



**M A G I S T E R A R B E I T**

**Feldstudie zur praktischen Anwendung  
von Metadaten Standards  
in Digitalen Bibliotheken und Archiven**

ausgeführt am Institut für  
**Distributed and Multimedia Systems**  
Universität Wien

unter Anleitung von Univ. Prof. Dr. Wolfgang Klas  
und unter Mitwirkung von Mag. Dipl.-Ing. Bernhard Haslhofer

durch  
**Maximilian Schwarzmaier**  
Matr. Nr.: 9225414  
Studienkennzahl: E175  
A-1130 Wien, Hanselmayergasse 17/31F

Wien, im Jänner 2008

## **Abstract**

Immer mehr Informationen und Daten werden geschaffen und digital publiziert. Die vorliegende Arbeit beschäftigt sich mit der aktuellen Verwendung von Metadaten, die helfen, Daten zu speichern, zu finden, auszutauschen und verwenden zu können. Es gibt zahlreiche Organisationen, die Metadatenstrukturen und -standards entworfen haben. Es sind dadurch eine Vielzahl von Metadaten Austauschformate entstanden. Der Kern der Arbeit wird sein, herauszufinden, welche Metadaten schemata genutzt werden. Die Analyse wird einerseits mit einer entwickelten Java Applikation verwirklicht. Mit dieser Anwendung werden verwendete Online Metadaten via Internet gefiltert, analysiert und die Verteilungen in Diagrammen dargestellt. Andererseits wird mit Hilfe eines Fragebogens die aktuelle Benützung von Metadatenstandards festgestellt. Die Ergebnisse werden ebenso ausgewertet und abgebildet.

# INHALTSVERZEICHNIS

<b>1 EINFÜHRUNG.....</b>	<b>7</b>
1.1 WAS SIND METADATEN .....	8
1.2 TYPEN VON METADATEN .....	9
1.3 INTRINSISCHE UND EXTRINSISCHE DATEN .....	10
1.4 DATEN GRANULARIÄT .....	10
1.5 METADATEN SPEICHERUNG .....	10
1.6 ANFORDERUNG AN METADATEN .....	11
1.7 INFORMATIONSFINDUNG .....	12
1.7.1 Das Ordnen elektronischer Ressourcen.....	12
1.7.2 Austausch von Metadaten .....	12
1.7.3 Transformation von Metadaten.....	13
1.7.4 Metadaten Registrierung.....	13
1.8 IDENTIFIKATION VON METADATEN.....	13
1.9 ARCHIVIERUNG UND ERHALTUNG VON METADATEN .....	13
1.10 METADATEN STRUKTURIERUNG .....	14
1.11 ERZEUGUNG VON METADATEN.....	14
1.11.1 Beispiel an der Fa. SAP .....	14
1.11.2 Beispiel an der Fa. Adobe.....	15
1.11.3 Erzeugungswerkzeuge.....	15
1.11.4 Qualitätskontrolle .....	15
1.12 ROLLE DER METADATEN IM MULTIMEDIABEREICH .....	16
1.12.1 Bedarf an Metadaten.....	16
1.12.2 Medien spezifische Daten.....	16
<b>2 METADATENSTANDARDS - EIN ÜBERBLICK.....</b>	<b>17</b>
2.1 VERWENDUNG IN DIGITALEN BIBLIOTHEKEN.....	17
2.2 STANDARDISIERTE BESCHREIBUNGSSCHEMATA .....	17
2.2.1 Dublin-Core .....	18
2.2.2 MARC und MARCII - Machine Readable Cataloging Record.....	20
2.2.3 EXIF - austauschbares Bildformat für digitale Standbilder .....	23
2.2.4 LOM - Learning Object Metadata .....	24
2.2.5 RAK und RAK-WB - Regeln für alphabetische Katalogisierung.....	24
2.2.6 MAB und MAB2 - Maschinelles Austauschformat für Bibliotheken .....	26
2.2.7 METS - Metadata Encoding and Transmission Standard.....	27
2.2.8 RSWK - Regelwerk für Schlagwortkatalogisierung .....	29
2.2.9 PICA - Project of Integrated Catalogue Automation.....	30
2.2.10 International Standard Organisation (ISO) - Standard 19115 .....	31
2.2.11 ISAD(G)- (General) International Standard Archival Description, Second edition.....	31

2.2.12 RFC1807- Request for Comments.....	32
2.2.13 JUNII und JUNI2.....	33
2.2.14 ETD-MS -Electronic Theses and Dissertations Metadata Set .....	34
2.2.15 QDC - Qualified Dublin Core.....	35
2.2.16 EPICUR - Enhancement of Persistent Identifier Services- Comprehensive Method for Unequivocal Resource Identification .....	36
2.2.17 DIDL- Digital Item Declaration Language .....	37
2.2.18 Metadaten in HTML.....	37
2.2.19 "Data Dictionary" - Technische Metadaten für digitale Bilder .....	37
2.2.20 RDF - Resource Description Framework .....	38
2.2.21 XMP - Extensible Metadaten Plattform.....	39
2.2.22 MPEG-7.....	44
2.2.23 MPEG-21 .....	46
<b>3 FELDSTUDIE .....</b>	<b>47</b>
3.1 AUFBAU.....	47
3.1.1 Ausgewählte Fragen .....	47
3.2 TEILNEHMER.....	52
3.3 ZIEL .....	52
3.1 AUSWERTUNGEN DER FELDSTUDIE MIT DIAGRAMMEN.....	53
3.1.1 Auswertung Top 10 Metadatenstandards.....	53
3.1.2 Auswertung nach Tätigkeitsumfeld .....	54
3.1.3 Auswertung nach Domänen .....	54
3.1.4 Auswertung nach Metadatenstandards .....	58
3.1.5 Auswertung: Zusammenführung von Systemen.....	68
<b>4 ONLINE ANALYSE .....</b>	<b>73</b>
4.1 ZIEL .....	73
4.2 IMPLEMENTIERUNG.....	73
4.3 AUSWERTUNG DER ONLINE ANALYSE .....	74
<b>5 VERGLEICH ONLINE ANALYSE UND FELDSTUDIE.....</b>	<b>82</b>
<b>6 DISKUSSION .....</b>	<b>86</b>
6.1 ZUKUNFT VON METADATEN UND -STANDARDS .....	86
6.2 KRITIK .....	86
6.2.1 Profit.....	86
6.2.2 Objektivität.....	87
6.2.3 Grenzen setzen .....	87
6.2.4 Einsatz.....	87
6.3 ÜBERLEGUNGEN .....	88

<b>7 LITERATUR UND WEBRESSOURCEN .....</b>	<b>89</b>
<b>APPENDIX A .....</b>	<b>95</b>
<i>Das Java Programm .....</i>	95
<b>APPENDIX B.....</b>	<b>108</b>
<i>Eine Auflistung einiger NISO Standards.....</i>	108
<b>APPENDIX C .....</b>	<b>111</b>
<i>Beispiele für die Definition der 1.1 DCMI Metadaten Begriffe .....</i>	111
<b>APPENDIX D .....</b>	<b>116</b>

