargestellt) falsch interpretiert. Die Kenntnis der Syntax vorausgesetzt, kann eine Maschine einzelne Bestandteile "erahnen". Die semantische Analyse kann die Erkennungsgenauigkeit verbessern. Dennoch sind solche Daten nicht immer für die Generierung von Quellenangaben in einem anderen Format geeignet. Diese Arbeit unterscheidet zwischen vielen Typen von Eingaben (Feldern) und ist dadurch besser lesbar. Die einzelnen Metadaten-Elemente können präziser eingesetzt werden. Sie sind unabhängiger von der zu verwendenden Zitiernorm. XML lässt eine Transformation der Daten zu, die Betriebssystem-unabhängig (universal und populär) ist und sogar im Internet-Browser stattfinden kann. Die Ausgangsdaten (Datenbasis) werden **auf Wunsch zu einer bestimmten Art von Quellenangaben umgestaltet** (Darstellung) – können aber auch in Rohform für eine weitere Verwendung verarbeitet werden. Die Idee der Quellenangabe als "für das Beschaffen bzw. Wiederfinden notwendiger Angaben zum Werk" wird durch die Verwendung von DOI, URI bzw. URN um den Aspekt der elektronischen Identität von Werken erweitert.

Die Eingaben mit hoher Granularität tragen zur höheren Präzision bei der Suche und einer Zitiernorm-Unabhängigkeit bei. Die einmal erfassten Daten gehen bei der Übertragung nicht verloren: Schlagworte, Notizen, Abstracts, Klassifikation, Alternativen etc werden in diesem Austauschformat zuverlässig vom Informationsanbieter (Server) zum Kunden (Client) transportiert. Damit werden beide Teilnehmer einer solchen Kommunikation bezüglich eines Werkes **auf den gleichen „Wissensstand“** gebracht. Je nach Konfiguration können Abhängigkeiten oder „Verwandschaften“ zwischen den Werken (ähnliche Werke) beliebig tief aufgelöst werden. Die

Zuordnung zum Gesamtwerk bietet eine gute Möglichkeit für eine spätere Suche nach verwandten Artikeln oder Büchern.

Die atomare Beschaffenheit der Datenfelder beugt einer ungenauen oder falschen Verwendung vor. Im Falle von Büchern und Periodika ist die Kenntnis einiger Grundregeln unerlässlich. In den meisten Fällen sind **die Namen der Datenfelder selbsterklärend und eindeutig**. Relationen bieten die Möglichkeit einer thesauriartigen Zuordnung der Daten. **Weitere benötigte Felder können bei einer Umsetzung in Form von XML durch einen weiteren Namensraum hinzugefügt werden**. Damit ist das Problem der gleichberechtigten Verwendung des in dieser Arbeit entworfenen Formates mit einem anderen XML-atrigen Format gut lösbar. Sowohl die parallele Verwendung (einfaches Format des Benutzers und detailreiches Format des Bibliothekars) als auch **die Einbettung in andere Formate** (als "bibliographische Datenbank" in einem Textdokument) sind bei der Verwendung von XML möglich.

Die Vermeidung von Redundanzen und eine strikte objektrelationale Ausrichtung sorgen für eine Minimierung des Datenumfangs. Bei einer Realisierung in Form von XML fällt dies vor allem durch die häufige Verwendung der "*ID*" und "*Referenz*"-Attribute auf. Die horizontalen Relationen können sogar für Menschen nachvollziehbar abgelegt werden, indem z. B. eine Serie die dazugehörigen Bücher beinhaltet (wortwörtlich: „umfasst“). Die Verwendung von internen Identifikatoren erlaubt die innere Vernetzung verschiedener Schreibweisen, die in anderen bibliographischen Formaten nicht verlinkt werden. Unabhängig von der Schreibweise (Inklusion oder Verweisung) kann ein Computerprogramm die Daten aus der einfachen bibliographischen Auszeichnungssprache dank ihrer Objektorientierung sehr einfach und effizient in die intern verwendete Datenstrukturen übersetzen.

6.2. Schwachstellen des neuen Formats

Das in dieser Arbeit erstellte Datenraster ist auf die Erfassung und Verarbeitung der zitierrelevanten Bestandteile (hinsichtlich seiner

Verwendung im wissenschaftlichen Umfeld) einer bibliographischen Beschreibung eines Werkes ausgelegt. Damit soll die Forderung nach der Einfachheit der Erfassung erfüllt werden. Diese Zielsetzung führt zu **gewissen Begrenzungen wie dem Fehlen einiger seltener Datentypen**.

Um die Abhängigkeitsbehandlung möglichst einfach zu gestalten, wurde in diesem Entwurf bewusst auf die vielfältigen Möglichkeiten eines Thesaurus oder der OWL (Web Ontology Language) verzichtet. Eine Einbindung solcher beschreibenden Merkmale kann in Zukunft helfen, genauere Aussagen zu Inhalten oder Zusammenhängen maschinell zu treffen. Aus Gründen der modellhaften Vereinfachung wurde auf eine Ausarbeitung solcher Datenfelder verzichtet. Einer Umsetzung steht bei der Verwendung von XML jedoch nichts im Wege.

Anders als in den meisten bibliographischen Formaten ist es möglich alle Sprachen des Inhalts zu erfassen und Angaben zur Sprache der Metadaten zu machen. Es ist möglich, den Inhalt des deutsch-polnischen Buches „*Die deutsch-polnischen Verträge vom 14.11.1990 und 17.6.1991. Traktaty polsko-niemieckie z 14.11.1990 r. i 17.6.1991 r.*“ als "deutschsprachig" und zugleich als "auf polnisch" anzugeben. Es ist jedoch nicht möglich das Feld "Titel" mit Hilfe vom XML-Attribut "lang" als in diesen Sprachen abgefasst anzugeben. Man könnte zwar den Titel in Titel und Untertitel auftrennen (und so diese Schwachstelle umgehen), würde aber zugleich die Absicht des Herausgebers, beide Sprachen als gleichberechtigt anzusehen, missachten. In diesem seltenen Fall ist es nicht möglich, eine hinreichende Lösung anzubieten, ohne zu erlauben das Feld "Titel" mehrfach zu verwenden.

Eine weitere Schwachstelle dieses Entwurfs ist die sehr allgemein verstandene Verwendung des Unified Resource Identifier. Die Einbindung eines elektronischen Verweises ist heutzutage Pflicht, zumal das Open-Access-Prinzip (d. h. das wissenschaftliche Publikationsmodell, welches auf das Internet als kostenlosen und freien Zugang zu Forschungsergebnissen setzt) durch die Berliner Erklärung vom 22.10.2003 zum Ziel der wichtigsten wissenschaftlichen Institute und Gesellschaften Deutschlands erklärt wurde.

Die Angabe einer URI (oder URL) ohne weitere Typisierung ist zu vieldeutig. Der Benutzer weiß nicht, um welchen Inhalt es sich hierbei handelt: den Volltext, einen Abstract, eine Rezension, einem Diskussionsforum zum Buch, die Verlagsseite mit weiteren Informationen oder eine Verkaufsseite des Werkes. Die Zahl der möglichen URI-Typen ist nur schwer zu ermitteln. Bis auf MODS und ONIX for books typisiert kein anderes Format die angeführte URL. Neue technische Entwicklungen (wie Ebook-Reader o.Ä.) oder neuartige Geschäftsmodelle machen es schwer, eine vollständige und verbindliche Liste festzulegen. Es bleibt dem Benutzer überlassen, den verlinkten Inhalt intellektuell zu erschließen und einem Typ zuzuordnen. In den meisten Fällen wird die Datenquelle (ein Katalog oder eine Lesezeichensammlung) eine solche Unterscheidung auch nicht kennen. Somit bleibt das nicht näher definierte URI-Feld die kleinste gemeinsame Schnittmenge. Besondere Fähigkeiten wie die Suche nach Werken, die einen Link zum geschützten Volltext besitzen (vgl. MedLine / PubMed), sind nur mit einer URL realisierbar. Eine Erweiterung des Formats auf solche besonderen Fälle ist dank des XML-Einsatzes nicht ausgeschlossen.

Die Angabe von Relationen zwischen einzelnen Werken ist stark vereinfacht. Mit Hilfe von Verweisungen zur übergeordneten Einheit, ist jedoch eine mehrfache Rückwärtsauflösung möglich. Äquivalenzen zwischen mehreren Elementen können durch eine konsequente Weiterverweisung untereinander ebenfalls erreicht werden.

Mangels eines allgemeinen Konsenses über die Handhabung elektronischer Quellen konnten hier nur einige modellhafte, auf Analogien basierende Ansätze des Umgangs dargestellt werden. Die in ISO 690-2 Teil 2 angegebenen Arten elektronischer Werke sind mit diesem Entwurf nicht vollständig abgedeckt. Der Einfachheit wegen wurde auf allgemeine Gültigkeit geachtet. Damit sollte es möglich sein, dieselben Regeln auf elektronische und Print-Werke anzuwenden.

6.3. Fazit

Die effizientere Nutzung bibliographischer Metadaten (und somit die Optimierung des Umgangs mit diesen) durch Verbesserung der Datenqualität ist das Ziel dieser Arbeit. Der hier erarbeitete Entwurf zeigt, wie die vorhandenen Technologien eingesetzt werden können, um dieses Ziel zu erreichen. Er sorgt für hohe Genauigkeit der Angaben und somit für besseres Recall-Precision-Verhältnis und ist bei Bedarf erweiterbar. Als Einsatzgebiet wird der wissenschaftliche Umgang mit Literatur im universitären Umfeld angesehen. Anders als in öffentlichen Bibliotheken wird hier oft in sog. Clustern (thematisch verwandten und homogenen Gruppen) gearbeitet. Der Bestand der Bibliotheken deckt sowohl Grenz- als auch Spezialgebiete ab. Diese werden mit Hilfe von Klassifikationen und Schlagwörtern geordnet und für den Kunden erschlossen. Hier sind auch sehr viele bibliographische Sonderformen anzutreffen. Es galt, diese Anforderungen zu berücksichtigen, sofern sie für die Arbeit des Wissenschaftlers von Belang sind. Die wissenschaftliche Kommunikation als eines der Haupttätigkeiten, welche von der Erstellung von Quellenangaben geprägt ist, hatte in dieser Arbeit vor der Genauigkeit einer bibliographischen Beschreibung den Vorrang: ein „Abstract“ – wenn es auch nur eine kurze und prägnante Notiz zum Inhalt – ist dem Kollegen aus der Arbeitsgruppe oft nützlicher als eine „Inhaltszusammenfassung“ oder „Rezension“, die das Problemfeld gar nicht erwähnt. Es galt das Prinzip des Nutzens in der wissenschaftlichen Arbeit – also auch der Zitierfähigkeit in unterschiedlichen Zitierformaten. Als die am schwierigsten umzusetzende Form scheint die DIN 1505-2 zu gelten. Sie konnte bis auf wenige Typen wie Zitate aus der Bibel, Gesetzen und Normen vollständig umgesetzt werden.

Als Mittel zur Erreichung dieses Zieles wurde eine XML-basierte Auszeichnungssprache entwickelt. Diese bezieht sich bewusst nur auf bibliographische Werke. Archivalien oder Multimedia-Inhalte konnten nicht berücksichtigt werden, um den Umfang dieser Arbeit nicht übermäßig auszudehnen. Die Vielfalt an Erscheinungsarten (und -modi) sollte auf das (aus Sicht des Verfassers) Notwendigste reduziert werden, ohne dass die Beschaffung oder das Wiederfinden eines Werkes erschwert wird. Mit vielen

eindeutigen benannten Datenfeldern sollte die Verwendung erleichtert werden. Die dahinter stehende Annahme, dass ein sehr „eng“ definiertes Feld zweckmäßiger ist als viele mehrdeutige Felder, wurde durch die Praxis der heute üblichen Formate und Regelwerke bestätigt. In einigen Fällen konnte die Verwendung einiger Felder nicht weiter eingegrenzt werden, um die Zukunftsfähigkeit dieses Entwurfes nicht zu gefährden. Auch diese Erkenntnis entstammt der Betrachtung des Einsatzes von MARC, MAB oder RAK-WB in einer elektronisch vernetzten Welt. Die neuen Möglichkeiten einer freien Publikation (ohne Verlag oder ausdrücklichen Nennung des Autors) im WWW konnte in diesen Entwurf einfließen – einem vergleichbaren Print-Werk entsprechend. Die Freiheit, die man durch das Hinterfragen von Regeln und Formaten gewinnt, wurde unter anderem dazu gebraucht, die Arbeit mit den wissenschaftlichen Veröffentlichungen zu reflektieren und die Zusammenhänge neu zu erkennen. So ist die Idee der Vermeidung von Redundanzen an die Verwendung von Relationen gebunden. Die Objektorientierung in Verbindung mit relationaler Sicht hilft dem Benutzer (und dem Computer), die Übersicht über Zusammenhänge und Abhängigkeiten zu behalten. Die horizontalen Relationen ermöglichen neue – bisher nach Ansicht des Autors kaum verwendete – Funktionen wie: die Suche nach Alternativen (Sprachen und Erscheinungsmodi eines Werkes) oder die Umsetzung einer einfachen Personennamen- und Schlagwortnormdatei. Die Angaben die in diesem Format übertragen werden, können sowohl von Benutzern elektronischer Kataloge wie von den Betreibern solcher benutzt werden, um das Wissen um die Werke "anzureichern" und eine bessere automatisierte Interpretation zu ermöglichen. Der Austausch und das Anreichern der Metadaten funktioniert in beide Richtungen und kann von beiden Seiten ebenfalls für eine intellektuelle Erschließung und Interpretation genutzt werden (das Community-Prinzip).