## ABBILDUNGSVERZEICHNIS

ABBILDUNG 1-1 METADATEN BESCHREIBUNG.....	8
ABBILDUNG 2-1 METS METADATENSTANDARD.....	28
ABBILDUNG 2-2 EINFACHE XMP TYPEN.....	41
ABBILDUNG 2-3 XMP STRUKTUREN.....	41
ABBILDUNG 2-4 XMP ARRAYS.....	42
ABBILDUNG 2-5 EIGENSCHAFTSQUALITÄTEN .....	42
ABBILDUNG 2-6 "EIGENSCHAFTSQUALIFIZIERER".....	43
ABBILDUNG 3-1 TEILNEHMER DER FELDSTUDIE .....	52
ABBILDUNG 3-2 ANTEILE ÜBER ALLE BRANCHEN UND TOP 10 METADATENSTANDARD .....	53
ABBILDUNG 3-3 VERWENDETE METADATEN FORMATE NACH DOMÄNE "BÜCHER" .....	54
ABBILDUNG 3-4 TOP 10 METADATENSTANDARDS / BIBLIOTHEKEN.....	55
ABBILDUNG 3-5 TOP 10 METADATENSTANDARDS / FORSCHUNG UND LEHRE.....	55
ABBILDUNG 3-6 TOP 10 METADATENSTANDARDS / MEDIEN UND VERLAGE.....	56
ABBILDUNG 3-7 TOP 10 METADATENSTANDARDS / VERWALTUNG .....	56
ABBILDUNG 3-8 TOP 10 METADATENSTANDARDS / KULTUR .....	57
ABBILDUNG 3-9 TOP 10 METADATENSTANDARDS / IT.....	57
ABBILDUNG 3-10 DUBLIN CORE METADATENSTANDARD VERWENDUNG NACH BRANCHEN .....	58
ABBILDUNG 3-11 HAUSINTERNE METADATENSTANDARD VERWENDUNG NACH BRANCHEN.....	59
ABBILDUNG 3-12 RAK METADATENSTANDARD VERWENDUNG NACH BRANCHEN .....	60
ABBILDUNG 3-13 MAB METADATENSTANDARD VERWENDUNG NACH BRANCHEN .....	61
ABBILDUNG 3-14 MAB2 METADATENSTANDARD VERWENDUNG NACH BRANCHEN .....	62
ABBILDUNG 3-15 METS METADATENSTANDARD VERWENDUNG NACH BRANCHEN .....	63
ABBILDUNG 3-16 RSWK METADATENSTANDARD VERWENDUNG NACH BRANCHEN .....	64
ABBILDUNG 3-17 ISAD(G) METADATENSTANDARD VERWENDUNG NACH BRANCHEN.....	65

ABBILDUNG 3-18 ISO METADATENSTANDARDS (19115,19119,19139) VERWENDUNG.....	66
ABBILDUNG 3-19 RAK-WB METADATENSTANDARD VERWENDUNG NACH BRANCHEN .....	67
ABBILDUNG 3-20 AUSWAHL ANWENDUNGSSZENARIOEN .....	68
ABBILDUNG 3-21 BEDARF AN ZUSAMMENFÜHRUNG VON METADATEN .....	69
ABBILDUNG 3-22 ANWENDUNGSSZENARIOEN DER ZUSAMMENFÜHRUNG VON METADATEN.....	70
ABBILDUNG 3-23 INTERESSE AN ANGEBOTEN .....	71
ABBILDUNG 3-24 INTERESSE AN DIENSTEN .....	72
ABBILDUNG 4-1 TOP 10 METADATENSTANDARDS PRO OAI REPOSITORY .....	74
ABBILDUNG 4-2 VERTEILUNG VON METADATENSTANDARDS ÜBER ALLE OAI-PMH REPOSITORIES.....	75
ABBILDUNG 4-3 ANZAHL VON METADATENSTANDARDS IN OAI-PMH REPOSITORIES.....	76
ABBILDUNG 4-4 METADATENSTANDARDS PRO OAI-PMH REPOSITORY MIT MEHR ALS 4 STANDARDS .....	77
ABBILDUNG 4-5 METADATENSTANDARDS PRO OAI-PMH REPOSITORY MIT 4 STANDARDS.....	78
ABBILDUNG 4-6 METADATENSTANDARDS / REPOSITORIES MIT 3 STANDARDS .....	79
ABBILDUNG 4-7 6 METADATENSTANDARDS / REPOSITORIES MIT 2 STANDARDS .....	80
ABBILDUNG 4-8 6 METADATENSTANDARDS / REPOSITORIES MIT EINEM STANDARD .....	81
ABBILDUNG 5-1 ANTEILE ÜBER ALLE BRANCHEN UND TOP 10 METADATENSTANDARD .....	82
ABBILDUNG 5-2 TOP 10 METADATENSTANDARDS PRO OAI-PMH REPOSITORY .....	84

## TABELLENVERZEICHNIS

TABELLE 1-1 UNTERSCHIEDUNG DER INTRINSISCHEN UND EXTRINSISCHE EINTEILUNG, TYPOLOGIE VON METADATEN [10] .....	10
TABELLE 2-1 ORDNUNGSZAHLEN MARC STANDARD.....	22
TABELLE 2-2 ORDNUNGSZAHLEN MARC STANDARD .....	23
TABELLE 5-1 2 ERKLÄRUNG ZU DEN METADATENSTANDARDS IM DIAGRAMM.....	83
TABELLE 5-2 METADATEN-NAMESPACE ABKÜRZUNGEN .....	84

## **Einführung**

### **1 Einführung**

Der Themenbereich der Arbeit beschäftigt sich mit der aktuellen Verwendung von Metadaten in öffentlichen Institutionen, Bibliotheken und Multimedia Archiven in Österreich und Europa. Digitale Archive für Bücher, Bilder, usw. verwenden verschiedene Metadatenstandards, um Daten zu speichern, wiederaufzufinden, auszutauschen und verwenden zu können. Es gibt zahlreiche Organisationen; die viele Metadatenstrukturen, -schemata und -standards entworfen haben. Es sind dadurch eine Vielzahl von Metadaten Austauschformate entstanden.

Der Kern der Arbeit wird einerseits sein, herauszufinden welche Metadatenstrukturen und -schemata genutzt, und andererseits gilt es festzuhalten, wie Metadaten in Wirklichkeit ausgetauscht bzw. transformiert und integriert werden. Die Recherche zur aktuellen Anwendung von Metadaten wird mit Hilfe eines Fragebogens gemacht. Die Analyse für die Suche im Internet nach verwendeten Metadaten wird mit Hilfe eines eigens entwickelten Java Programms verwirklicht. Mit dem Programm werden verwendete Online Metadaten herausgefunden und in der Diplomarbeit aufgezeigt.

Die Motivation ist auch hier einerseits einen Überblick zu schaffen, welche Metadatenformen, -strukturen, -schemata und -standards in Bibliotheken und Institutionen in der Realität genutzt werden. Folgende Fragestellung werden in dieser Diplomarbeit beantwortet: Welche wichtigen Metadaten Schemas gibt es und welche werden in der Arbeitswelt benutzt? In welchem Ausmaß werden die standardisierten Metadaten Schemas von Institutionen angewendet? Welche Probleme bzw. welche Herausforderungen gibt es beim Austausch zwischen den Systemen? Welche Metadatenprotokolle werden Online verwendet? Auf technischer Ebene wird eine Java Applikation entworfen, welche das Internet nach verwendeten Metadaten Standards durchsucht. Diese Informationen werden ausgewertet, beschrieben und dargestellt.

## Einführung

### 1.1 Was sind Metadaten

Metadaten sind strukturierte Informationen, die Informationsressourcen beschreiben, erklären, und lokalisieren. Die Wiederauffindbarkeit, die effiziente Nutzung und die richtige Verwaltung der Ressourcen sind wichtige Punkte, warum Metadaten geschaffen wurden.

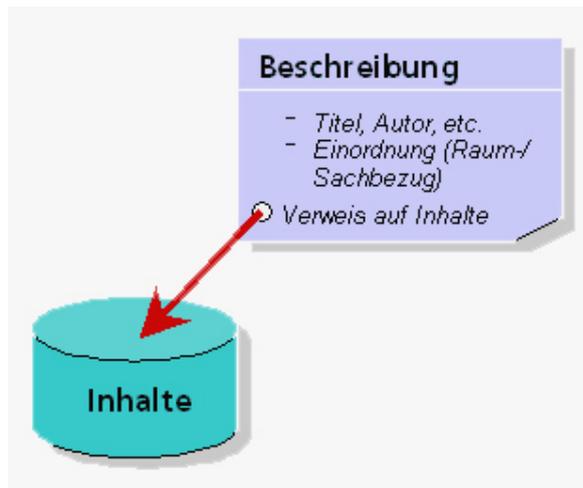


Abbildung 1-1 Metadaten Beschreibung

Metadaten sind also Daten, die Daten beschreiben, bzw. Informationen zu Informationsressourcen oder anders formuliert: "Daten über Daten". Diese Informationen werden zusätzlich zu den normalen Informationen generiert und gespeichert. Metadaten beinhalten semantische Informationen über digitalisierte und nicht digitalisierte Objekte. Metadaten werden auch unter den Gesichtspunkten des Findens und der Suche, der Nutzung, dem Zugang, der Beziehung zu anderen Objekten und der Einordnung der Objekte in einen größeren Rahmen gesehen.