Entgegen der weit verbreiteter Praxis, die englische Sprache als die 'Lingua Franca' der Wissenschaft zu verwenden, wurden alle Sprachen als gleichwertig behandelt und dürfen frei verwendet werden (z. B. bei

Personennamen). Der Einsatz von Unicode soll sicherstellen, dass jede Sprache dieser Welt (in ihrem eigenen Zeichensatz) verwendet werden kann. Mit einigen Hilfsmitteln ist es gelungen, die Verwendung mehrerer Sprachen gleichzeitig zu ermöglichen.

Nicht alle Anforderungen bzw. bekannten Probleme älterer Regeln und Formate konnten zufriedenstellend gelöst werden. Dennoch kann dieser Entwurf mit seinen wenigen Datenfeldern und Erfassungsregeln einen neuen Weg bestreiten. Es ist nicht auf nur ein Einsatzgebiet begrenzt (wie ONIX for books), kennt keine Altlasten (wie das MARC-ausgerichtete MODS) und deckt mehr als nur die Zitat-relevanten Datenfelder ab (wie BibTeXML, RISXML u. ä.). Durch die Verwendung von XML als Basis (Syntax und Notation) kann eine Vielzahl an XML-Techniken auf dieses Format angewendet werden (XPath, XSLT).

Die hintergründige Idee eines persönlichen bibliographischen Informationssystems kann nach Überzeugung des Verfassers trotz einiger Schwächen als realisiert angesehen werden. Mit diesem Dateiformat als Datenquelle ist es seiner Ansicht nach möglich, ein System zu erstellen, welches ein gutes Recall-Precision-Verhältnis mit einer automatischen Auswertung von Zusammenhängen sicherstellt, Zeichensatz-unabhängig ist und Zugriff auf Volltexte und andere verwandten elektronischen Ressourcen bietet. Interpretationen können anhand von Bemerkungen, Schlagwortung oder Klassifizierung maschinell und intellektuell vorgenommen werden. Die innere Vernetzung mit Alternativen und verwandten Werken kann darin persistent abgelegt werden. Die Quellenangaben können maschinell in beliebigen Formaten generiert werden. Der Austausch der Daten unter Wissenschaftlern kann ohne Verlust relevanter Informationen stattfinden. Diese Merkmale tragen zur effizienteren Nutzung der bibliographischen Metadaten bei.

7. Definition des Datenformats

7.1. Datenblätter

7.1.1. Hauptklassen

Datenblatt für die Hauptklasse "Serie"

Zulässige Kinder-Klassen:

- Buch
- Periodikum
- Artikel

Anzahl der Kinder-Elemente: ≥ 2

Eigenschaften:

- **Titel**
- **Untertitel** (optional)
- **Herausgeber** (optional)
- **Kürzel** (optional)
- **Organisation** (optional, mehrfach wiederholbar; wenn nicht der Herausgeber selbst)
- **Mitarbeiter** (optional, unter Angabe der Tätigkeit, mehrfach wiederholbar)
- **Verlag** (optional)
- **Sprache** der Inhalte (optional, mehrfach wiederholbar)
- **Klassifikation** (optional, mehrfach wiederholbar)
- **Schlagwort** (optional, mehrfach wiederholbar)
- **Abstract** (optional, mehrfach wiederholbar)
- **ID** (optional, beginnend mit einem "s")
- **Lizenz** (optional)
- **Identifikator**: URI, DOI, ISSN, ISSN-L, Signatur (optional, mehrfach wiederholbar)
- **Relation**: "fortführend", "äquivalent", "original" (optional, mehrfach wiederholbar)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

Datenblatt für die Hauptklasse Buch

Zulässige Kinder-Klassen:

- Artikel

Anzahl der Kinder-Elemente: ≥ 0

Eigenschaften:

- **Titel**
- **Untertitel** (optional)
- **Herausgeber** (optional, mehrfach wiederholbar)
- **Autor** (optional, mehrfach wiederholbar)
- **Organisation** (optional, mehrfach wiederholbar; wenn nicht Herausgeber)
- **Mitarbeiter** (optional, unter Angabe der Tätigkeit, mehrfach wiederholbar)
- **Band** (optional)
- **Auflagennummer** (optional)
- **Edition** (optional)
- **Erscheinungsjahr** (optional)
- **Seitenumfang**
- **Verlag** (optional)
- **Kürzel** (optional)
- **Typ** (optional; Auswahl: "Monographie", "Dissertation", "Abschlussarbeit", "Report", "Handbuch", "Lexikon", "Katalog", "Atlas" oder Freitext; wenn nicht vorgegeben gilt "Monographie")
- **Sprache des Inhalts** (optional, mehrfach wiederholbar)
- **Klassifikation** (optional, mehrfach wiederholbar)
- **Schlagwort** (optional, mehrfach wiederholbar)
- **Abstract** (optional, mehrfach wiederholbar)
- **ID** (optional, beginnend mit einem "b")
- **Lizenz** (optional)
- **Identifikator**: URI, DOI, ISBN, ISSN-L, Signatur (optional, mehrfach wiederholbar)
- **Volltext** (optional, mehrfach wiederholbar: URI, DOI)
- **Relation**: "gehört_zu", "fortführend", "äquivalent", "original" (optional, mehrfach wiederholbar)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

Datenblatt für die Hauptklasse Periodikum

Zulässige Kinder-Klassen:

- Artikel

Anzahl der Kinder-Elemente: ≥ 1

Eigenschaften:

- **Titel** (optional)
- **Untertitel** (optional)
- **Herausgeber** (optional [nur wenn bereits in der Serie enthalten], mehrfach wiederholbar)
- **Organisation** (optional, mehrfach wiederholbar; wenn nicht Herausgeber)
- **Mitarbeiter** (optional, unter Angabe der Tätigkeit, mehrfach wiederholbar)
- **Band** (optional)
- **Nummer** (optional)
- **Jahrgang** (optional)
- **Datum** (optional)
- **Seitenumfang** (als Anfang- und Ende-Seitennummer [Buchstaben und Zahlen erlaubt] bzw. eine Ganzzahl)
- **Verlag** (optional)
- **Kürzel** (optional)
- **Typ** (optional; Auswahl: "Journal", "Konferenz", "Bericht", "Jahrbuch" oder Freitext; wenn nicht angegeben gilt "Journal")
- **Sprache** des Inhalts (optional, mehrfach wiederholbar)
- **Klassifikation** (optional, mehrfach wiederholbar)
- **Schlagwort** (optional, mehrfach wiederholbar)
- **Abstract** (optional, mehrfach wiederholbar)
- **ID** (optional, beginnend mit einem "p")
- **Lizenz** (optional)
- **Identifikator**: URI, DOI, ISSN, ISBN, ISSN-L, Signatur (optional, mehrfach wiederholbar)
- **Volltext** (optional, mehrfach wiederholbar: URI, DOI)
- **Relation**: "gehört_zu", "fortführend", "äquivalent", "original" (optional, mehrfach wiederholbar)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

Datenblatt für die Hauptklasse Artikel

Zulässige Kinder-Klassen:

- keine

Anzahl der Kinder-Elemente: 0

Eigenschaften:

- **Titel**
- **Untertitel** (optional)
- **Autor** (mehrfach wiederholbar)
- **Organisation** (optional, mehrfach wiederholbar)
- **Mitarbeiter** (optional, unter Angabe der Tätigkeit, mehrfach wiederholbar)
- **Seitenumfang** (als Anfang- und Ende-Seitennummer oder als Aufzählung von Seiten [Buchstaben und Zahlen erlaubt])
- **Kürzel** (optional)
- **Typ** (optional; Auswahl: "Artikel", "Loseblattausgabe" oder Freitext; wenn nichts angegeben gilt "Artikel")
- **Sprache** des Inhalts (optional, mehrfach wiederholbar)
- **Klassifikation** (optional, mehrfach wiederholbar)
- **Schlagwort** (optional, mehrfach wiederholbar)
- **Abstract** (optional, mehrfach wiederholbar)
- **ID** (optional, beginnend mit einem "a")
- **Lizenz** (optional)
- **Identifikator**: URI, DOI, ISSN, Signatur (optional, mehrfach wiederholbar)
- **Volltext** (optional, mehrfach wiederholbar: URI, DOI)
- **Relation**: "gehört_zu", "fortführend", "äquivalent", "original" (optional, mehrfach wiederholbar)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

1. Hilfsklassen

Datenblatt für die Hilfsklasse Person

Zulässige Kinder-Klassen:

- keine

Anzahl der Kinder-Elemente: 0

Eigenschaften:

- **Rufname** (optional, an Stelle des üblichen Namens)
- **Präfix** (optional)
- **Vorname** (mehrfach wiederholbar in richtiger Reihenfolge)
- **Infix** (optional)
- **Nachname**
- **Suffix** (optional)
- **Anmerkungen** (optional)
- **ID** (optional, beginnend mit einem "m")
- **Identifikator** (zu externen Datenbanken, optional)
- **Relation:** "original" (immer zum Geburtsnamen hin verlinkt wird, optional, mehrfach wiederholbar)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

Wiederholbarkeit:

- Mehrfach

Reihenfolge bei der Erfassung und Sortierung:

- Die Reihenfolge des Vorkommens im Datenbestand ist beliebig
- Sortierung nur für Quellenangaben von Bedeutung
- Maschinell nach Landessprache, Norm und Stil-Vorlage sortierbar

Datenblatt für die Hilfsklasse Organisation**Zulässige Kinder-Klassen:**

- keine

Anzahl der Kinder-Elemente: 0**Eigenschaften:**

- **Name**
- **Abteilung** (optional)
- **Ort** (optional)
- **Funktion**: Kurzbezeichnung für die Funktion der genannten Organisation im Bezug auf das Entstehen des Werkes, wenn nicht Herausgeber
- **ID** (optional, beginnend mit einem "o")
- **Identifikator** (zu externen Datenbanken, optional)
- **Relation**: "gehört_zu", „fortführend“ (optional, mehrfach wiederholbar)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

Wiederholbarkeit:

- Mehrfach

Reihenfolge (Sortierung):

- Beliebig (darf umgestellt werden)

Datenblatt für die Hilfsklasse Verlag**Zulässige Kinder-Klassen:**

- keine

Anzahl der Kinder-Elemente: 0**Eigenschaften:**

- **Name**
- **Ort**
- **ID** (optional, beginnend mit einem "v")

Wiederholbarkeit:

- Nicht wiederholbar

Datenblatt für die Hilfsklasse Abstract**Zulässige Kinder-Klassen:**

- keine (im hierarchischen Sinne)

Anzahl der Kinder-Elemente: 0**Eigenschaften:**

- **Text**
- **Person** (optional, Angabe zum Autor)
- **Relation:** "betrifft" (optional, wenn Abstract außerhalb des Werkes)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

Wiederholbarkeit:

- Mehrfach

Reihenfolge (Sortierung):

- Beliebig (darf umgestellt werden)

2. Datentypen

Datenblatt für den Datentyp Relation

Inhalt:

- ID des betreffenden Objektes

Eigenschaften:

- **Typ** mit einem der folgenden Werte: „gehört_zu“, „äquivalent“, „original“, „betrifft“, „fortführend“
- **Referenz:** ID des betreffenden Objektes

Wiederholbarkeit:

- Mehrfach bei äquivalent und betrifft

Reihenfolge (Sortierung):

- Beliebig (darf umgestellt werden)

Datenblatt für den Datentyp Klassifikation

Inhalt:

- Klassenidentifikator

Eigenschaften:

- **Typ** (Kürzel der verwendeten Klassifikation wie: DDC, IPC, ICD10, Colon, Pica oder andere)

Wiederholbarkeit:

- Mehrfach

Reihenfolge (Sortierung):

- Beliebig (darf umgestellt werden)

Datenblatt für den Datentyp Schlagwort

Inhalt:

- Ein freies oder Thesaurus-gebundenes Schlagwort (je nach Usus)

Eigenschaften:

- **ID** (optional, beginnend mit einem "w")
- **Typ** mit einem der folgenden Werte: „gehört_zu“, „äquivalent“, „original“, „fortführend“ (optional - nur zusammen mit "Referenz"; für die Bildung einfacher Thesauri)
- **Referenz** auf ein anderes Schlagwort (optional; zusammen mit der Eigenschaft "Typ" für die Bildung einfacher Thesauri; ohne die Eigenschaften "Typ" und "ID" für bloße Zuweisung)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Erfassung, optional)

Wiederholbarkeit:

- Mehrfach

Reihenfolge (Sortierung):

- Beliebig (darf umgestellt werden)

Datenblatt für den Datentyp Autor

Inhalt:

- Objekt der Klasse Person (inline)

Eigenschaften (alternativ zum Inhalt):

- **Referenz:** Referenz auf die ID eines Objektes der Klasse Person

Wiederholbarkeit:

- Mehrfach

Reihenfolge (Sortierung):

- Wie angegeben

Datenblatt für den Datentyp Herausgeber

Inhalt:

- Objekt der Klasse Person oder der Klasse Organisation (inline)

Eigenschaften (alternativ zum Inhalt):

- **Referenz** (Referenz auf die ID eines Objektes der Klasse Person oder der Klasse Organisation)

Wiederholbarkeit:

- Mehrfach

Reihenfolge (Sortierung):

- Wie angegeben

Datenblatt für den Datentyp Mitarbeiter

Inhalt:

- Objekt der Klasse Person (inline)

Eigenschaften:

- **Referenz** (alternativ zum Inhalt): Referenz auf die ID eines Objektes der Klasse Person
- **Funktion** (Kurzbezeichnung für die Funktion der genannten Person im Bezug auf das Entstehen des Werkes)
- XML-Attribut "lang" (zur Auszeichnung der Sprache der Typ-Eigenschaft)

Wiederholbarkeit:

- Mehrfach

Reihenfolge (Sortierung):

- Wie angegeben

7.2. XML-Schema für das Datenformat

```
<?xml version="1.0" encoding="UTF-8"?>
<grammar
  xmlns="http://relaxng.org/ns/structure/1.0"
```

```
xmlns:xml='http://www.w3.org/XML/1998/namespace'
datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes">

<start>
  <element name="Bibliographie">
    <interleave>
      <zeroOrMore>
        <ref name="Serie"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Buch"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Periodikum"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Artikel"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Person"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Organisation"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Verlag"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Abstract"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Schlagwort"/>
      </zeroOrMore>
    </interleave>
  </element>
</start>

<define name="ID" datatypeLibrary="http://www.w3.org/2001/XMLSchema-
datatypes">
  <data type="ID"/>
</define>

<define name="IDREF" datatypeLibrary="http://www.w3.org/2001/XMLSchema-
datatypes">
  <data type="IDREF"/>
</define>

<define name="IDREFS" datatypeLibrary="http://www.w3.org/2001/XMLSchema-
datatypes">
  <data type="IDREFS"/>
</define>

<define name="int" datatypeLibrary="http://www.w3.org/2001/XMLSchema-
datatypes">
  <data type="int"/>
</define>

<define name="language" datatypeLibrary="http://www.w3.org/2001/XMLSchema-
datatypes">
  <data type="language"/>
</define>

<define name="xml_lang">
```

```
<optional>
  <attribute name="xml:lang">
    <ref name="language"/>
  </attribute>
</optional>
</define>

<define name="Author0derHerausgeber">
  <choice>
    <group>
      <group>
        <oneOrMore>
          <ref name="Autor"/>
        </oneOrMore>
      </group>
      <group>
        <zeroOrMore>
          <ref name="Herausgeber"/>
        </zeroOrMore>
      </group>
    </group>
    <ref name="Herausgeber"/>
  </choice>
</define>

<define name="Herausgeber">
  <oneOrMore>
    <element name="Herausgeber">
      <optional>
        <attribute name="ID">
          <ref name="ID"/>
        </attribute>
      </optional>
      <choice>
        <ref name="Person"/>
        <ref name="Organisation"/>
        <attribute name="Referenz">
          <ref name="IDREF"/>
        </attribute>
      </choice>
    </element>
  </oneOrMore>
</define>

<define name="Autor">
  <element name="Autor">
    <choice>
      <ref name="Person"/>
      <attribute name="Referenz">
        <ref name="IDREF"/>
      </attribute>
    </choice>
  </element>
</define>

<define name="Klassifikation">
  <element name="Klassifikation">
    <optional>
      <attribute name="Typ">
        <choice>
          <value>DDC</value>
          <value>BK</value>
        </choice>
      </attribute>
    </optional>
  </element>
</define>
```

```
<value>IPC</value>
<value>ICD10</value>
<value>Colon</value>
<text/>
</choice>
</attribute>
</optional>
<text/>
</element>
</define>

<define name="Identifikator">
<attribute name="Typ">
<choice>
<value>DOI</value>
<value>URI</value>
<value>ISBN</value>
<value>ISSN</value>
<value>ISSN-L</value>
<value>Signatur</value>
</choice>
</attribute>
<text/>
</define>

<define name="Identifikator_Serie">
<attribute name="Typ">
<choice>
<value>DOI</value>
<value>URI</value>
<value>ISSN</value>
<value>ISSN-L</value>
<value>Signatur</value>
</choice>
</attribute>
<text/>
</define>

<define name="Identifikator_Artikel">
<attribute name="Typ">
<choice>
<value>DOI</value>
<value>URI</value>
<value>ISSN-L</value>
<value>Signatur</value>
</choice>
</attribute>
<text/>
</define>

<define name="Person">
<choice>
<element name="Person">
<optional>
<attribute name="ID">
<ref name="ID"/>
</attribute>
</optional>
<choice>
<group>
<zeroOrMore>
<element name="Prefix">
<ref name="xml_lang"/>

```

```
<text/>
</element>
</zeroOrMore>
<oneOrMore>
  <element name="Vorname">
    <ref name="xml_lang"/>
    <text/>
  </element>
</oneOrMore>
<zeroOrMore>
  <element name="Infix">
    <ref name="xml_lang"/>
    <text/>
  </element>
</zeroOrMore>
<element name="Nachname">
  <ref name="xml_lang"/>
  <text/>
</element>
<zeroOrMore>
  <element name="Suffix">
    <ref name="xml_lang"/>
    <text/>
  </element>
</zeroOrMore>
</group>
<element name="Rufname">
  <ref name="xml_lang"/>
  <text/>
</element>
</choice>
<zeroOrMore>
  <element name="Anmerkung">
    <ref name="xml_lang"/>
    <text/>
  </element>
</zeroOrMore>
<zeroOrMore>
  <element name="Identifikator">
    <optional>
      <attribute name="Typ">
        <text/>
      </attribute>
    </optional>
    <text/>
  </element>
</zeroOrMore>
<optional>
  <ref name="Relation_original"/>
</optional>
</element>
<element name="Person">
  <attribute name="Referenz">
    <ref name="IDREF"/>
  </attribute>
</element>
</choice>
</define>

<define name="Organisation">
  <choice>
    <element name="Organisation">
      <ref name="xml_lang"/>
    </element>
  </choice>
</define>
```

```
<element name="Name">
  <ref name="xml_lang"/>
  <text/>
</element>
<optional>
  <element name="Abteilung">
    <ref name="xml_lang"/>
    <text/>
  </element>
</optional>
<optional>
  <element name="Ort">
    <ref name="xml_lang"/>
    <text/>
  </element>
</optional>
<optional>
  <attribute name="ID">
    <ref name="ID"/>
  </attribute>
</optional>
<optional>
  <attribute name="Funktion">
    <text/>
  </attribute>
</optional>
<zeroOrMore>
  <element name="Identifikator">
    <optional>
      <attribute name="Typ">
        <text/>
      </attribute>
    </optional>
    <text/>
  </element>
</zeroOrMore>
<zeroOrMore>
  <ref name="Relation_fortführend"/>
</zeroOrMore>
<zeroOrMore>
  <ref name="Relation_gehört_zu"/>
</zeroOrMore>
</element>
<element name="Organisation">
  <attribute name="Referenz">
    <ref name="IDREF"/>
  </attribute>
  <optional>
    <attribute name="Funktion">
      <text/>
    </attribute>
  </optional>
</element>
</choice>
</define>

<define name="Relation_gehört_zu">
  <element name="Relation">
    <attribute name="Typ">
      <value>gehört_zu</value>
    </attribute>
    <attribute name="Referenz">
      <ref name="IDREF"/>
    </attribute>
  </element>
</define>
```

```
</attribute>
</element>
</define>
<define name="Relation_fortführend">
  <element name="Relation">
    <attribute name="Typ">
      <value>fortführend</value>
    </attribute>
    <attribute name="Referenz">
      <ref name="IDREF"/>
    </attribute>
  </element>
</define>
<define name="Relation_äquivalent">
  <element name="Relation">
    <attribute name="Typ">
      <value>äquivalent</value>
    </attribute>
    <attribute name="Referenz">
      <ref name="IDREF"/>
    </attribute>
  </element>
</define>
<define name="Relation_original">
  <element name="Relation">
    <attribute name="Typ">
      <value>original</value>
    </attribute>
    <attribute name="Referenz">
      <ref name="IDREF"/>
    </attribute>
  </element>
</define>
<define name="Relation_betrifft">
  <element name="Relation">
    <attribute name="Typ">
      <value>betrifft</value>
    </attribute>
    <attribute name="Referenz">
      <ref name="IDREF"/>
    </attribute>
  </element>
</define>

<define name="Serie_Relationen">
  <optional>
    <ref name="Relation_fortführend"/>
  </optional>
  <zeroOrMore>
    <ref name="Relation_äquivalent"/>
  </zeroOrMore>
  <optional>
    <ref name="Relation_original"/>
  </optional>
</define>

<define name="Buch_Relationen">
  <optional>
    <ref name="Relation_gehört_zu"/>
  </optional>
  <optional>
    <ref name="Relation_fortführend"/>
  </optional>
</define>
```

```
<zeroOrMore>
  <ref name="Relation_äquivalent"/>
</zeroOrMore>
<optional>
  <ref name="Relation_original"/>
</optional>
</define>

<define name="Periodikum_Relationen">
  <optional>
    <ref name="Relation_gehört_zu"/>
  </optional>
  <optional>
    <ref name="Relation_fortführend"/>
  </optional>
  <zeroOrMore>
    <ref name="Relation_äquivalent"/>
  </zeroOrMore>
  <optional>
    <ref name="Relation_original"/>
  </optional>
</define>

<define name="Artikel_Relationen">
  <optional>
    <ref name="Relation_gehört_zu"/>
  </optional>
  <optional>
    <ref name="Relation_fortführend"/>
  </optional>
  <zeroOrMore>
    <ref name="Relation_äquivalent"/>
  </zeroOrMore>
  <optional>
    <ref name="Relation_original"/>
  </optional>
</define>

<define name="Verlag">
  <choice>
    <element name="Verlag">
      <optional>
        <attribute name="ID">
          <ref name="ID"/>
        </attribute>
      </optional>
      <element name="Name">
        <ref name="xml_lang"/>
        <text/>
      </element>
      <element name="Ort">
        <ref name="xml_lang"/>
        <text/>
      </element>
    </element>
    <element name="Verlag">
      <attribute name="Referenz">
        <ref name="IDREF"/>
      </attribute>
    </element>
  </choice>
</define>
```



```
<define name="Mitarbeiter">
  <element name="Mitarbeiter">
    <choice>
      <ref name="Person"/>
      <attribute name="Referenz">
        <ref name="IDREF"/>
      </attribute>
    </choice>
    <attribute name="Funktion">
      <text/>
    </attribute>
    <ref name="xml_lang"/>
  </element>
</define>

<define name="Abstract">
  <element name="Abstract">
    <ref name="xml_lang"/>
    <interleave>
      <element name="Text">
        <text/>
      </element>
      <optional>
        <ref name="Autor"/>
      </optional>
      <zeroOrMore>
        <ref name="Relation_betrifft"/>
      </zeroOrMore>
    </interleave>
  </element>
</define>

<define name="Schlagwort">
  <element name="Schlagwort">
    <optional>
      <attribute name="ID">
        <ref name="ID"/>
      </attribute>
    </optional>
    <optional>
      <choice>
        <group>
          <attribute name="Typ">
            <choice>
              <value>gehört_zu</value>
              <value>original</value>
              <value>äquivalent</value>
              <value>fortführend</value>
            </choice>
          </attribute>
          <attribute name="Referenz">
            <ref name="IDREF"/>
          </attribute>
        </group>
        <attribute name="Referenz">
          <ref name="IDREF"/>
        </attribute>
      </choice>
    </optional>
    <ref name="xml_lang"/>
    <text/>
  </element>
</define>
```

```
<define name="Serie">
  <choice>
    <element name="Serie">
      <optional>
        <attribute name="ID">
          <ref name="ID"/>
        </attribute>
      </optional>
      <interleave>
        <zeroOrMore>
          <choice>
            <ref name="Buch"/>
            <ref name="Periodikum"/>
            <ref name="Artikel"/>
          </choice>
        </zeroOrMore>
        <element name="Titel">
          <ref name="xml_lang"/>
          <text/>
        </element>
        <optional>
          <element name="Untertitel">
            <ref name="xml_lang"/>
            <text/>
          </element>
        </optional>
        <zeroOrMore>
          <ref name="Herausgeber"/>
        </zeroOrMore>
        <optional>
          <element name="Kürzel">
            <text/>
          </element>
        </optional>
        <zeroOrMore>
          <ref name="Organisation"/>
        </zeroOrMore>
        <zeroOrMore>
          <ref name="Mitarbeiter"/>
        </zeroOrMore>
        <optional>
          <ref name="Verlag"/>
        </optional>
        <zeroOrMore>
          <element name="Sprache">
            <ref name="language"/>
          </element>
        </zeroOrMore>
        <zeroOrMore>
          <ref name="Klassifikation"></ref>
        </zeroOrMore>
        <zeroOrMore>
          <ref name="Schlagwort"/>
        </zeroOrMore>
        <zeroOrMore>
          <ref name="Abstract"/>
        </zeroOrMore>
        <optional>
          <element name="Lizenz">
            <text/>
          </element>
        </optional>
      </interleave>
    </element>
  </choice>
</define>
```

```
</optional>
<zeroOrMore>
  <element name="Identifikator">
    <ref name="Identifikator_Serie"/>
  </element>
</zeroOrMore>
<ref name="Serie_Relationen"/>
</interleave>
</element>
<element name="Serie">
  <attribute name="Referenz">
    <ref name="IDREF"/>
  </attribute>
</element>
</choice>
</define>

<define name="Buch">
  <choice>
    <element name="Buch">
      <interleave>
        <optional>
          <attribute name="ID">
            <ref name="ID"/>
          </attribute>
        </optional>
        <optional>
          <attribute name="Typ">
            <choice>
              <value>Monographie</value>
              <value>Dissertation</value>
              <value>Abschlussarbeit</value>
              <value>Report</value>
              <value>Handbuch</value>
              <value>Lexikon</value>
              <value>Katalog</value>
              <value>Atlas</value>
            <text/>
          </choice>
        </attribute>
      </optional>
      <zeroOrMore>
        <ref name="Artikel"/>
      </zeroOrMore>
      <element name="Titel">
        <ref name="xml_lang"/>
        <text/>
      </element>
      <optional>
        <element name="Untertitel">
          <ref name="xml_lang"/>
          <text/>
        </element>
      </optional>
      <ref name="AuthorOderHerausgeber"/>
      <zeroOrMore>
        <ref name="Organisation"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="Mitarbeiter"/>
      </zeroOrMore>
      <optional>
        <element name="Band">
```

```
<text/>
</element>
</optional>
<optional>
  <element name="Kürzel">
    <text/>
  </element>
</optional>
<optional>
  <element name="Auflagenummer">
    <ref name="int"/>
  </element>
</optional>
<optional>
  <element name="Edition">
    <text/>
  </element>
</optional>
<optional>
  <element name="Erscheinungsjahr"
datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes">
    <data type="gYear"/>
  </element>
</optional>
<element name="Seitenumfang">
  <ref name="int"/>
</element>
<optional>
  <ref name="Verlag"/>
</optional>
<zeroOrMore>
  <element name="Sprache">
    <ref name="language"/>
  </element>
</zeroOrMore>
<zeroOrMore>
  <ref name="Klassifikation"></ref>
</zeroOrMore>
<zeroOrMore>
  <ref name="Schlagwort"/>
</zeroOrMore>
<zeroOrMore>
  <ref name="Abstract"/>
</zeroOrMore>
<optional>
  <element name="Lizenz">
    <text/>
  </element>
</optional>
<zeroOrMore>
  <element name="Identifikator">
    <ref name="Identifikator"/>
  </element>
</zeroOrMore>
<zeroOrMore>
  <element name="Volltext">
    <attribute name="Typ">
      <choice>
        <value>DOI</value>
        <value>URI</value>
      </choice>
    </attribute>
    <text/>
  </element>
</zeroOrMore>
```

```
        </element>
      </zeroOrMore>
      <ref name="Buch_Relationen"/>
    </interleave>
  </element>
  <element name="Buch">
    <attribute name="Referenz">
      <ref name="IDREF"/>
    </attribute>
  </element>
</choice>
</define>

<define name="Periodikum">
  <choice>
    <element name="Periodikum">
      <interleave>
        <zeroOrMore>
          <ref name="Artikel"/>
        </zeroOrMore>
        <optional>
          <attribute name="ID">
            <ref name="ID"/>
          </attribute>
        </optional>
        <optional>
          <attribute name="Typ">
            <choice>
              <value>Journal</value>
              <value>Konferenz</value>
              <value>Bericht</value>
              <value>Jahrbuch</value>
            <text/>
          </choice>
        </attribute>
        </optional>
        <optional>
          <element name="Ttitel">
            <ref name="xml_lang"/>
            <text/>
          </element>
        </optional>
        <optional>
          <element name="Untertitel">
            <ref name="xml_lang"/>
            <text/>
          </element>
        </optional>
        <zeroOrMore>
          <ref name="Herausgeber"/>
        </zeroOrMore>
        <zeroOrMore>
          <ref name="Organisation"/>
        </zeroOrMore>
        <zeroOrMore>
          <ref name="Mitarbeiter"/>
        </zeroOrMore>
        <optional>
          <element name="Band">
            <text/>
          </element>
        </optional>
      </optional>
    </element>
  </choice>
</define>
```

```
<element name="Kürzel">
  <text/>
</element>
</optional>
<choice>
  <element name="Datum"
datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes">
    <data type="date"/>
  </element>
  <element name="Nummer">
    <text/>
  </element>
</interleave>
  <element name="Datum"
datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes">
    <data type="date"/>
  </element>
  <element name="Nummer">
    <text/>
  </element>
</interleave>
</choice>
<optional>
  <element name="Jahrgang">
    <text/>
  </element>
</optional>
<element name="Seitenumfang">
  <choice>
    <group>
      <element name="Anfang">
        <text/>
      </element>
      <element name="Ende">
        <text/>
      </element>
    </group>
    <group>
      <attribute name="Anfang">
        <text/>
      </attribute>
      <attribute name="Ende">
        <text/>
      </attribute>
    </group>
    <ref name="int"/>
  </choice>
</element>
<optional>
  <ref name="Verlag"/>
</optional>
<zeroOrMore>
  <element name="Sprache">
    <ref name="language"/>
  </element>
</zeroOrMore>
<zeroOrMore>
  <ref name="Klassifikation"></ref>
</zeroOrMore>
<zeroOrMore>
  <ref name="Schlagwort"/>
</zeroOrMore>
<zeroOrMore>
```

```
<ref name="Abstract"/>
</zeroOrMore>
<optional>
  <element name="Lizenz">
    <text/>
  </element>
</optional>
<zeroOrMore>
  <element name="Identifikator">
    <ref name="Identifikator"/>
  </element>
</zeroOrMore>
<zeroOrMore>
  <element name="Volltext">
    <attribute name="Typ">
      <choice>
        <value>DOI</value>
        <value>URI</value>
      </choice>
    </attribute>
    <text/>
  </element>
</zeroOrMore>
  <ref name="Periodikum_Relationen"/>
</interleave>
</element>
<element name="Periodikum">
  <attribute name="Referenz">
    <ref name="IDREF"/>
  </attribute>
</element>
</choice>
</define>
```

```
<define name="Artikel">
  <choice>
    <element name="Artikel">
      <interleave>
        <optional>
          <attribute name="ID">
            <ref name="ID"/>
          </attribute>
        </optional>
        <optional>
          <attribute name="Typ">
            <choice>
              <value>Artikel</value>
              <value>Loseblattausgabe</value>
            </choice>
          </attribute>
        </optional>
        <element name="Titel">
          <ref name="xml_lang"/>
          <text/>
        </element>
        <optional>
          <element name="Untertitel">
            <ref name="xml_lang"/>
            <text/>
          </element>
        </optional>
      </interleave>
    </element>
  </choice>
</define>
```

```
<oneOrMore>
  <ref name="Autor"/>
</oneOrMore>
<zeroOrMore>
  <ref name="Organisation"/>
</zeroOrMore>
<zeroOrMore>
  <ref name="Mitarbeiter"/>
</zeroOrMore>
<element name="Seitenumfang">
  <choice>
    <group>
      <element name="Anfang">
        <text/>
      </element>
      <element name="Ende">
        <text/>
      </element>
    </group>
    <group>
      <attribute name="Anfang">
        <text/>
      </attribute>
      <attribute name="Ende">
        <text/>
      </attribute>
    </group>
    <oneOrMore>
      <element name="Seite">
        <text/>
      </element>
    </oneOrMore>
  </choice>
</element>
<optional>
  <element name="Kürzel">
    <text/>
  </element>
</optional>
<zeroOrMore>
  <element name="Sprache">
    <ref name="language"/>
  </element>
</zeroOrMore>
<zeroOrMore>
  <ref name="Klassifikation"></ref>
</zeroOrMore>
<zeroOrMore>
  <ref name="Schlagwort"/>
</zeroOrMore>
<zeroOrMore>
  <ref name="Abstract"/>
</zeroOrMore>
<optional>
  <element name="Lizenz">
    <text/>
  </element>
</optional>
<zeroOrMore>
  <element name="Identifikator">
    <ref name="Identifikator_Artikel"/>
  </element>
</zeroOrMore>
```



```

    <zeroOrMore>
      <element name="Volltext">
        <attribute name="Typ">
          <choice>
            <value>DOI</value>
            <value>URI</value>
          </choice>
        </attribute>
        <text/>
      </element>
    </zeroOrMore>
    <ref name="Artikel_Relationen"/>
  </interleave>
</element>
<element name="Artikel">
  <attribute name="Referenz">
    <ref name="IDREF"/>
  </attribute>
</element>
</choice>
</define>

</grammar>