Metadatenbeschreibungen umfassen typischerweise Felder wie "Titel" und "Autor" der Ressource. Weiters noch das Thema bzw. Stichwörter, eine inhaltliche Beschreibung, ein Erstellungs- bzw. Änderungsdatum, sowie Verweise auf andere Objekte zum Thema. Eine Ressource wird somit durch einige beschreibende Attribute dargestellt. Metadaten können direkt im Objekt gespeichert werden oder in einer eigenen Datei mit einem Link zu der Ressource.

Metadaten stellen allgemein sicher, dass Informationsressourcen zugänglich für Benutzer und Systeme bleiben. Metadaten sind die Basis für die Suche in der Ressourcenbeschreibung und sollten mit entsprechenden Verfahren der Informationsvermittlung, die auf einen effizienten und kostengünstigen Einsatz in elektronischen Netzen optimiert sind, verwendet werden. So soll in Zukunft eine Basis für ein maschinenverständliches Netz in Bezug auf Informationen aller Art geschaffen

## **Einführung**

werden. Diese soll durch die Nutzung geeigneter Technologien, wie beispielsweise XML und RDF erreicht werden. [23][2]

### **1.2 Typen von Metadaten**

Es gibt drei Hauptgruppen von Metadaten. Beschreibende Metadaten dienen einer Ressource zum Zweck des Wiederauffindens und der Identifikation. Auch gibt es die Struktur-Metadaten. Diese geben Aufschluss über die Zusammensetzung der Objekte, z.B. wie die einzelnen Seiten eines Buches ein Kapitel mit welcher Überschrift ordnet. Dann gibt es noch die administrative Metadaten, welche Informationen liefern, wie man Ressourcen am besten verwalten kann. Es existieren dazu die Rechteverwaltungsmetadaten, in denen die Urheberrechte der Informationsressource verwaltet werden und die Aufbewahrungsmetadaten, das Daten zur Archivierung und Speicherung sind. [3]

Eine Einteilung der Metadaten je nach Anwendungsgebiet ist eine andere Art der Strukturierung. Einerseits kann mit Bedingungen beschrieben werden, wie gewisse Ressourcen verwendet werden dürfen. Diese Bedingungen könnten etwa Listen enthalten, in der Zugriffsrechte bzw. Nutzungsbedingungen definiert sind. Zusätzlich existieren Administrative Daten, die angeben, wer für die Bereitstellung eines bestimmten Dokumentes verantwortlich sein soll, wann es geändert, erstellt, etc. wurde. Und es gibt eine Inhaltliche Bewertung von Metadaten. Die Plattform für Internet Content Selektion (PICS) beschäftigt sich unter anderem mit der Verwaltung von Zugriffsrechten auf gewisse Ressourcen, so kann z.B. je nach Art der Ressource der Zugriff auf Grund von Jugendschutzbestimmungen untersagt werden. Die Einführung der Bewertung (Rating) von Ressourcen spielt in diesem Zusammenhang eine wichtige Rolle. Auch gibt es einige Merkmale die von Bedeutung sind, zum Beispiel. die Herkunft und Ursprung der Inhalte werden deshalb benötigt, um angeben zu können woher bestimmte Daten kommen. Darüber hinaus sind die Daten über Beziehungen sehr wichtig, um z.B. auf Versionen eines Dokumentes in einer anderen Sprache verweisen zu können. Und ein weiterer wichtiger Punkt sind die strukturbeschreibenden Informationen, die darstellen wie auf komplexe Dokumente zugegriffen werden kann. [4][5][24]

## Einführung

### 1.3 Intrinsische und Extrinsische Daten

Eine andere zusätzliche Einteilung von Metadaten ist die intrinsische und extrinsische. Die intrinsische Betrachtungsweise stellt spezifische innere, zugehörige, eigene Werte des Objektes dar, die unabhängig von jedem anderem Objekt sind. Extrinsische Metadaten sind Informationen, welche jeder Ressource zugeordnet werden können. Es sind Daten, die unabhängig von der eigentlichen Information der Ressource sind.

	Manuell bestimmt	Automatisch bestimmt
Intrinsisch	Titel, Autor, Schlüsselworte, Kategorie, ...	Anzahl der Bilder, Satz von Bildern, Anzahl der Links, Erstellungsdatum, ...
Extrinsisch	Anmerkungen, Syntaxhervorhebung, ...	Anzahl der Zugriffe, Datum des letzten Zugriffs, Anzahl der Versionen, ...

**Tabelle 1-1 Unterscheidung der intrinsischen und Extrinsische Einteilung, Typologie von Metadaten [10]**

### 1.4 Daten Granularität

Ressourcen können auf verschiedenen Ebenen der Gliederungstiefe bzw. Datenhierarchie beschrieben werden, d.h. wenn viele Informationsbeschreibungen für wenige Daten existieren, ist das ein hoher Detaillierungsgrad. Andererseits kann eine Zusammenstellung vieler Informationsressourcen Metadaten entstehen lassen, die wenig Detail Informationen aufweisen. Dadurch ergibt sich eine Einteilung der Granularität. Die feinste Granularität ist somit der Bruchteil, weiters existiert eine Granularität auf einer quasi 1:1 Beschreibung von Metadaten zur Ressource. Zusätzlich kann es eine Beschreibung unter mehreren für eine größere Ressource geben, dass wiederum eine gröbere Granularität darstellt. Und eine ganze Sammlung von Ressourcen mit einem Metadatum zu beschreiben stellt die gröbste Einteilung von Granularität dar [8].

### 1.5 Metadaten Speicherung

Metadaten können von den eigentlichen Informationen getrennt oder zusammen in derselben Datei gespeichert werden. Manche Daten können oder sollten nicht mit den Metadaten in einer Datei gespeichert werden, d.h. sie müssen extern gespeichert werden. Die Metainformation wird dann eben extern vom Dokument gespeichert und

## **Einführung**

kann unabhängig abgerufen werden. Somit werden die Metadaten getrennt von der zugehörigen Ressource abgelegt, wie man es von Buchkatalogen in Bibliotheken her kennt. Die Speicherung kann z.B. als XML erfolgen. Werden Metadaten als eigenes Dokument zusammengefasst, so verfügen sie über einen Link mit einer bestimmten Identifikationsnummer (ID), der direkt auf das eigentliche Dokument verweist. Dabei werden die Metadaten in einer Datenbank gespeichert, was die beste Performance beim Zugriff gewährleistet. Dies hat aber auch zur Folge, dass die Metadaten von den eigentlichen Daten getrennt sind. Man muss sich also zusätzlich um Konsistenz und Synchronisation der Datenbestände kümmern. Die Metainformation und das Dokument sind gemeinsam in einem Container verpackt, welcher die Informationen bereitstellt, wenn sie benötigt werden. Die Metadaten sind direkt in der Ressource eingebettet. Dies ist allerdings nicht bei jeder Art von Ressource möglich, sehr wohl aber bei Textdokumenten wie z. B. Word- oder HTML-Dateien. Ein anderer Vorteil ist die Vereinfachung des Metadaten Managements und die zugehörige Suche und das Wiederauffinden. Die eingebettete Speicherung der Metadaten hat den Vorteil, dass die Metadaten nur mit den Daten selbst verloren gehen können. Weiters entstehen keine Probleme mit dem Verweis (Linkin) auf die Ressource und das Update der Metadaten passiert zusammen mit den Daten. Beispiele dafür sind HTML, SMIL, und ID3. [3][11]

### **1.6 Anforderung an Metadaten**

Die gegenwärtige Metadaten-Diskussion erhebt den Anspruch, dass bessere Erschließungs- und Retrievalmechanismen notwendig sind, als sie bisher im Internet vorhanden sind. Ein großer Bereich des Informationsangebots von Bibliotheken wird aus diesem Medium wie auch aus anderen elektronischen Quellen gespeist. Die Diskussion um Metadaten ist für die Bibliotheken sehr wichtig, da sich die Bibliotheken sich immer weniger allein auf ihre herkömmlichen Regelwerke beziehen können. [41] Das Erzeugen von Metadaten hat unter anderem den Zweck Informationsressourcen zu finden, elektronische Dokumente zu organisieren, Informationsaustausch und -transformation diverser Daten zu ermöglichen, gültige Ressourcen zu integrieren, digitale Daten zu identifizieren und Daten zu erhalten und zu archivieren. Wie wir bereits wissen, gibt es verschiedene Wege Metadaten nutzbar zu machen. Es müssen zuerst die Daten geschaffen werden, das können beschreibende Metadaten sein. Diese ermöglichen das Finden und Entdecken relevanter Informationen. Eine andere Möglichkeit kann sein, dass es strukturierende Metadaten sind, welche den Zugriff zu Informationen schaffen. Weiters gibt es administrative Metadaten, die Information verwalten [5][3]

Ein anderer Zugang zu Metadaten ist die automatisierte technische Bereitstellung von Metadaten über Objekte mit Funktionen. Die Funktionen werden über Metadaten beschrieben und dadurch für andere Anwendungen nutzbar. Grundsätzlich werden Input und Output Parameter in Form von Metadaten zur Verfügung gestellt. Die Firma

## **Einführung**

SAP stellt beispielsweise eine Metadatenklasse zur Verfügung, die genau diese Funktion bereitstellt. [44]

### **1.7 Informationsfindung**

Das Finden von Ressourcen passiert anhand zutreffender Suchkriterien. Die Ressourcen werden identifiziert und ähnliche Ressourcen werden zusammengeführt und unterteilt. Zum Wiederauffinden der Ressource müssen die Zugriffsinformationen entsprechend bereitgestellt und mit geeigneten Technologien zur Verfügung gestellt werden. [3]

#### **1.7.1 Das Ordnen elektronischer Ressourcen**

Metadateninformationen werden zur Organisation elektronischer Dokumente verwendet. Einerseits werden Informationen von den vorhandenen Dokumenten ausgewertet und werden zum Dokument hinzugefügt. Aus den vorhandenen Informationen werden zum Beispiel Metadaten, die im Dokument gespeichert oder aus den Daten abgeleitet werden können. "Hardcoded" Informationen können ausgewertet werden, genauso wie z.B. HTML Metadaten. Hinzugefügt werden Erläuterungen, wie Bemerkungen und relevante Informationen zu den verschiedenen Fachgebieten. [12]

#### **1.7.2 Austausch von Metadaten**

Jedes bibliothekarische Regelwerk setzt den effektiven Einsatz von Metadaten in einem gewissen Standardisierungsgrad voraus. Doch dieser Standardisierungsgrad hat bei den herkömmlichen Bibliotheksregelwerken inzwischen ein so hohes Komplexitätsniveau erreicht, dass es sich nicht auf die dramatisch steigende Fülle von Dokumenten in elektronischen Netzen wie dem derzeitigen Internet übertragen lässt. Ein Grund dafür sind u.a. die geringe finanzielle und personelle Kapazität der Bibliotheken.

Unter "Interoperability" versteht man den effizienten Austausch von Informationen über verschiedene Systeme hinweg, unter der Bedingung keine Informationsverluste bei der Übertragung zu haben. Genau dieser Austausch ist ein sehr wichtiger Aspekt bei der Verwendung von Metadaten mehrerer Systeme über das WWW. Metadaten können durch diesen Austausch von Mensch und Maschine verstanden werden. Um den richtigen Austausch von Metadaten zu garantieren muss bei der Generierung der Metadaten auf dieses austauschbare Format, das eben den reibungslosen Austausch bereitstellt, geachtet werden. [15] Der Dublin Core Standard bietet unter anderem standardisierte Transformationen dazu an. Der effiziente Austausch von Metadaten zwischen Business Warehouse Systemen ist auch im SAP System realisiert worden. Es wird der Common Warehouse Metamodel Standard benutzt um Daten auszutauschen. [14]

## **Einführung**

### **1.7.3 Transformation von Metadaten**

Wenn Metadaten in einem nicht austauschbaren Format vorliegen, muss eine Transformation der Daten durchgeführt werden. Im Allgemeinen wird so eine Transformation im Englischen als "Crosswalk" bezeichnet. Die Daten bzw. das Schema eines Systems werden im anderen System abgebildet. In diesem Zusammenhang könnte eine vollständige Transformationsmatrix natürlich auch eine Abbildung eines Metadatenstandards in einen anderen Standard erlauben. [16]

### **1.7.4 Metadaten Registrierung**

Eine Registrierung der Metadaten begünstigt das Finden und die Wiederverwendung existierender Metadaten Definitionen. Einerseits wird das Feststellen von Ähnlichkeiten bzw. Übereinstimmungen von Termen und Beziehungsdefinitionen vereinfacht und andererseits werden die Beziehungen zwischen den Termen aufgezeigt. Für Anwendungen gibt es in diesem Zusammenhang die Metadatenstandards (z.B. Dublin Core), Web Services, EDI, usw. Die Dublin Core Metadaten Initiative hat in diesem Zusammenhang eine Metadaten Registrierungsstelle ins Leben gerufen. [18][19] Zusätzlich wird ein ISO Standard diesbezüglich angeboten. ISO/IEC 11179 Metadaten Registrierung (MDR) Standard [20] Die Online Analyse, welche mittels eines Java Programms realisiert wird, verwendet zum Beispiel das OAI-PMH Repository. In diesem Repository können sich Benutzer von Metadatenstandards öffentlich registrieren. [26]

### **1.8 Identifikation von Metadaten**

Für die digitale Identifikation werden Elemente verwendet, die das Objekt eindeutig identifizieren und unter Umständen die Speicherstelle feststellen können. Allgemein gebräuchlich sind Dateinamen, die eben Objekte eindeutig identifizieren. Es werden URI (Uniform Resource Identifier) verwendet, die URL (Uniform Resource Locator), welche die Ressource identifiziert, und auch die so genannte URN (Uniform Resource Name) wird verwendet, welche den Namen erkennt. PURL (persistent Uniform Resource Locator) und DOI (Digital Object Identifier) ermöglichen ebenso die Bestimmung von Metadaten und sind unbedingt zu erwähnen. [3]

### **1.9 Archivierung und Erhaltung von Metadaten**

Digitale Information kann absichtlich oder unabsichtlich verändert oder fehlerhaft werden. Wenn zum Beispiel Hardware, Software oder Speichermedien getauscht oder kaputt werden, ist die darauf gespeicherte Information ebenso nicht in Ordnung und kann höchstwahrscheinlich nicht mehr verwendet werden. Eine aktuelle Vorgehensweise zur Archivierung und Erhaltung ist die Migration und Emulation bestehender Formate. Metadaten können in diesem Zusammenhang benutzt werden,

## **Einführung**

um einerseits die Abstammung, den Verlauf und auch die charakteristischen Details eines Dokumentes zu bestimmen. [3]

### **1.10 Metadaten Strukturierung**

Metadaten Schemata sind Sätze von Elementen, die für einen bestimmten Sinn und Zweck geschaffen wurden. Die semantischen Schemata sind bekannt dafür, dass jedes Element eine genaue Definition oder Bedeutung hat. Eine Option für die Strukturierung und Benennung der Felder besteht darin, Elemente hinzuzufügen, welche Inhalts- und Syntaxregeln spezifizieren. Metadaten Schemata können desweiteren in jeder definierten Syntax kodiert werden. Zurzeit werden unter anderem Arten von SGML bzw. XML verwendet, um relativ allgemeine Schemata zu erzeugen. [3] Welche Formate wie verwendet werden, wird in den nächsten Kapiteln auch mit Beispielen genauer beschrieben. Siehe Kapitel 2 und Appendix C bzw. D.