```

7.3. Verwendungsbeispiel

```

<?xml version="1.1" encoding="UTF-8" standalone='yes'?>
<Bibliographie>

  <!-- Personen -->
  <Person ID="m0001">
    <Rufname xml:lang="de">Benedikt XVI.</Rufname>
    <Relation Typ="original" Referenz="m0002"/>
  </Person>
  <Person ID="m0001lat">
    <Rufname xml:lang="lat">Benedict XVI</Rufname>
    <Relation Typ="original" Referenz="m0002"/>
  </Person>
  <Person ID="m0002">
    <Vorname>Joseph</Vorname>
    <Infix>Kardinal</Infix>
    <Nachname>Ratzinger</Nachname>
  </Person>
  <Person ID="m0003">
    <Vorname>Scott</Vorname>
    <Vorname>W.</Vorname>
    <Nachname>Means</Nachname>
  </Person>

  <!-- Serien -->
  <Serie ID="s0001"><!-- eine noch unvollständige mit einem Buch -->
    <Titel>In a Nutshell</Titel>
    <Kürzel>Nutshell</Kürzel>
    <Verlag Referenz="v0ReillyVerlag"/>
    <!-- Buch inline, mit ID -->
    <Buch ID="b0001" Typ="Monographie">
      <Titel>XML in a Nutshell</Titel>
      <Autor>
        <Person>
          <Vorname>Elliotte</Vorname>
          <Vorname>Rusty</Vorname>
          <Nachname>Harold</Nachname>
        </Person>