### **1.11 Erzeugung von Metadaten**

Metadaten beschreiben Ressourcen auf verschiedenen Ebenen der Zusammensetzung, z.B. ein Photo in einem Artikel, der in einer Zeitung steht, kann mit unterschiedlichem Detaillierungsgrad in Form von Metadaten dokumentiert sein. Die Entscheidung wie man die Metadaten katalogisiert, bleibt bei dem Erschaffer der Metadaten. Die IFLA (International Federation of Library Association and Institutions) hat ein Modell zur Methode zur Erschaffung von Metadaten entworfen. Es gibt unterschiedliche Methoden zur Erzeugung von Metadaten. Eine Methode besteht darin die Daten manuell einzugeben. Typische Felder wie Kommentare, Rechte und Author können auf diese Art eingegeben werden. Mit der automatischen Art Metadaten zu erzeugen werden Felder wie Erstellungsdatum, Größe oder andere extrahierte Merkmale befüllt. Eine Mischform der beiden vorangegangenen Methoden Metadaten zu erzeugen, ist die halb-automatische Vorgehensweise. Es ist die Kombination der manuellen und automatischen Technologie. [6][7] Auf welche Art und Weise Metadaten erzeugt werden hängt von verschiedenen Faktoren ab. Diese Faktoren werden unter anderem vom Anwendungsgebiet, der zu beschreibenden Ressource, der zur Verfügung stehenden Software Werkzeuge und dem gewünschten Output und Ergebnis bestimmt. [3]

#### **1.11.1 Beispiel an der Fa. SAP**

Es gibt ein Metadaten Repository, welches dem Benutzer ermöglicht auf Metadaten Objekte von einem zentralen Punkt im BW (Business Information Warehouse) zuzugreifen. Die Metadaten beinhalten wichtige Objekteigenschaften und Beziehungen zu anderen Objekten. Die Funktionen bestehen aus der Informationsanzeige, der Suche im Metadaten Repository, der graphischen Anzeige, dem HTML-Export, dem

## **Einführung**

Austausch der Metadaten im XMI Format, die Erzeugung eines Metadaten-Dokumentes und der Bereitstellung des Repository als HTTP Service. [22]

### **1.11.2 Beispiel an der Fa. Adobe**

Das Adobe XMP Entwicklungszentrum stellt eine Technologie bereit, welche Metadaten in ein existierendes File integrieren kann. Die unterstützten Betriebssysteme sind Macintosh, Windows, und UNIX. Das XMP Toolkit hat Beispiele, Dokumentation, Quell Code und Projektdateien. Das Toolkit ist unter der BSD Lizenz verfügbar. [52].

### **1.11.3 Erzeugungswerkzeuge**

Es gibt verschiedene Software Werkzeuge zur Erschaffung von Metadatenschemata oder Elementensätzen. Welche Software Werkzeuge verwendet werden hängt von den gewünschten spezifischen Metadaten ab. Vorlagen geben dem Benutzer die Möglichkeit Metadaten Werte in vorgegebene Felder, welche Elementensätze entsprechen, einzugeben. Es werden bei diesem Vorgang Metadaten Sätze generiert. Mark-up Tools strukturieren die Metadaten in spezifizierte Schema-Sprachen und einige Vorlagen unterstützen diese markierten Ergebnisse. Die erzeugten Formate sind zb. XML, SGML, DTD, etc. Extraktionswerkzeuge kreieren durch Analyse digitaler Ressourcen automatisch Metadaten. Die Qualität ändert sich offensichtlich dadurch, welche Algorithmen und welche Typen der digitalen Ressource verwendet werden. Transformations- bzw. Konvertierungswerkzeuge übersetzen Metadaten von einem Format in ein anderes. [3]

### **1.11.4 Qualitätskontrolle**

Gemäß dem Werk "Framework of Guidance for Building Good Digital Collections", das auf der NISO Webseite zur Verfügung steht, sollen Metadaten einige Anforderungen erfüllen: Die Anforderungen für gute Metadaten sind eine Art von Daten, die dem verwendeten Material, Benutzer, Absicht und der aktuellen und zukünftigen Verwendung möglichst entsprechen. Gute Metadaten sollen zwischen Systemen austauschbar (interoperable) sein, nur Standard Vokabular verwenden, klare Aussagen über die allgemeinen Benützungsbedingungen beinhalten, gute Qualitäten bei der Archivierbarkeit besitzen, Beständigkeit haben, und eindeutige Identifikation aufweisen, weil Metadatensätze selber digitale Objekte sind. Des Weiteren sollen sie allgemein maßgebend, nachvollziehbar und beweisbar sein und ein Management über einen langfristigen Zeitraum gewährleisten.

Laufende Anstrengungen, wie man mit Metadatenqualität umgeht und diese erhöhen kann, gehen in die Richtung, Erschaffungswerkzeuge mit Vorlagen, Auswahllisten und Überprüfungs-Regeln auszustatten. Außerdem versucht man die Qualität zu

## **Einführung**

verbessern, indem man Software-Schnittstellen bildet (interoperability), um ein möglichst einfaches Verständnis für Metadaten auch den Inhaltsmachern von Webseiten und Büchern zur Verfügung zu stellen. vordringen zu lassen. Des Weiteren sollen festgelegte Vokabeln einer breiteren Öffentlichkeit zugänglich gemacht werden, Communities entwickelt und Metadaten Schemata ständig verbessert werden. [3]

### **1.12 Rolle der Metadaten im Multimediabereich**

Metadaten für multimedialen Inhalt sind viel wichtiger als zum Beispiel für traditionell gut strukturierte Text Dokumente. Die Gründe dafür sind, dass einerseits unterschiedliche Abfragen verwendet werden und dadurch eine ganz genaue Übereinstimmung im Falle von multimedialem Inhalt nicht immer gegeben ist. Andererseits ist die Abarbeitungstechnik für inhaltsbasierte Daten für multimediale Daten unzulänglich und beschränkt. Dazu kommt, dass die inhaltsbasierte Suche kostenaufwändig ist und an Effizienz mangelt. Die Semantik von multimedialem Inhalt wird abgeleitet und interpretiert, d.h. der Kontext und die Semantik haben offensichtlich einen sehr hohen Wert. [3]

#### **1.12.1 Bedarf an Metadaten**

Ein Problem ist die Bearbeitung großer Mengen von digitalen Medieninhalten. Der Bedarf an Informationen von Inhalten und von anwendungsspezifischen Daten ist gegeben, d.h. die Inhalte und Informationen müssen vorher erfasst werden, um zu Metadaten verarbeitet werden zu können. Multimediale Metadaten können für verschiedene Multimedia Inhaltverzeichnisse, Beschreibungen und Anmerkungen verwendet werden. Auf die Metadaten soll bequem und einfach zugegriffen werden können. Zusätzlich sollten die Daten ausgewählt, selektiert und aufbereitet werden können. [3]

#### **1.12.2 Medien spezifische Daten**

Neben der normalen Art Metadaten zu generieren, die aus dem Inhalt des Mediums stammen, gibt es auch medien-spezifischen Metadaten. Meistens sind diese Metadaten auf das ausgewählte Format des Mediums bezogen und unabhängig vom Inhalt des Objektes. Beispiele sind je nach Medienart unterschiedlich. Bei Text wird nach Codierung, z.B. ASCII, UTF-8, Unicode, etc. und der verwendeten Hypertext Sprache, z.B. SGML, XML, HTML. geordnet. Bei Bildern kommen die Bits/Sample z.B. 8, 16, 24 ebenso die Farbpalette, welche die Farben für nicht-RGB Bilder beinhaltet. Auch werden die Größe, Druckgröße in technischen Bildmetadaten gespeichert. Bei Bild Dokumenten gibt es wie bei Ton und Bild die Bits/Sample, z.B. 8, 24. Wichtige Metadaten sind der verwendete Codec (DIVX, MPEG2, Windows Media, usw.), Datenrate (in kbit), Framerate (Bilder pro Sekunde), die Auflösung, die Dauer, usw. Bei Tondaten sind wieder die Bits/Sample zu erwähnen, auch der verwendete Codec

## **Metadatenstandards - ein Überblick**

(MP3), die Datenrate (64 128, 256kps) und Dauer sind wichtige Felder der Meta-Tondateien. [3]

## **2 Metadatenstandards - ein Überblick**

### **2.1 Verwendung in digitalen Bibliotheken**

Es treten in der Literatur immer wieder drei Begriffe in Verbindung mit Digitalen Bibliotheken auf: die digitale Bibliothek, die elektronische Bibliothek und die virtuelle Bibliothek. Alle diese Begriffe bezeichnen eine Einrichtung, welche schriftliche, visuelle, bzw. akustische Zeichen auf Basis neuer Kommunikationstechnologien sammelt, aufbewahrt und zur Verwendung bereitstellt. Die Attribute "digital", "elektronisch" und "virtuell" bezeichnen die drei verschiedenen Aspekte dieser Institution. Die Bezeichnung "Digitale Bibliothek" wird mit Dokumenten, wie Texte, Fotos, Filme und Animationen in Verbindung gebracht. Es werden jedoch keine Aussagen über das Speichermedium gemacht. [3][27]

### **2.2 Standardisierte Beschreibungsschemata**

Standards kommen zwar in unterschiedlichsten Bereichen zum Einsatz, weisen aber trotzdem einige gemeinsame Ausprägungen auf. Alle Standards haben gemeinsam, dass sie eine fixe auf den Anwendungsbereich zugeschnittene Menge von Beschreibungselementen aufweisen. Gewöhnlich ist die Grundlage dieser Standards eine sehr einfache. Die Basis sind Attribut-Wert Paare, eine einfache Typisierung und es werden keine Beziehungen und komplexe Datentypen verwendet. Zwei Metadatenstandards werden laut Literatur besonders stark eingesetzt: Der Dublin Core (DC) und der Machine Readable Cataloging Record (MARC). Im diesem Kapitel werden diese Standards noch genauer betrachtet. [3]

Ein Problem bei den standardisierten Beschreibungsschemata liegt in der schwierigen Anwendbarkeit auf verschiedene Fachgebiete. Meistens sind sie auch schwierig an spezielle Anforderungen anpassbar. Deswegen sind die verwendeten Datenmodelle sehr allgemein gehalten und wenig ausdrucksstark. Demzufolge können komplexe Themenstellungen nur mit Aufwand abgebildet werden. Zusätzlich fehlen gemeinsame Grundlagen für Datenmodelle und Austauschformate. Der Einsatz generischer Werkzeuge wird dadurch sehr erschwert. Austausch und Transformation erfolgt oft mittels einer Übersetzungstabelle der Attribute der jeweiligen Standards in den anderen Standard, somit kann eine Transformation erfolgen. [3]

## Metadatenstandards - ein Überblick

### 2.2.1 Dublin-Core

#### 2.2.1.1 Beschreibung

Im Rahmen der Metadatenentwicklung gibt es in den letzten Jahren eine Reihe von Initiativen, von denen das so genannte Dublin Core Metadata Element Set nicht nur am bekanntesten ist, sondern in der internationalen Diskussion auch eine immer größere Rolle spielt und inzwischen sicherlich zum am meisten diskutierten Ansatz geworden ist. Das DC Metadatenmodell wuchs seit 1995 in interdisziplinärem Konsens zwischen Informatikern, Wissenschaftlern und Bibliothekaren und besteht zurzeit aus einigen definierten Elementen zur Ressourcenbeschreibung. Es ist einfach gehalten, sodass die Produzenten von Dokumenten die entsprechenden Metadaten gegebenenfalls selber generieren können, ohne dabei auf aufwendige und damit teure Verfahren durch geschultes Personal zurückgreifen müssen. Die DC-Initiative stellt eine ambitionierte Reaktion auf die Herausforderung durch das World Wide Web dar und bietet einen zukunftsweisenden Ansatzpunkt für eine verbesserte Informationssuche im Internet. [29] Es gibt noch eine Vielzahl von weiteren Metadatenformaten neben dem Dublin-Core-Schema, z.B. die Text Encoding Initiative (TEI) oder Government Information Locator Service (GILS) und andere Formate, die entweder an fachspezifischen Erschließungsanforderungen (etwa die Content Standards for Digital Geospatial Metadata in den Geowissenschaften oder die Metadaten bei IEEE, dem Institute of Electrical and Electronic Engineers) oder an komplexeren bibliothekarischen Regelwerken, etwa PICA+ oder die existierenden MARC-Formate wie MARC, UNIMARC, USMARC, ... ausgerichtet sind. Gegenüber diesen Varianten steht hinter dem Dublin-Core-Ansatz der Versuch, ein kostengünstiges, leicht handhabbares und doch wirkungsvolles Verfahren zu entwickeln, das auch vor der großen Zahl an Dokumenten in elektronischen Netzen nicht kapitulieren muss. Weiterhin wird darauf geachtet, dass die Anbindung an andere, zumeist komplexere Formate möglich ist (Stichwort: interoperability). Dazu wurde nicht nur die so genannte Warwick-Framework-Architektur [73] entwickelt, die inzwischen dabei ist, vom noch innovativeren Resource Description Framework (RDF) abgelöst zu werden, sondern es wird auch verstärkt an Konvergenzen (Mappings/Crosswalks) zu anderen Formaten gearbeitet.

Die DCMI Metadaten bestehen aus allen Metadaten Bezeichnungen, die von der Dublin Core Metadaten Initiative (DCMI) gepflegt werden. Diese umfassen die Dublin Core Metadaten Elementsätze, die Felderweiterungs- und Feldkodierungs-Schemavermerke (Element refinements and encoding schemes-Qualifiers) und das DCMI Typ Verzeichnis (DCMI type vocabulary).[30] Die Dublin Core Metadaten Feld Sätze beschreiben den einfachen oder unqualifizierten Dublin Core Standard. Das Vokabular besteht zB. aus den Feldern (in Deutsch übersetzt, original nur in Englisch): Titel, Autor, Thema, Beschreibung, Veröffentlichter, Mitwirkende, Datum, Typ,

## Metadatenstandards - ein Überblick

Format, Identifier, Quelle, Sprache, Beziehung, Umfang, Rechte. Diese Element Sätze werden für die Ressourcen Beschreibung verwendet und sind ein Teil der technischen Spezifikation der Dublin Core Metadaten Initiative (DCMI). Die NISO (mit dem NISO Standard Z39.85.2001 im September 2001) und die ISO (mit dem ISO Standard 15836-2003 im Februar 2003) standardisierten den DCMI. Siehe dazu Appendix B für genauere Informationen. Die Element- Verbesserungs- (Erweiterungs-) und Kodierungsschema-Vermerke (Element refinements and encoding schemas - qualifiers) erweitern und spezialisieren den Dublin Core Standard. Zusätzliche Bezeichnungen werden verwendet um den ursprünglichen "einfachen" Satz an Elementen zu veredeln, z.B. wird das Datums Feld um die folgende Begriffe erweitert: Verfügbarkeit (available), erzeugt am (Created), allgemein akzeptiertes Datum (Date accepted), Urheberschutzdatum (Copyright Date) , Absendedatum (Date submitted), veröffentlicht (issued), geändert (modified) und gültig (valid). Die Element Erweiterungs- und Kodierungsschema-Vermerke empfehlen das Kodierungsschema für Element Werte, zb. das Datums Element soll mit dem W3C-DTF oder DCMI Punkt (period) Schema kodiert werden. Das DCMI Typ Vokabular gibt eine kontrollierte und eine branchenunabhängige Liste von anerkannten Bezeichnungen vor. Diese Vokabeln können als Werte für die Ressourcentypen eingesetzt werden. [74]

Der Dublin Core Standard ist sehr einfach gehalten. Elementareinträge lassen sich aus vorhandenen Metadatenbanken oder Medienbeschreibungen anderer standardisierter Beschreibungsschemata befüllen. Die Werte der Metadatenelemente und deren Attribute werden als Strings abgebildet. Somit kann Freitext eingegeben werden. Als negativen Punkt hinsichtlich des DC Standards kann man erwähnen, dass das zugrunde liegende Datenmodell sehr einfach und trivial gehalten ist, und somit die detaillierte Ausdrucksfähigkeit darunter leidet. [35][74]

### 2.2.1.2 Ziel

Das Entwicklungsziel von Dublin Core Metadaten für Ressourcen lassen sich in folgende Punkte einteilen: Einerseits ist der DC Standard einfach gehalten. Die meisten Elemente weisen eine allgemein verständliche und überschaubare Syntax auf, was DC sowohl für Spezialisten und auch für Anfänger geeignet macht. Hinzu kommt, dass die semantische Kompatibilität eine möglichst umfassende ist, d.h. die Metadaten können über verschiedene Disziplinen hinweg verwendet werden. Somit wird eine qualifizierte Suche im Internet möglich. Ein weiterer Punkt ist der internationale Konsens und die Anwendung des Standards in über 30 Ländern. Hervorzuheben ist weiters, dass die feinen semantischen Strukturen durch flexible Erweiterungen auszubilden sind und zusätzlich die Interoperabilität mit anderen Indiziersystemen im Internet unterstützt wird. Z.B. kann der Dublin Core Standard RDF verwenden. [3]