```

```
</Autor>
<Autor>
  <Person Referenz="m0003"/>
</Autor>
<Seitenumfang>702</Seitenumfang>
<Sprache>de</Sprache>
<Mitarbeiter Funktion="Übersetzer ins Deutsche">
  <Person>
    <Vorname>Katharina</Vorname>
    <Nachname>Udemadu</Nachname>
  </Person>
</Mitarbeiter>
<Mitarbeiter Funktion="Übersetzer ins Deutsche">
  <Person>
    <Vorname>Kathrin</Vorname>
    <Nachname>Lichtenberg</Nachname>
  </Person>
</Mitarbeiter>
<Mitarbeiter Funktion="Übersetzer ins Deutsche">
  <Person>
    <Vorname>Lars</Vorname>
    <Nachname>Schulten</Nachname>
  </Person>
</Mitarbeiter>
<Auflagenummer>3</Auflagenummer>
<Identifikator Typ="ISBN">3-89721-339-7</Identifikator>
<Erscheinungsjahr>2005</Erscheinungsjahr>
<Kürzel>XML-Nutshell</Kürzel>
</Buch>
</Serie>

<Serie>
  <Verlag ID="v_heise">
    <Name>heise Zeitschriften Verlag</Name>
    <Ort>Hannover</Ort>
  </Verlag>
  <Titel>c't</Titel>
  <Untertitel>magazin für computer technik</Untertitel>
  <Kürzel>ct</Kürzel>
  <Sprache>de</Sprache>
  <Identifikator Typ="URI">http://www.ct.de</Identifikator>
  <Herausgeber>
    <Person>
      <Vorname>Christian</Vorname>
      <Nachname>Heise</Nachname>
    </Person>
  </Herausgeber>
  <Herausgeber>
    <Person>
      <Vorname>Ansgar</Vorname>
      <Nachname>Heise</Nachname>
    </Person>
  </Herausgeber>
  <Herausgeber>
    <Person>
      <Vorname>Christian</Vorname>
      <Nachname>Persson</Nachname>
    </Person>
  </Herausgeber>
  <Periodikum>
    <Seitenumfang>226</Seitenumfang>
    <Identifikator Typ="ISSN">0723-8679</Identifikator>
    <Jahrgang>2010</Jahrgang>
```

```
<Nummer>13</Nummer>
<Datum>2010-06-07</Datum>
<Artikel>
  <Autor><Person>
    <Vorname>Holger</Vorname>
    <Nachname>Bleich</Nachname>
  </Person></Autor>
  <Titel>Ein falscher Klick...</Titel>
  <Untertitel>Das Internet ist ein juristischen Minenfeld</Untertitel>
  <!--
  <Seitenumfang>
    <Anfang>76</Anfang> <Ende>81</Ende>
  </Seitenumfang>
  oder alternativ -->
  <Seitenumfang Anfang='76' Ende='81' />
  <Schlagwort ID='wVerschlüsselung' Typ="äquivalent"
    Referenz="wKryptographie">Verschlüsselung</Schlagwort>
  <Schlagwort ID='wAbofalle'>Abofalle</Schlagwort>
  <Klassifikation Typ="DDC">340</Klassifikation>
  <!-- hier beispielhaft dem "Recht" zugeordnet -->
</Artikel>
</Periodikum>
</Serie>

<Serie>
  <Titel>Grundlagen der partischen Information und Dokumentation</Titel>
  <Herausgeber ID="hKuhlen">
    <Person ID="mRainerKuhlen"><!-- inline, mit ID -->
      <Prefix>Prof.</Prefix><!-- wird in Zitaten ignoriert -->
      <Vorname>Rainer</Vorname>
      <Nachname>Kuhlen</Nachname>
      <Anmerkung>1980-2010 Professor für Informationswissenschaft an der
        Universität Konstanz
      </Anmerkung>
      <!-- Identifikator (hier URL in wikipedia) -->
      <Identifikator Typ="Wikipedia">
        http://de.wikipedia.org/wiki/Rainer_Kuhlen
      </Identifikator>
    </Person>
  </Herausgeber>
  <Herausgeber ID="hSeeger">
    <Person ID="mThomasSeeger">
      <Vorname>Thomas</Vorname><Nachname>Seeger</Nachname>
    </Person>
  </Herausgeber>
  <Herausgeber ID="hStrauch">
    <Person ID="mDietmarStrauch">
      <Vorname>Dietmar</Vorname>
      <Nachname>Strauch</Nachname>
    </Person>
  </Herausgeber>
  <!-- weitere beteiligte namentlich genannten Personen -->
  <Mitarbeiter Funktion="Begründer">
    <Person>
      <Vorname>Klaus</Vorname>
      <Nachname>Laisiepen</Nachname>
    </Person>
  </Mitarbeiter>
  <Mitarbeiter Funktion="Begründer">
    <Person>
      <Vorname>Ernst</Vorname>
      <Nachname>Lutterbeck</Nachname>
    </Person>
  </Mitarbeiter>
</Serie>
```

```
</Mitarbeiter>
<Mitarbeiter Funktion="Begründer">
  <Person>
    <Vorname>Karl-Heinrich</Vorname>
    <Nachname>Meyer-Uhlenried</Nachname>
  </Person>
</Mitarbeiter>
<Schlagwort ID="wInfowiss">
  Informationswissenschaft
</Schlagwort>
<Klassifikation Typ="DDC">020</Klassifikation>
<Sprache>de</Sprache>
<Buch ID="bKSS" Typ="Handbuch">
  <Titel>Handbuch zur Einführung in die Informationswissenschaft und
    -praxis</Titel>
  <Herausgeber Referenz="hKuhlen"/>
  <Herausgeber Referenz="hSeeger"/>
  <Herausgeber Referenz="hStrauch"/>
  <Auflagenummer>5</Auflagenummer>
  <Edition>völlig neu erfasste Ausgabe</Edition>
  <Band>1</Band>
  <Verlag><Name>K G Saur</Name><Ort>München</Ort></Verlag>
  <Erscheinungsjahr>2004</Erscheinungsjahr>
  <Seitenumfang>762</Seitenumfang>
  <Identifikator Typ="ISBN">3-598-11674-8</Identifikator>
</Buch>
<Buch ID="bKSS-Glossar" Typ="Lexikon">
  <Herausgeber Referenz="hKuhlen"/>
  <Herausgeber Referenz="hSeeger"/>
  <Herausgeber Referenz="hStrauch"/>
  <Titel>Handbuch zur Einführung in die Informationswissenschaft
    und -praxis</Titel>
  <Auflagenummer>5</Auflagenummer>
  <Edition>völlig neu erfasste Ausgabe</Edition>
  <Band>2</Band>
  <Verlag><Name>K G Saur</Name><Ort>München</Ort></Verlag>
  <Erscheinungsjahr>2004</Erscheinungsjahr>
  <Seitenumfang>148</Seitenumfang>
  <Mitarbeiter Funktion="Redaktion">
    <Person Referenz="mDietmarStrauch"/>
  </Mitarbeiter>
  <Identifikator Typ="ISBN">3-598-11674-8</Identifikator>
</Buch>
</Serie>

<!-- Bücher -->
<Buch ID="b0002"><!-- wird gedeutet als "Typ='Monographie'" -->
  <Titel>Der Geist der Liturgie</Titel>
  <Untertitel>Eine Einführung</Untertitel>
  <Seitenumfang>208</Seitenumfang>
  <Verlag ID="vHerder_Freiburg">
    <Name>Herder</Name>
    <Ort>Freiburg</Ort>
  </Verlag>
  <Erscheinungsjahr>2000</Erscheinungsjahr>
  <Auflagenummer>1</Auflagenummer>
  <Autor Referenz="m0002"/>
  <Sprache>de</Sprache>
  <Abstract>
    <Autor>
      <Person ID="mRK">
        <Vorname>Robert</Vorname>
        <Nachname>Kolatzek</Nachname>
      </Person>
    </Autor>
  </Abstract>
</Buch>
```

```
</Person>
</Autor>
<Text>Ausgewählte Themen der geschichtlichen Entwicklung der
katholischen Liturgie und deren Relevanz für das heutige
Verständnis und Praxis.</Text>
</Abstract>
<Schlagwort Typ="gehört_zu" Referenz="wReligion">Liturgie</Schlagwort>
<Schlagwort>Einführung</Schlagwort>
<Schlagwort ID="wKatholisch" Typ="äquivalent"
Referenz="wKatholizismus">Katholisch</Schlagwort>
<Identifikator Typ="ISBN">3-451-27247-4</Identifikator>
</Buch>

<Buch>
<Titel>Berührt vom Unsichtbaren</Titel>
<Untertitel>Jahreslesebuch</Untertitel>
<Autor Referenz="m0002"/>
<Herausgeber>
<Person <!-- inline, ohne ID -->
<Vorname>Ludger</Vorname>
<Nachname>Hohn-Morisch</Nachname>
</Person>
</Herausgeber>
<Seitenumfang>400</Seitenumfang>
<Verlag Referenz="vHerder_Freiburg"/>
<Erscheinungsjahr>2000</Erscheinungsjahr>
<Auflagenummer>1</Auflagenummer>
<Identifikator Typ="ISBN">3-451-26425-0</Identifikator>
<Abstract>
<Autor>
<Person Referenz="mRK"/>
</Autor>
<Text>Eine Zusammenstellung kurzer Textpassagen aus den Werken Joseph
Ratzingers.</Text>
</Abstract>
<Schlagwort Referenz="wKatholisch"/>
</Buch>

<!-- Artikel -->
<Artikel ID="a0001">
<Titel>XSL Transformations (XSLT)</Titel>
<Autor Referenz="m0003"/>
<Seitenumfang>
<Anfang>146</Anfang>
<Ende>163</Ende>
</Seitenumfang>
<Relation Typ="gehört_zu" Referenz="b0001"/>
<Schlagwort ID="wXML">XML</Schlagwort>
<Schlagwort ID="wXSLT">XSLT</Schlagwort>
</Artikel>

<!-- Verlag -->
<Verlag ID="v0ReillyVerlag">
<Name>O'Reilly Verlag</Name>
<Ort>Köln</Ort>
</Verlag>

<!-- Organisationen -->
<Organisation ID="o0001">
<Name>Der Völkerbund</Name>
<Ort>Genf</Ort>
</Organisation>
```

```
<!-- Schlagworte -->
<Schlagwort ID="wKatholizismus" Typ="gehört_zu"
  Referenz="wReligion">Katholizismus</Schlagwort>
<Schlagwort ID="wReligion">Religion</Schlagwort>
<Schlagwort ID="wKryptographie">Kryptographie</Schlagwort>

</Bibliographie>
```

8. Literaturverzeichnis

ANSI/NISO Z39.88-2004: *The OpenURL Framework for Context-Sensitive Services*). URL: http://www.niso.org/standards/standard_detail.cfm?std_id=783 Stand: 31.12.2007

Bekavac, Bernard: *Information und Kommunikationstechnologien*. In: KSS

Berners-Lee, Tim; Hendler, James; Lassila, Ora: *The semantic web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. Scientific American. 5 / 2001

Bibliographisches Institut & F. A. Brockhaus AG (Hrsg.): *Postscript®*. URL: <http://lexikon.meyers.de/index.php?title=PostScript%C2%AE&oldid=41521> zuletzt online: 23.03.2009

Bibliographisches Institut & F. A. Brockhaus AG (Hrsg.): *Transkription*. In: *Meyers Lexikon online*. Stand: 25.08.2006. URL: <http://lexikon.meyers.de/index.php?title=Transkription&oldid=54765> zuletzt online: 23.03.2009

Bibliographisches Institut & F. A. Brockhaus AG (Hrsg.): *Transliteration*. In: *Meyers Lexikon online*. Stand: 25.08.2006. URL: <http://lexikon.meyers.de/index.php?title=Transliteration&oldid=54768> zuletzt online: 23.03.2009

CSS (*Cascading Style Sheet*) In: *InfoWissWiki*. URL: http://server02.is.uni-sb.de/courses/wiki/index.php?title=CSS_%28Cascading_Style_Sheets%29&oldid=6098

-
- Deutsches Bibliotheksinstitut (Hrsg.): *Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken*. 2., überarbeitete Ausgabe. Berlin, 1993. - ISBN 3-87068-436-4
- Deutsches Institut für Normung (Hrsg.): *Format für den Austausch von bibliographischen Daten. DIN 1506*. (Verabschiedet im März 1978) In: Deutsches Institut für Normung e. V. (Hrsg.): *Publikation und Dokumentation* 2. 3. Auflage. Berlin: Beuth, 1989 (DIN-Taschenbuch 154). - ISBN 3-410-12343-1. Kurz: „**DIN-Taschenbuch-154**“
- Deutsches Institut für Normung (Hrsg.): *Titelangaben von Dokumenten: Titelaufnahme von Schrifttum. DIN 1505 Teil 1*. (Verabschiedet im Mai 1984) In: *DIN-Taschenbuch-154*
- Deutsches Institut für Normung (Hrsg.): *Titelangaben von Dokumenten: Zitierregeln. DIN 1505 Teil 2*. (Verabschiedet im Januar 1984) In: *DIN-Taschenbuch-154*
- Deutsche Nationalbibliothek (Hrsg.): *MABXML*. URL: <http://www.d-nb.de/standardisierung/formate/mabxml.htm> Stand: 01.12.2007
- Deutsche Nationalbibliothek (Hrsg.): *MABxml-Datei*. http://www.d-nb.de/standardisierung/formate/mabxml_beispiel_ebene1.xml Stand 20.01.2011
- Deutsche Nationalbibliothek (Hrsg.): *DNB, Personennamendatei (PND)*. URL: <http://www.d-nb.de/standardisierung/normdateien/pnd.htm> Stand: 16.11.2008
- Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (Hrsg.): *Funktionelle Anforderungen an bibliografische Datenätze. Abschlussbericht der IFLA Study Group on the Functional Requirements for Bibliographic Records*. urn:nbn:de:1111-20040721195 (URL: http://www.d-nb.de/standardisierung/pdf/frbr_deutsch.pdf)
- Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (Hrsg.): *Individualisierungsrichtlinie für die Personennamendatei*. S. 4 ff. URL: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:1111-20040721161>
- Die Deutsche Bibliothek (Hrsg.): *Projekt Umstieg auf internationale Formate und Regelwerke (MARC21, AACR2) – Abschlussbericht*. URL: http://www.ddb.de/standardisierung/pdf/dfg_abschlussbericht.pdf Stand: 21.09.2006
- Diller, Antoni: *LaTeX wiesz po wierszu / Jełowiecki Jan (Übers.)*. Gliwice: Helion, 2001. - Originaltitel: *LaTeX Line by Line*. - ISBN 83-7197-341-1.

-
- Dublin Core Metadata Initiative (Hrsg.): „*DCMI Metadata Terms: A complete historical record*“ <http://dublincore.org/usage/terms/history/> Stand: 12.12.2008
- EDItEUR (Hrsg.): *ONIX for Books*. URL: <http://www.editeur.org/onix.html> Stand: 22.12.2007
- EDItEUR (Hrsg.): ONIX for Books XML-Schema. URL: http://www.editeur.org/onix/2.1/reference/ONIX_BookProduct_Release2.1_reference.xsd Stand: 22.12.2007
- Eversberg, B.: *Was sind und was sollen Bibliothekarische Datenformate?*. Braunschweig : Univ.-Bibliothek der TU, 1994. - ISBN 3-927115-21-5. Als WWW-Version mit Ergänzungen (2003) URL: <http://www.allegro-c.de/formate/> Stand: 12.12.2008
- Flimm, Oliver: *Die Open-Source-Software OpenBib an der USB Köln - Überblick und Entwicklungen in Richtung OPAC 2.0* . In: Bibliothek. Forschung und Praxis, Jg. 31 (2007) Nr. 2 . S. 9ff URL: <http://eprints.rclis.org/archive/00009891/01/openbib-opac20-2007.pdf> Stand: 20.06.2008
- Funke, Fritz: *Buchkunde : Die historische Entwicklung des Buches von der Keilschrift bis zur Gegenwart*. VMA-Verlag, Wiesbaden 2006, ISBN 3-928127-95-0
- Gömpel, Renate ; Frodl, Christine ; Hengel, Christel ; Kutz, Martin ; Münnich, Monika ; Werner, Claudia: *Aus den Veranstaltungen der Division IV Bibliographic Control des Weltkongresses Bibliothek und Information, 69. IFLA-Generalkonferenz in Berlin*. In: Bibliotheksdienste 2003, Nr. 10.
- Harms, Ilse (Hrsg.) ; Luckhardt, Heinz-Dirk (Hrsg.) ; Giessen Hans W. (Hrsg.): *Information und Sprache : Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern : Festschrift für Harald H. Zimmermann*. München: K G Saur, 2006. – ISBN-13 978-3-598-11754-1. Kurz: "**Festschrift Zimmermann**"
- Harold, Elliotte R. ; Means, W. Scott: *XML in a Nutshell* / Deutsche Übersetzung von Katharina Udemadu, Kathrin Lichtenberg und Lars Schulten. 3. Auflage. Köln : O'Reilly, 2005. - ISBN 3-89721-339-7
- Havemann, Frank; Kaufmann, Andrea: *Der Wandel des Benutzerverhaltens in Zeiten des Internet – Ergebnisse von Befragungen an 13 Bibliotheken*. In: *Festschrift für Walther Umstätter*, S. 65-89. (Als PDF-Datei der CD-ROM-Ausgabe oder im Print: ISBN 978-3-88347-248-5, Bad Honnef: Bock + Herchen, 2006.)