## Metadatenstandards - ein Überblick

### 2.2.2 MARC und MARCII - Machine Readable Cataloging Record

Das Kürzel MARC steht für Machine Readable Cataloging. Das MARC-Format ist eines der ältesten und umfangreichsten Standards im Bereich bibliographischer Metadaten. In seiner ursprünglichsten Form wurde das MARC-Format Mitte der Sechzigerjahre auf Initiative der amerikanischen Library Of Congress entwickelt. Die ersten Entwicklungsschritte von MARC gehen auf das Projekt MARC I zurück, welches die Erstellung und Nutzung von Katalogdaten in maschinenlesbarer Form versuchte. Parallel zu diesen Aktivitäten in USA versuchte das britische Council of the British National Bibliography ebenfalls technische Lösungen zur maschinenlesbaren Verwaltung bibliographischer Metadaten zu entwickeln. Die Zielsetzung der britischen Initiative war, die Produktion der British National Bibliography (BNB) mit Hilfe automatisierter Verfahren zu vereinfachen. Dieses Unterfangen firmierte unter dem Namen BNB MARC Project. Im Jahre 1968 fusionierten die amerikanischen und britischen Bestrebungen unter dem Projektnamen MARC II mit dem Ziel, ein standardisiertes Austauschformat für bibliographische Metadaten zu machen. Die Zielsetzung dieser Kooperation wurde wegen abweichender Anforderungen der nationalen Bibliotheken und abweichender Praktiken bei der Katalogisierung verfehlt. Die Folge war die Herausbildung eines jeweils eigenen Standards für Großbritannien und die Vereinigten Staaten. So gingen aus dem MARC II-Projekt der UKMARC für Großbritannien sowie der USMARC für den Einsatz in Nordamerika hervor. Das veröffentlichte Regelwerk zur Katalogisierung bibliographischer Daten, die Anglo-American Cataloguing Rules (AACR) erschienen einmal als britische und einmal als nordamerikanische Edition. Ableger des MARC Formats folgten am Ende der Sechzigerjahre. In Australien etablierte sich der Aus MARC, in Kanada bildete sich das Derivat CanMARC heraus. Trotz des von Beginn der MARC-Initiative an herrschenden Trends zur Diversifikation in heterogene nationale Formate wurde das Grundbestreben des MARC Formats als willkommener Beitrag in der internationalen Standardisierungsarbeit wahrgenommen. Die Anwendung der elektronischen Datenverarbeitung sowie einer eindeutigen Kennzeichnung des einzelnen Buches wird es wahrscheinlich ermöglichen, nationale Netze mit einem oder mehreren bibliographischen Zentren zu schaffen, die die bibliographischen Informationen über Neuerscheinungen des eigenen Landes sammeln, maschinell lesbar machen und auch in dieser Form verteilen könnten. Die nationalen Zentren werden untereinander die Daten ihrer eigenen, nationalen Bibliotheken verteilen und austauschen. [85]

Im Jahr 1969 startete das Library of Congress (LoC) mit der Verbreitung maschinenlesbarer Daten auf Magnetbändern. Der Name des Formats war MARC II, das der Nachfolger von MARC I war. Als Grundlage der Daten wurde das Regelwerk Anglo-American Cataloguing Rules von 1967 verwendet, welches das Format stark geprägt haben dürfte. Der Durchbruch war, dass Organisationen das Format aufgegriffen und aktiv propagierten. Es entstanden dann Varianten des Formats für

## Metadatenstandards - ein Überblick

Zeitschriften, Mikrofilme, Archivmaterial, audiovisuelle Medien, Karten, Noten und Tonträger. Alle diese Varianten wurden in mehrjähriger Arbeit unter dem Thema "Format Integration" 1989 zusammengefasst. An dieser Stelle möchte ich Stichworte wie Machine Readable (z.B. ein Rechner der Daten lesen und interpretieren kann) und Cataloging Record (Karteikarte) hervorheben. Wie jeder andere Standard ist auch der MARC Standard in verschiedene Untergliederung gefasst. Eine davon ist die "Bibliographic" Gliederung, die definierte Elemente für bibliographische Daten enthält. Eine andere Gliederung ist die Authority, die Angaben zum Namen und Organisation, etc. enthält. Holdings stellt eine weitere Einteilung dar und speichert strukturelle Eigenschaften in Datenform. Die Struktur des MARC hat einen zusätzlichen Gliederungspunkt, die Classification, diese macht, wie der Name schon sagt, eine Klassifikation von Inhalten. Ein zusätzliches Element zur Gliederung ist die Community, welches Angaben zu nicht-bibliographischen Informationsquellen macht. Der allgemeine Aufbau des MARC Standard ist so definiert, dass jeder MARC-Datensatz in logische Felder unterteilt ist. Jedes der Felder hat einen genau definierten Inhalt. Tags, welche aus dreistelligem Zahlen bestehen, kennzeichnen die Felder. Der Großteil der Felder hat einen Indikator und können auch in Unterfelder unterteilt werden. Der Großteil der Informationen ist wiederum in den Unterfeldern gespeichert. Im Feld 100 (Autor) wird beispielsweise der Name des Autors gespeichert. Grundsätzlich kann man sagen, dass die meisten Felder Unterfelder enthalten.

100 1# \$a Pirsig, Robert M.,

\$d 1966

Die Zahl 100 ist das so genannte *Tag*, welches das Feld identifiziert. Das Feld enthält den Namen des Autors. 1# enthält zwei Zeichen für Indikatoren, wobei # angibt, dass der zweite Indikator leer ist, also nicht verwendet wird. \$a und \$d sind Unterfelder. \$a stellt Hauptinformationen dar, und im Unterfeld (subfield) \$d wird eine zusätzliche Information gespeichert, nämlich in diesem Fall die weitere Zahl "1966".

## Metadatenstandards - ein Überblick

Tags bestehen immer aus einer dreistelligen Zahl, welche das Feld bezeichnet. Das Tag könnte als Ordnungszahl angesehen werden. Die 1000er Stelle gibt den Kontext des Feldes an.

0xx	Control Information
1xx	Main entry
2xx	Titles
3xx	Physics description, etc.
4xx	Series statements
5xx	Notes
6xx	Subject added entries
7xx	Added entries other than subject or series
8xx	Series added entries
9xx	Benutzerdefiniertes Feld, wird z.B. für Barcodes verwendet

**Tabelle 2-1 Ordnungszahlen MARC Standard**

## Metadatenstandards - ein Überblick

Die Einser und Hunderter Stelle des Tags beschreiben den Inhalt der Felder. Folgende Tabelle stellt die genauen Details dar:

X00	Personal Name
X10	Cooperate names
X11	Meeting names
X30	Uniform titles
X40	Bibliographic titles
X50	Topical terms
X51	Geographic names

**Tabelle 2-2 Ordnungszahlen MARC Standard**

Vgl. [37][38]

### 2.2.3 EXIF - austauschbares Bildformat für digitale Standbilder

Das EXIF Format ist ein Dateiformat, welches andere Formate wieder verwendet und deswegen zu existierenden Formaten "kompatibel" ist. Dieses Format ermöglicht spezifische Metadaten Tags hinzuzufügen und wird sehr häufig von Digitalkameras verwendet. Die EXIF Version 2.2 wurde von der Japan Electronics and Information-Technology Industries Association (JEITA) im April 2002 spezifiziert . Das EXIF Format beinhaltet Bild- und Tondatei- Spezifikationen und zusätzlich gibt es Auskunft über Beziehungs-Informationen, die zwischen Bild- und Ton-Dateien vorhanden sind. [40]

#### 2.2.3.1 EXIF Bilddatei Spezifikation

Das EXIF Bilddatei Format enthält die Struktur von Bilddateien, die verwendeten *Tags* von EXIF und die Definition, bzw. die Verwaltung der Format-Versionen. Komprimierte Dateien sind als JPEG gespeichert (ISO/IEC 10918-1) und werden mit kennzeichnenden Segmenten (APP1 und APP2) eingefügt. Unkomprimierte Dateien werden im TIFF Rev. 6.0 Format gespeichert. Die Tag Strukturen werden dem TIFF Format entnommen. Zu erwähnen ist, dass EXIF keine Unterstützung für JPEG2000, PNG oder GIF Formate hat. Die Metadaten Tags decken das breite Spektrum von

## **Metadatenstandards - ein Überblick**

Datum und Zeit ab, auch die Kamera Einstellungen, darunter fallen Informationen wie das Kameramodell und Hersteller, und genaue Bild Informationen: Ausrichtung, Blende, Verschlusszeit, Brennweite und ISO Geschwindigkeit sind vorhanden. Weiters sind Urheberschutzinformationen, Beschreibungen und Vorschaubilder in den Metadaten Tags enthalten. [40]