-
- Helmes, Leni; Steidl, Nicole: *Webservices praktisch angewandt. FIZ Karlsruhe entwickelt Automatisierung der Informationsstruktur*. S. 421-428. In: Information. Wissenschaft und Praxis. 2003
- Herb, Ulrich: *Ohne Web 2.0 keine Bibliothek 2.0*. URN: urn:nbn:de:bsz:291-scidok-12917
URL: <http://scidok.sulb.uni-saarland.de/volltexte/2007/1291/> Stand: 15.12.2008
- Herzog Gottfried ; Wiesner Hans-Jörg: *Normung*. In: KSS.
- Hilberer, Thomas: *Aufwand vs. Nutzen : Wie sollen deutsche wissenschaftliche Bibliotheken künftig katalogisieren?* In: Bibliotheksdienst 2003, Nr. 6
- Hobohm, Hans-Christoph: *Persönliche Literaturverwaltung im Umbruch. Vom Bibliographie-Management zum Social Bookmarking. Anmerkungen zu EndNote, Reference Manager, RefWorks und Connotea*. In: Information. Wissenschaft und Praxis. 7 / 2005. S. 358-388
- International Standard Organisation (Hrsg.): *Information and documentation – Bibliographic references - Content, form and structure. ISO 690*.
- Kailer, Michaela: *Gütezeichen für barrierefreies Internet : Eine theoretische und empirische Analyse*. Wien, Universität Wien, Fakultät für Wirtschaftswissenschaften, Diplomarbeit, 2009.
- Kalbitz, Andreas: *LaTeX-Geschichte* URL: http://www.selflinux.org/selflinux/html/latex_geschichte01.html Stand: 14.10.2006
- Kerschis, Annett: *Literaturverwaltung und Wissensorganisation im Vergleich. Das Angebot von Literaturverwaltungsprogrammen und Social Bookmarking in Bezug auf die Benutzbarkeit in Bibliotheken*. Potsdam, Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Diplomarbeit, 2007
- Kett, Jürgen: *MABxml-1. Dokumentation*. URL: http://www.d-nb.de/standardisierung/pdf/mabxml_1_dok.pdf Version 1.2, Stand 01.12.2007
- Kett, Jürgen: *Regeln zur Übertragung von MAB2-Datensätzen nach MABxml-1*. URL: http://www.d-nb.de/standardisierung/pdf/mabxml_1_uebertr.pdf Version 1.0, Stand: 01.12.2007

Kompetenzzentrum Interoperable Metadaten (Hrsg.): *Umfragerport zur Nutzung von Metadaten*. URL: <http://www.kim-forum.org/material/pdf/KIM-Umfragerport.pdf>
Stand: 28.06.2008

Krause, Jürgen: *Shell Model, Semantic Web ans Web Information Retrieval*. In: *Festschrift Zimmermann*

Kuhlen, Reiner: *Informationsaufbereitung III: Referieren (Abstracts – Abstracting – Grundlagen)*. In: *KSS*

Kuhlen, Reiner (Hrsg.) ; Seeger, Thomas (Hrsg.) ; Strauch, Dietmar (Hrsg.): *Grundlagen der praktischen Information und Dokumentation. Bd. 1 : Handbuch zur Einführung in die Informationswissenschaft und -praxis*. 5. Auflage. München: K G Saur, 2004. - ISBN 3-598-11674-8. Kurz: „**KSS**“

Kuhlen, Reiner (Hrsg.) ; Seeger, Thomas (Hrsg.) ; Strauch, Dietmar (Hrsg.): *Grundlagen der praktischen Information und Dokumentation. Bd. 2 : Glossar*. 5. Auflage. München: K G Saur, 2004. - ISBN 3-598-11674-8. Kurz „**KSS-Glossar**“. Darin:

Auszeichnungssprache.

Client-Server-Architektur.

DTD.

Dewey-Dezimalklassifikation.

Dezimalklassifikation.

Formalerschließung.

Klassifikation.

Medical Subject Headings.

Metadaten.

Open Access.

Protokolle.

RDF.

Schlagwort.

Unicode.

XML.

Lang, Elke: *Datenbanken und Datenbank-Management-Systeme*. In: *KSS*.

Lautenschlager, Michael; Sens, Iris: *Konzept zur Zitierfähigkeit wissenschaftlicher Primärdaten*. In: *Information. Wissenschaft und Praxis*, 2003.

Lepsky, Klaus ; Zimmermann Harald H.: *Katalogerweiterung durch Scanning und automatische Dokumenterschließung . Ergebnisse des DFG-Projekts KASCADE*. In: *Zeitschrift für Bibliothekswesen und Bibliographie* 4/00, S. 305-316. URL: <http://is.uni-sb.de/zimmermann/pdf/2000a.pdf> Stand: 15.12.2008

Lessig, Lawrence: *Freie Kultur. Wesen und Zukunft der Kreativität*. Deutsche..Online-Ausgabe. URL: https://www.opensourcepress.de/freie_kultur/index.php Stand: 08.04.2010

Library of Congress (Hrsg.): *CQL: the Contextual Query Language: Specifications (SRU: Search/Retrieval via URL, Standards, Library of Congress)*. URL: <http://www.loc.gov/standards/sru/specs/cql.html> Stand: 11.11.2007

Library of Congress (Hrsg.): *Examples - MODS User Guidelines Version 3: Metadata Object Description Schema: MODS (Library of Congress)*. URL: <http://www.loc.gov/standards/mods/v3/mods-userguide-examples.html> Stand: 15.12.2008

Library of Congress (Hrsg.): *MARC in XML*. URL: <http://www.loc.gov/marc/marcxml.html> Stand: 01.12.2007

Library of Congress (Hrsg.): *Schema MARC21slim.xsd*. URL: <http://www.loc.gov/standards/marcxml/xml/spy/spy.html> Stand: 01.12.2007

Library of Congress (Hrsg.): *MODS: Uses and Features*. URL: <http://www.loc.gov/standards/mods/mods-overview.html> Stand: 14.12.2007

Library of Congress, Cataloging in Publication Division (Hrsg.): *The Cataloging in Publication Program*. URL: <http://cip.loc.gov/> Stand: 20.11.2006

Manecke, Hans-Jürgen: *Klassifikation, Klassieren*. In: KSS.

McCallum; Sally H.: *A Look at New Information Retrieval Protocols: SRU, OpenSearch/A9, CQL, and XQuery*. (Vortrag an: World Library And Information Congress: 72nd IFLA General Conference And Council, 20-24.08.2006, Seoul, Korea)

Müller, Heike: *Erstellung von Bibliographien auf Basis von XML und XSLT*. Diplomarbeit. Fachhochschule Stuttgart (Hochschule der Medien). 2001

National Center for Biotechnology Information (NCBI) (Hrsg.): *MEDLINE Display*. URL: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helppubmed.table.pubmedhelp.T44> Stand: 12.12.2008

National Center for Biotechnology Information (NCBI) (Hrsg.): *PubMed Data Provider Documentation: NLM Standard Publisher Data Format*. URL: <http://www.ncbi.nlm.nih.gov/entrez/query/static/spec.html> Stand: 12.12.2008

National Information Standards Organization (Hrsg.): *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*. Bethesda: NISO Press, 2003 - ISBN: 1-880124-55-6.

Online Computer Library Center, Inc. (Hrsg.): *OpenURL ContextObject in SPAN (COinS)*. URL: <http://ocoins.info/> Stand: 31.12.2007

Open-Source-Software. In: Meyers Lexikon Online. URL: <http://lexikon.meyers.de/meyers/checked/Open-Source-Software> Stand: 24.7.2008

OpenOffice.org (Hrsg.): *Bibliographic: CiteProc - Main Page*. URL: <http://bibliographic.openoffice.org/citeproc/> Stand: 31.12.2007

OpenOffice.org (Hrsg.): *OpenOffice Developers Guide*. Kapitel 12.2: Data Sources in OpenOffice.org API. URL: <http://api.openoffice.org/docs/DevelopersGuide/checked/Database/Database.xhtml> Stand: 12.12.2008

Paepke, Andreas: *Digital Libraries: Searching is not enough : What We Learned On-Site*. In: *D-Lib Magazine*, 1996, Mai. ISSN: 1082-9873. URL: <http://www.dlib.org/dlib/may96/checked/stanford/05paepcke.html> Stand: 05.09.2006

Panyr, Jiri: *Thesauri, Semantische Netze, Frames, Topic Maps, Taxonomien, Ontologien - begriffliche Verwirrung oder konzeptionelle Vielfalt?* In: *Festschrift Zimmermann*

Payer, Margarete: *AACR2 - RAK : grundsätzliche Unterschiede*. URL: <http://www.payer.de/einzel/aacr2rak.htm> Stand: 09.07.2006

Payer, Margarete: *Computervermittelte Kommunikation. – Kapitel 113. OSI-Schicht 7: Application Layer – Anwendungsschicht ; Teil 3: Information-Retrieval-Protokoll Z39.50*. URL: <http://www.payer.de/cmccmcs1303.htm> Stand: 2.1.2011

Payer, Margarete ; Payer, Alois: *Datenbankaufbau : Skript. - Kapitel 7: Formate in bibliographischen Datenbanken*. URL: <http://www.payer.de/dbaufbau/dbauf07.html> Stand: 11.10.2006

Siebert, Irmgard: *Positionen zu RAK/AACR. Vortrag auf der Sitzung des DBV, Sektion IV, Heidelberg, 19.05.2005*. URL: http://www.bibliotheksverband.de/fileadmin/user_upload/Sektionen/sektion4/Tagungen/2005-05_Siebert.pdf Stand: 11.01.2011

Schneider, Wolfram: *Ein verteiltes Bibliotheks-Informationssystem auf Basis des Z39.50 Protokolls*. Berlin, Technische Universität Berlin, Fachbereich Informatik, Dipl.-Arb., 1999. URL: http://www.zib.de/groetschel/students/diplom_schneider.pdf Stand: 20.12.2008

Schütz, Thomas: *Dokumentenmanagement*. In: *KSS*.

Schwens, Ute; Liegmann, Hans: *Langzeitarchivierung digitaler Ressourcen*. In: *KSS*.

Surowiecki, James: *The Wisdom of Crowds. Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday, 2004; ISBN: 0385503865 bzw. 9780385503860.

Taube, Anke: *Konzept und Implementierung eines Moduls zum Abgleich lokaler Zeitschriftenbestände am DKFZ mittels Z39.50 Protokoll*. Darmstadt, Fachhochschule Darmstadt, Fachbereich Informations- und Wissensmanagement, Dipl.-Arb., 2004. URL: http://www.dkfz-heidelberg.de/spec/dipl_diss/anke.taube.dipl.pdf Stand: 10.12.2008

The International DOI Foundation (Hrsg.): *The Digital Object Identifier System*. URL: <http://doi.org/> Stand: 29.12.2007

-
- Thomson Reuters (Hrsg.): *RIS Format Specifications*. URL: http://www.refman.com/support/risformat_sample_04.asp Stand: 14.10.2006
- Torvalds, Linus: *Offener Brief von Linux-Entwicklern zum Thema "Software-Patente"*. E-Mail vom 21. September 2003 an den Präsidenten des Europäischen Parlaments, Herrn Pat Cox, und an die Mitglieder des Europäischen Parlaments. Archivierte Kopie unter http://www.ffi.org/patentit/patents_torvalds_cox.html Stand: 2.10.2010
- UNESCO (Hrsg.): *Recommendation concerning the International Standardization of Statistics Relating to Book Production and Periodicals*, II. 6. a, URL: http://portal.unesco.org/en/ev.php-URL_ID=13068&URL_DO=DO_TOPIC&URL_SEC TION=201.html Stand: 10.05.2008
- U.S. National Library of Medicine (Hrsg.): *MEDLINE®/PubMed® XML Data Elements*. URL: http://www.nlm.nih.gov/bsd/licensee/data_elements_doc.html Stand: 12.12.2008
- Wagner-Döbler, Roland: *Umberto Ecos Betrachtung einer benutzerfeindlichen Bibliothek – 25 Jahre danach*. In: *Festschrift für Walther Umstätter*; CD-ROM
- Wenz, Christian: *AJAX. schnell + kompakt*. Frankfurt am Main: EntwicklerPress, 2006. - ISBN 3-935042-92-2
- Werner, Max ; Gramm, Gerhard: *Literaturflut – Informationslawine – Wissensexplosion. Wächst der Wissenschaft das Wissen über den Kopf?* URL: <http://www.fkf.mpg.de/ivs/literaturflut.html>. Stand: 05.9.2006
- Wiegand, Dorothee: *Gut zitiert ist halb geschrieben*. In: *c't* (2006), Nr. 7, S. 160-165
- Wikimedia Foundation Inc. (Hrsg.): *Arabischer Name*. URL: http://de.wikipedia.org/w/index.php?title=Arabischer_Name&oldid=47726855
- Wikimedia Foundation Inc. (Hrsg.): *Binärformat*. URL: <http://de.wikipedia.org/w/index.php?title=Bin%C3%A4rformat&oldid=6801306>
- Wikimedia Foundation Inc. (Hrsg.): *Datenformat*. URL: <http://de.wikipedia.org/w/index.php?title=Datenformat&oldid=21368358>
- Wikimedia Foundation Inc. (Hrsg.): *Escape-Sequenz*. URL: <http://de.wikipedia.org/w/index.php?title=Escape-Sequenz&oldid=19939353>

Wikimedia Foundation Inc. (Hrsg.): *Machine-Readable Catalog*. URL:

http://de.wikipedia.org/w/index.php?title=Machine-Readable_Catalog&oldid=21799071

Wikimedia Foundation Inc. (Hrsg.): *Normalisierung (Datenbank)*. URL:

[http://de.wikipedia.org/w/index.php?title=Normalisierung_\(Datenbank\)&oldid=53514147](http://de.wikipedia.org/w/index.php?title=Normalisierung_(Datenbank)&oldid=53514147)

Wikimedia Foundation Inc. (Hrsg.): *Peer-to-Peer*. URL: [http://de.wikipedia.org/w/index.php?](http://de.wikipedia.org/w/index.php?title=Peer-to-Peer&oldid=52452889)

[?title=Peer-to-Peer&oldid=52452889](http://de.wikipedia.org/w/index.php?title=Peer-to-Peer&oldid=52452889)

Wikimedia Foundation Inc. (Hrsg.): *Uniform Resource Identifier*. URL:

http://de.wikipedia.org/w/index.php?title=Uniform_Resource_Identifier&oldid=53228903

Wiesenmüller, Heidrun: *Informationsaufbereitung I: Formale Erfassung*. In: *KSS*

Womser-Hacker, Christa: *Zur Rolle von Eigennamen im Cross-Language Information Retrieval*. In: *Festschrift Zimmermann*

Ziegler, Cai: *Sinn oder nicht Sinn. Vom Suchen und Finden der Semantik im Web*. c't 21 / 2007.

9. Glossar

(AACR2) Anglo-American Cataloging Rules – Regeln zur formalen Erfassung und Katalogisierung des Bestandes von Bibliotheken. Gültig vor allem im englischsprachigem Bereich. AACR gilt als Fundament für viele nationale Regelwerke. Im Gegensatz zu RAK legt es einen großen Wert auf die Verwendung der lokal üblichen Namen und Bezeichnungen als Ansetzungsform.

AJAX steht für **Asynchronous JavaScript And XML** und bezeichnet eine neu Art der Kommunikation zwischen dem Benutzer (dem genutzten Browser) und dem Server sowie eine neue „Verhaltensweise“ von Internetanwendungen, die ähnlich einer lokalen Anwendung Funktionen oder Bestandteile bei Bedarf nachladen, statt den ganzen Inhalt des Fensters (alle Daten und Funktionen) neu anzufordern.

Auszeichnungssprache – Eine Auszeichnungssprache hat die Aufgabe, die logischen Bestandteile eines Dokuments zu beschreiben. Sie enthält Befehle zum Markieren typischer Elemente eines Dokuments wie Überschriften, Textabsätze, Listen, Tabellen oder Grafikreferenzen. HTML ist eine derartige Auszeichnungssprache, die für Dokumente im WWW verwendet wird. Dabei ist zu beachten, dass die Bedeutung und der zulässige Verwendungskontext (die Syntax) aller Auszeichner (en. Markups) in einem Schema (XMLS, Schematron, RelaxNG) oder einer Definition (DTD) hinterlegt wird.

Atomare Feldstrukturen sind solche Strukturen von Datenfeldern, die nicht in kleinere Teile zerlegt werden können. (Analog zur Prädikatenlogik verwendet: Eine Aussage ist dann atomar, wenn sie nicht weiter zerlegt werden kann.) Z. B.

Erscheinungsjahr (ohne Erscheinungsort), Buchtitel (ohne den Untertitel oder den Titel der Gesamtausgabe).

Bibsonomy – ein Dienst auf WWW-Basis für eine gemeinsame Verwaltung und Nutzung von Metadaten von Online- und Print-Publikationen. Die Erfassung und Erschließung in Form von Verschlagwortung geschieht auf Community-Basis und wird somit zu Social-Bookmarking-Systemen gezählt.

CEN, DIN, ISO – Normierungsgremien auf nationaler (Deutschland), EU und internationaler Ebene.

International Organization for Standardization (ISO) entstand 1947. Sie entwickelt unter Beteiligung der meisten Länder Standards, welche als international geltende Normen verabschiedet werden. In Deutschland werden die ISO-Normen in den nationalen Kontext umgesetzt und verabschiedet. Dies geschieht durch das **Deutsche Institut für Normung e. V.** – kurz **DIN** – (gegründet 1917), welches auch die deutschen Interessen bei ISO vertritt. Zwischen der nationalen und der internationalen Ebene gibt es seit 1961 das **Europäische Komitee für Normung (CEN)**. Im Bereich der Elektrotechnik existieren weitere übernationale Organisationen, die Normen für ihren Fachbereich verabschieden.

Citation Style Language (CSL) – beschreibt in einer XML-Datei die Formatierungsanweisungen für Quellenangaben. Sie wurde von Bruce D'Arcus definiert und wird z.B. in Firefox' Zotero-Add-On eingesetzt. CiteProc setzt diese in Form von XSL um.

cite-key (en.) oder "**Zitierschlüssel**" ist ein Kürzel aus Autornennamen, Titel und/oder Jahreszahl, die Stellvertretend für die gesamte bibliographische Beschreibung genannt wird.

Client-Server-Modell ist ein in der Informatik entwickeltes Modell, in dem ein Rechner die Verbindung initiiert und Anfragen sendet (der sog. Client), auf die ein anderer Rechner mit Antworten reagiert (der sog. Server).

Context Objects in Spans (COinS) ist eine Methoden zur Einbettung von bibliographischen Daten in HTML-Dateien unter Nutzung des OpenURL-Standards. Diese werden im title-Attribut des span-Elementes notiert.

Contextual Query Language (CQL) ist eine formale Sprache für die Beschreibung von Anfragen. Ihr Nutzungsgebiet ist die Übermittlung von Suchanfragen an heterogene

Datenbanksysteme (ähnlich dem Bib-1 in Z39.50). Es ist in vielen Kontexten verwendbar und ist nicht auf bestimmte Feldtypen begrenzt.

Ein **Dateiformat** legt fest, wie die in einer Datei abzuspeichernden Daten geordnet sein müssen, damit sie von einem Programm verstanden und verarbeitet werden können. Obwohl es nicht ganz korrekt ist, wird in dieser Arbeit der Begriff „**Datenformat**“ dem Dateiformat gleichgesetzt. Es soll dem Missverständnis vorgebeugt werden, dass unter Dateiformat nur Dateien aber keine Datenströme verstanden werden.

EndNote – eine PC-Anwendung der Firma Thomson Reuters zur persönlichen Literaturverwaltung mit cite-while-you-write- und Z39.50-Client-Funktion. Aufgrund hoher Nachfrage im geisteswissenschaftlichen Bereich auch im Rahmen einer Campus-Lizenz an der Universität des Saarlandes angeboten. EndNote bietet eine große Auswahl an international verwendeten Zitierstilen sowie Unterstützung für unterschiedliche Datenformate. In dieser Arbeit wird die Version 4 beschrieben.

Extensible Stylesheet Language (XSL) ist eine XML-basierte Transformationssprache zur Definition von Darstellung für XML-Dokumente. Ähnlich wie CSS (Cascading Style Sheets) beschreibt sie den Darstellungsmodus einzelner Elemente. XSL wurde vom W3C im Oktober 2001 verabschiedet und Enthält XSLT, XSL-FO sowie XPath.

(FTP) File Transfer Protokoll ist ein Protokoll zur Übertragung von binären und Textdaten welches bei Schmalband-Internetzugängen gerne benutzt wird, da es durch die Nutzung des UDP statt des TCP auf Integritätsprüfung von Dateien verzichtet und somit höheren Netto-Datendurchsatz erreicht als HTTP. Es ist ein sehr einfaches Protokoll und bietet Vorteile nur bei reinen Übertragung von Dateien.

Functional Requirements for Bibliographic Records (FRBR) – ein von IFLA entwickeltes Model zu Darstellung von Relationen zwischen Personen, Organisationen und bibliographischen Entitäten untereinander. Dabei wird zwischen dem Werk, einem Ausdruck (wie Übersetzung, Interpretation), der Manifestation (wie Auflage oder Erscheinungsweise) und dem Exemplar unterschieden. Vgl. dazu http://www.ifla.org/VII/s13/frbr/frbr_current_toc.htm und DNB, Arbeitsstelle für Standardisierung (Hrsg.): *Funktionelle Anforderungen an bibliographische Datenätze. Abschlussbericht der IFLA Study Group on the Functional Requirements for Bibliographic*

Records . urn:nbn:de:1111-20040721195 S. 12 f. FRBR ist die Grundlage für das neu entstehende Regelwerk Resource Description and Access (RDA), welches AARC2 und MAB2 ablösen soll. Vgl. Library of Congress (Hrsg.) : *Testing Resource Description and Access (RDA)*. URL: <http://www.loc.gov/bibliographic-future/rda/> Stand: 13.08.2011

(Web-) **Gateway** (de. Protokollumsetzer) ist bedeutet eine Internetseite, die nur wenige eigene Informationen anbietet, aber zu solchen hinführt oder solche auf Anfrage aus fremden Quellen zusammenstellt.

Google-Scholar ist eine von Google Inc. entwickelte und betriebene Suchmaschine für Recherche wissenschaftlicher Literatur. Sie bietet Verlinkung zum Volltext bei Online-Quellen, Metadaten, Export-Funktionen für diese (BibTeX, RIS und EndNote) sowie automatisch generierte Zitationsindexe (i10-, h- und Gesamtzitationsindex der letzten 5 Jahre).

Granularität ist das Maß für die Unterteilung der Daten. Ein System verfügt über hohe Granularität, wenn die Datenfelder so eng gefasst sind, dass sie im verwendeten Kontext nicht weiter unterteilt werden können oder müssen.

Harvesting (Webharvesting) ist ein automatisches Einsammeln von Internetseiten mittels eines Roboters (en. "crawler").

Hyperlink (oder kurz **Link**) - eine gerichtete oder ungerichtete Verbindung zwischen zwei Hypertext-Knoten. Eine solche Verweisung auf (X)HTML-Seiten im WWW kann fakultativ typisiert oder mit weiteren Informationen angereichert werden.

ILTIS – **Integrierte Literatur-, Tonträger- und Musikalien-Informationssystem** der Deutschen Nationalbibliothek. In seinem Zentralkatalog findet man neben den Normdaten ca. 5. Millionen Titel der Deutschen Nationalbibliographie seit 1945.

International Standard Bibliographic Description (ISBD) ist ein internationaler Standard für die Beschreibung von Werken wie Monographien (ISBD-M) oder kartographischem Material (ISBD-CM). Mit Hilfe von wenigen Deskriptionszeichen werden darin die grundlegenden bibliographischen Merkmale eines Werkes beschrieben.

(ISSN-L) **ISSN-Linking** ist eine Weiterentwicklung des ISSN, die es ermöglicht, parallele Ausgaben (wie Print, E-Book, CD-ROM)

miteinander zu verknüpfen. Die ISSN-L ist ein Identifikator, der für alle Medienarten desselben Werkes gilt und wird zusammen mit der üblichen ISSN verwendet. Sie wurde in ISO 3297:2007 (als zwei vierstellige und miteinander durch ein Bindestrich verbundene Zahlenblöcke die der Zeichenfolge „ISSN-L“ folgen) definiert.

Linkresolver – ein System zur Verfügbarkeitsanzeige von Publikationen, der durch die Suche der übergebenen Metadaten (z.B. openURL) im lokalen Bestandsdaten gemäß der Lizenzinformationen einen direkten Hyperlink auf passende Dienste generiert.

Lokale Verarbeitung – Unter lokaler Verarbeitung von Daten versteht man jede Art der Datenverarbeitung wie Eingabe, Suche, Lesen, Speichern und Umwandlung der Inhalte einer oder mehrerer Dateien, die immer auf einem und demselben Rechner geschieht.

Machine-Readable Cataloging (MARC, MARC21, UKMARC, UNIMARC, USMARC etc.) ist ein Text-basierendes Datenformat für den Austausch bibliographischer Daten. Es wurde unter der Federführung der Library of Congress entwickelt und implementiert den ISO 2709-Standard „Format für Informationsaustausch“ sowie sein amerikanisches Gegenstück ANSI/NISO Z39.2. MARC wurde in einigen Ländern weiterentwickelt. Daraus entstanden: **USMARC** (für die Vereinigten Staaten), **CAN/MARC** (für Canada), **MARC21** (die harmonisierte Version von USMARC und CAN/MARC, die diese Versionen ersetzt), **UKMARC** (für England), **INTERMARC** (der Französischen Nationalbibliothek), **NORMARC** (für Norwegen), **DANMARC** (für Dänemark) und **UNIMARC** (eine internationale Version der IFLA von 1997, wird jedoch fast ausschließlich in Europa verwendet).

Maschinelles Austauschformat für Bibliotheken (MAB, MAB2) – Ein in Deutschland von der Deutschen Nationalbibliothek entwickeltes (nicht MARC-kompatibles) Format für den Austausch bibliographischer Metadaten. Das **MAB** basiert auf DIN 1506 (entspricht ISO 2709). Die neueste Version des MAB-Formats stammt von 1995 und wird als **MAB2** bezeichnet. Unter dem Begriff MAB2 versteht man seit 1998 folgende Formate: **MAB-Titel** (Format für bibliographische Daten), **MAB-PND** (Format für Personennamen), **MAB-GKD** (Format für Körperschaftsnamen), **MAB-SWD** (Format für Schlagwörter), **MAB-Lokal** (Format für lokale Daten), **MAB-Adress** (für Adress- und Bibliotheksdaten) und **MAB-NOTAT** (Format für Notations- und Klassifikationsdaten).

Mashups – ein Verfahren der Zusammenstellung von Daten mehrerer Quellen im Internet. Die technische Basis dafür stellen die sog. APIs (Programmierschnittstellen). Mashups stellen einen wesentlichen Teil des "Web 2.0" dar.

Metadaten sind eigenschaftsbeschreibende Daten über andere Daten. Strukturellen Metadaten geben z.B. an, wie die beschriebenen Daten beschaffen sind, administrative Metadaten hingegen, wem die Daten gehören und wer Zugriff darauf hat. Im Bereich der Veröffentlichungen (im Print- sowie online-Bereich) wird von **bibliographischen Metadaten** gesprochen.

Eine **Ontologie** – der Terminus ist der Philosophie entlehnt und beschreibt die Dinge einer bestimmten Domäne. Eine **Web-Ontologie** beschreibt ein komplexes Gebilde aus Datenformat (OWL), Datenbasis und Verknüpfungslogik.

Das **Open System Interconnection (OSI)** Referenzmodell ist ein von ISO standardisiertes Schichtenmodell, welches die unterschiedlichen Protokollfunktionen eines Computernetzwerks (TCP/IP, UDP/IP) hierarchisch anordnet. Dazu zählen (aufsteigend und aufeinander aufbauend): die Bitübertragungsschicht (physikalische Verbindung), die Sicherungsschicht (Protokolle für die Überprüfung der Zustellung), die Vermittlungsschicht (Adressierung und Routing = das IP-Protokoll), die Transportschicht (TCP, UDP), die Kommunikationsschicht (Dienstprotokolle wie FTP oder HTTP, die die Kommunikation steuern), die Darstellungsschicht (die Formate für die zu übertragenden Daten) und die Anwendungsschicht (Anwendungen, die die Daten verarbeiten). Bei der Kommunikation über WWW werden (absteigend) folgende Protokolle und Formate verwendet: Hypertext, HTML, TCP, IP, ISDN (bzw. ADSL oder Modem), Netzkabel. Vgl. dazu: Bekavac, Bernard: *Informations- und Kommunikationstechnologien*. In: KSS, S. 336f

openURL – ein Standard zur Link-Auflösung, der für jedes Dokument einen abstrakten Namen vergibt. Dieser Name wird von einem Linkresolver auf eine URL aufgelöst. OpenURL wurde von der National Information Standards Organization entwickelt und unter ANSI/NISO Z39.88 standardisiert.

Ein **Parser** ist ein Programm, das Zeichenketten analysiert. Ein Parser bereitet die Daten für weitere Verwendung im Programm vor, indem es die Eingaben aufteilt und den erwarteten Datentypen zuordnet. Beim XML werden zwei Typen von Parsern verwendet: der **einfache** – für das Lesen von XML-Dateien

bzw. -Datenströmen und der **validierende**, der die gelesenen XML-Daten gegen die hinterlegten XML-Schemata validiert.

Peer-to-Peer (P2P) steht für ein relativ neues Modell der Kommunikation, in dem jeder Rechner sowohl Anfragen als auch Antworten generieren kann, zu einem anderen Teilnehmer (sog. „Peer“) Anfragen schicken kann (wie ein Client) als auch die an ihn gerichtet Anfragen beantworten kann (wie ein Server). In diesem Modell fehlt die zentrale Instanz eines Servers, der als einziger die Antworten gibt. Vielmehr entwickeln sich autonome und gleichberechtigte Peers zu einem dezentralen („anarchischen“) Netzwerk, in welchem die gesuchten Informationen oft auf mehreren Computern gefunden werden können.

Protokolle regeln die Kommunikation in Rechnernetzen. Solche Protokolle werden im Vorfeld der Entstehung einer Client- oder Server-Software definiert und können in unterschiedlichen Versionen vorliegen, wobei die Abwärtskompatibilität meist gegeben ist. In dieser Arbeit wird das Wort **Austauschprotokoll** synonym zum Wort Protokoll verwendet.

Quellenangabe oder **Quellennachweis** gibt den Ort an, aus dem ein Zitat stammt. Je nach Veröffentlichungsort oder Forschungsbereich kann die Quellenangabe sehr unterschiedlich formatiert werden. Neben **ISBD** (International Standard Bibliographic Description) existieren viele weitere sog. **Zitierstile** wie **DIN 1505-2**, **ISO 690**, **APA** etc.

Reference Manager – eine PC-Anwendung der Firma Thomson Reuters zur persönlichen Literaturverwaltung mit cite-while-you-write- und Z39.50- bzw. PubMed-Client-Funktionen. Anders als EndNote (vom selben Hersteller) ist diese Software auf die Verwaltung großer Datenmengen ausgelegt und bietet Unterstützung bei der Erfassung von Personennamen sowie eine große Auswahl an international verwendeten Zitierstilen. In dieser Arbeit wird die Version 10 und 11 beschrieben.

Regeln zur Alphabetischen Katalogisierung (RAK) – Ein Regelwerk zur formalen Erfassung von Schrifttum. RAK gibt Normen vor, nach welchen der Bestand der Bibliotheken und Archiven einheitlich katalogisiert werden soll. Es gilt nur im deutschsprachigen Bereich. Es entstand 1976 / 1977 und löste die „Preußischen Instruktionen“ ab. Neben der RAK-WB (für Wissenschaftliche Bibliotheken) existieren folgende weitere RAK-Regelwerke: Musik, Karten, Nichtbuchmaterialien, für Parlaments- und Behördenbibliotheken, für alte Drucke und zur Erschließung

von Nachlässen. Ähnliche Katalogisierungsregeln für das englischsprachige Bereich stellt das AACR2 dar.

Relax NG ist eine vereinfachte Grammatik für die Definition von XML-Formaten. Sie basiert auf XML und zählt zu den XML-Schemas. Anders als bei XML-Schema ist es nicht möglich genaue Anzahl der Vorkommen eines Elementes festzulegen oder Regeln für den Inhalt von Elementen und Attributen aufzustellen. Wegen der Vereinfachung erfreut sich dieses Schema einer großen Beliebtheit.

Ressource Description Framework (RDF) ist ein XML-basiertes Format, welches von W3C entwickelt wurde. Es stellt ein besonders einfaches Datenmodell dar, welches mit Hilfe von Tripeln Ressourcen beschreibt (Tripel ist ein dreigliedriger Satz in Form von: "Ressource X hat die Eigenschaft Y mit einem Wert Z").

Schlagwortwolke, Begriffswolke, Tag-Wolke (en. „tag cloud“) – Eine Art der zweidimensionalen Visualisierung der Relevanz von einzelnen Wörtern in einem bestimmten Umfeld (z.B.: verwandte Begriffe wie das gesuchte Wort etc).

Schlagwortnormdatei – (SWD) ist ein kontrolliertes Schlagwortsystem, der bei der Sacherschließung in den deutschen Bibliotheken eingesetzt wurde. Die Aufnahme in die SWD geschieht entsprechend den Regeln für den Schlagwortkatalog (RSWK). Mit der Inbetriebnahme der Gemeinsamen Normdatei (GND) wird die SWD abgelöst.

Search / Retrieve via URL (SRU) ist ein Recherche-Protokoll für Suchen im Internet, welches für die Suche CQL (Contextual Query Language) und die Antwort in einem XML-Dialekt ausliefert. Es setzt das Z39.50-Protokoll auf Basis des HTTP um.

Standard Generalized Markup Language (SGML) – eine Metasprache, mit deren Hilfe verschiedene Auszeichnungssprachen definiert werden können. Sie wurde als ISO 8879 verabschiedet und dient als Grundlage für viele verschiedene Dateiformate wie z. B. TeX oder XML.

Structured Query Language (SQL) – eine weit verbreitete Abfragesprache für Datenbanken, die sich an der natürlichen Sprache (englisch) orientiert.

Thesaurus – ein präskriptives System von Begriffen und Bezeichnungen (meist natürlichsprachlich) für die Erschließung und Retrieval in einem Dokumentationsgebiet. Durch terminologische Kontrolle werden Homonyme, Synonyme und Polyseme durch eine Vorzugsbenennung aufeinander bezogen bzw. voneinander unterschieden. Zusätzlich werden die Beziehungen zwischen den Begriffen hergestellt und beschrieben.

Trunkierung – Methode zur Maskierung von fakultativen oder unbekanntem Teilen eines Wortes oder einer Wortgruppe im Retrieval, die mit Stellvertreterzeichen den Ort und die Menge angibt.

Unicode ist eine Zeichenkodierung (**ISO 10646**), die mehr als ein Byte pro Zeichen verwenden kann und damit mehr als 256 verschiedene Zeichen abbilden kann. Die bekanntesten Unterarten des Unicode sind **UTF-8** und **UTF-16** (UCS2).

Uniform Resource Identifier (URI) (engl. „einheitlicher Bezeichner für Ressourcen“). Es ist ein Identifikator für elektronische oder physische Ressourcen. URI besteht aus dem Namen des Schema, begleitet von Doppelpunkt und dem schemaspezifischen Teil. Z. B.: <http://de.wikipedia.org/wiki/URL> oder <urn:isbn:3598116748>. URI wurde definiert in **RFC 3986**.

XML - Extensible Markup Language ist eine SGML-basierte, von W3C entwickelte Metasprache zur Definition von Auszeichnungssprachen. XML ist sozusagen die Grammatik, die es erlaubt, eine beliebige Sprache (die ähnlich einer natürlichen Sprache aus Semantik und Syntax besteht) zu entwerfen. Eine Besonderheit von XML ist die Fähigkeit, viele solcher „Sprachen“ (in ihrem Verwendungs- und Bedeutungskontext) in einem Dokument nutzbar machen zu können - die sog. **Namensräume** (en. "namespaces"). Ein XML-Wortschatz wird auch als **XML-Format** bzw. **Dialekt** bezeichnet.

XML-Schema (des W3C) ist eine XML-basiertes Format, welches der Definition von XML-Formaten dient. XML-Schema ist sehr ausdrucksstark und kann neben der Reihenfolge und Anzahl von Elementen ihren Inhalt genau beschreiben (z. B. "Datum", "Jahreszahl", "Ganzzahl" etc). Es erfreut sich einer sehr guten Unterstützung und wird meist für XML-Formate verwendet, die Daten mit hoher Genauigkeit aufnehmen sollen.

XPath ist eine Sprache, die der Identifizierung von XML-Elementen dient. Die Schreibweise erinnert an Pfadangaben eines Unix-Systems, die entweder absolut (mit einem „/“ am Anfang) oder relativ (ohne „/“ am Anfang) sein können. XPath wird vor allem für die

Transformation von XML mit Hilfe von XSL-T oder XSL-FO benötigt.

XSL - Formatting Objects (XSL-FO) ist eine Programmiersprache, die beschreibt, wie ein XML-Dokument in andere (Darstellungsorientierte) Dateitypen (wie PDF, PostScript oder MS Word) umgewandelt werden soll. XSL-FO wurde vom W3C innerhalb von XSL spezifiziert.

XSLT-Prozessor – Software die XML-Daten anhand der XSL-Vorlagen verarbeitet. Diese ist als eigenständiges Programm, eine Programmbibliothek oder als Funktion in modernen Webbrowsern realisiert.

XSL Transformations (XSLT) – ist eine Programmiersprache zur Transformation von XML-Dateien in andere Text-basierte Dateien wie HTML, SVG, LaTeX. XSLT ist ein Teil des XSL.

Z39.50 – Protokoll für den datenbankunabhängigen Information Retrieval. Es ist ein nordamerikanischer (**ANSI/NISO Z39.50**) und internationaler (**ISO 23950**) Standard, welcher seit 1984 von der Library of Congress entwickelt und ständig verbessert wurde (Version 1: 1988; Version 2: 1992). Es ist ein Protokoll der siebten Ebene des **OSI-Modells** (s. **OSI**) welches auf dem TCP/IP-Protokoll aufbaut. Eine durch ein Z39.50-Protokollfähiges Programm erstellte Anfrage wird nach der Authentifizierung auf dem Server an ein Programm übergeben, welches die gewünschten Daten in der Datenbank sucht, sie entsprechend einem Datenformat (genauer „Übertragungsformat“ wie z. B. **UNIMARC**) zusammenstellt und an den Benutzer über dieses Protokoll ausliefert.

ZACK – ein verteiltes Bibliotheks-Informationssystem auf Basis des Z39.50-Protokolls, welches über eine WWW-Schnittstelle verwendet werden kann. Anders als gewöhnliche Z39.50-Gateways bietet es eine verteilte Suche auf mehreren (ausgewählten) Z39.50-Gateways und Deduplizierung der Treffer im Ergebnis.

ZING steht für „**Z39.50 International New Generation**“ - ein Nachfolgemodell, welches an die Funktionalität von Z39.50 anknüpft. Seit 2009 hat sich der Begriff "**SRU/W**" anstelle des ZING durchgesetzt.

Zitat – die wörtliche Übernahme einer fremden Aussage, die mit Anführungszeichen markiert wird und mit Angaben zur Quelle (s. **Quellenangabe**) versehen wird. In dieser Arbeit wird der

Begriff "**Zitieren**" im weiteren Sinne verwendet: als Synonym für das Erstellen von Quellenangaben (ungeachtet dessen, ob der Inhalt der Quelle wörtlich wiedergegeben wird oder nicht).

Zitierstil ist eine formalisierte Beschreibung eines bibliographischen Werkes, die das Beschaffen oder das Wiederfinden ermöglicht. Je nach Forschungsbereich (bzw. Wissenschaftlergemeinschaft), Verlag oder Zeitschrift (in der ein Aufsatz erscheint) werden unterschiedliche Zitierstile verwendet: z. B. **APA**, **Harvard** oder **DIN 1505-2**.