### **2.2.3.2 EXIF Audiodatei Spezifikation**

Die EXIF Audio Datei Spezifikation beschreibt die Struktur von Audiodateien, Einheiten (chunks) und auch Definition und Verwaltungsinformationen der Formatversionen. Das RIFF WAVE Tondateiformat nützt in unkomprimierter Form den PCM bzw. den  $\mu$ -Law PCM (ITU-T G.711) Codec und in komprimierter Form den IMA-ADPCM Codec. Allgemeine Eigenschaftsinformationen (Metadaten) sind in der Listeneinheit "INFO" gespeichert und auf die gleiche Weise werden EXIF-spezifische Einheiten definiert. [40]

### **2.2.4 LOM - Learning Object Metadata**

LOM ist der IEEE Standard 1484.12 für Learning Object Metadata. Die Definition eines lernenden Objektes ist, dass jede Entität erlaubt ist, ob sie nun digital oder nicht-digital ist. Das Objekt kann mithilfe von technikgestütztem Lernen benützt, wiederbenützt oder referenziert werden. Wenn man sich auf minimale Eigenschaften konzentriert, welche ein LOM benötigt, so sind das folgende: Ein LOM muss verwaltet, lokalisiert und bewertet werden können. Ein LOM muss die Wiederverwendbarkeit unterstützen und es muss auch die Auffindbarkeit der eigenen Ressource unterstützt werden. Allgemein werden LOMs in Learning Management Systems verwendet. Standards für LOMs gibt in Form des 1484.12.2. ISO/IEC Standards 11404, der das Modell für lernende Objekt Metadaten darstellt. 1484.12.3. ist der Standard für Learning Technology-Extensible Mark-up Language (XML), welcher die Schema Definition Binding for Learning Object Metadata definiert. Ein anderer wichtiger Standard ist der 1484.12.4: Das ist der Standard für Ressource Beschreibungssprache (Resource Description Framework-RDF). [41]

### **2.2.5 RAK und RAK-WB - Regeln für alphabetische Katalogisierung**

RAK bedeutet das "Regelwerk für die alphabetische Katalogisierung" und RAK-WB mit dem Zusatz in "wissenschaftlichen Bibliotheken". RAK ist wahrscheinlich der Schritt in die richtige Richtung, nämlich die schrittweise Ablösung der aus dem 19. Jahrhundert stammenden „Preußischen Instruktionen“ (PI) durch die in den 60er und 70er Jahren unseres Jahrhunderts entwickelten „Regeln für alphabetische Katalogisierung“ (RAK) und dem Mitte der 70er Jahre daraus hervorgegangenen „Regelwerk für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken (RAK-WB)“. Entscheidend ist für den durchaus großen Erfolg der RAK-WB die

## Metadatenstandards - ein Überblick

großflächige Einführung der EDV-Katalogisierung im deutschsprachigen Raum im Rahmen der Errichtung von Bibliotheksverbänden. Erforderlich war der EDV-Einsatz eines angepassten Regelwerkes, welches den Bedürfnissen und Anforderungen der Datentechnik entspricht. Eine Voraussetzung, die die RAK-WB im Gegensatz zu den PI erfüllten. Im neuen Regelwerk wurden international anerkannte Vorschriften berücksichtigt. So war die Einführung der RAK-WB erstmals eine solide, von vielen akzeptierte Basis für die weitere gemeinsame ‚Reiseplanung‘, denn die Schaffung und Anwendung eines bundesweit einheitlichen Katalogisierungsstandards gehörte zu den unabdingbaren Voraussetzungen, bevor überhaupt an die Realisierung weitergehender Zukunftsvisionen gedacht werden konnte.[76]

Zum Beispiel ist für das BSZ (Bibliotheksservice-Zentrum Baden-Württemberg) die Datenbasis der Verbundkatalog des Südwestdeutschen Bibliotheksverbands, der mit OPAC online verfügbar ist. Die Erschließung erfolgt nach speziellen bibliothekarischen Regeln, nach dem Regelwerk für die alphabetische Katalogisierung [78]. Der Internetbenutzer kann mit einem Web-Browser über ein Portal auf diese Datenbank und die enthaltenen Metadaten zugreifen. Die Daten und digitalisierten Objekte selbst, die durch die Metadaten beschrieben werden, verbleiben im Internetangebot und damit auf dem World Wide Web (WWW)-Server der jeweiligen Institution. Der Benutzer kann so gleichzeitig Informationen über Bücher zu dem Thema erhalten, das ihn interessiert. Verwendet für die Erschließung wird auch der Standard, der später in diesem Kapitel noch beschrieben wird, nämlich das standardisierte „Maschinenlesbare Datenaustauschformat für Bibliotheken“ (MAB). [76][77]

Im Gegensatz zu Archiven und Museen sind Bibliotheken schon lange im Internet präsent, z. B. über ihre OPAC Einrichtungen und virtuelle Bibliotheksverbände. Die Gründe dafür sind vielfältig. Bibliotheksdaten benötigen gewöhnlich keine provenienzorientierte Erschließung mit der Notwendigkeit einer tektonischen Einbindung. Für die Recherche eignen sich Datenbanklösungen, ohne dass eine kontextorientierte Navigation erforderlich ist.

Die Archiv- und Museumswelt, sind dagegen sehr heterogen. Es gibt wenige allgemein anerkannte oder genormte Erschließungsstandards und Datenaustauschformate. Zu verschieden sind die Vorstellungen und Traditionen, besonders die Verwaltungstraditionen der einzelnen Archive und Archivlandschaften. Im amerikanischen Raum gibt es als Standard die „Encoded Archival Description“ (EAD). EAD ist seit 1998 standardisiert und wird in den USA und teilweise darüber hinaus verwendet [77]. Des Weiteren ließ die PICA-Einführung die Erschließungspraxis der Formalkatalogisierung zumindest für die bereits vorher RAK-WB anwendenden Bibliotheken weitgehend unverändert. So hatte der Einsatz des Systems für die Sacherschließung aller Teilnehmer Konsequenzen: Mit Gründung des Verbundes wurde das so genannte Konzept der „kooperativen Sacherschließung“ von der

## Metadatenstandards - ein Überblick

„Projektgruppe Sacherschließung“ entwickelt und durch ministeriellen Erlass für alle niedersächsischen Teilnehmerbibliotheken angeordnet. Das Konzept ermöglicht so die Übernahme der Sacherschließungsleistungen der Deutschen Bibliothek. Daneben können interessierte Bibliotheken auch Sacherschließung nach eigenen Vorstellungen machen. Im Rahmen der kooperativen Sacherschließung im Verbund kommt der SWD als ‚Quelle‘ RSWK-gerechter Schlagwörter zentrale Bedeutung für die verbale Erschließung zu. Schon kurz vor der verbundweiten PICA-Einführung wurde die SWD als laufend zu pflegende Normdatei in der Verbunddatenbank vorgehalten. [77]

### 2.2.6 MAB und MAB2 - Maschinelles Austauschformat für Bibliotheken

MAB ist die Abkürzung für Maschinelles Austauschformat für Bibliotheken. Mit MAB können alle im Bibliotheksbereich erzeugten Daten ausgetauscht werden, nämlich bibliografische, Norm- und Lokaldaten. Mit dem Ausbau und der Weiterentwicklung von MAB zu einem allgemeinen Kommunikationsstandard ergaben sich neue zusätzliche Anforderungen, für die MAB nicht eingerichtet war. Daher entwickelte der MAB-Ausschuss nach zweijähriger Entwicklungsarbeit eine neue Formatversion: MAB2.

Der MAB2 enthält alle erforderlichen inhaltlichen, strukturellen und technischen Erweiterungen, um MAB auch als Austauschformat in Online-Umgebungen einsetzen zu können. MAB2 besteht heute aus fünf einzelnen Datenformaten, die auf einer einheitlichen, integrierten und für alle Formate gültigen Feldstruktur aufsetzen: das MAB-Format für bibliografische Daten (MAB-TITEL), das MAB-Format für Personennamen (MAB-PND), das MAB-Format für Körperschaftsnamen (MAB-GKD), das MAB-Format für Schlagwörter (MAB-SWD) und das MAB-Format für Lokaldaten (MAB-LOKAL). Darüber hinaus gibt es zwei MAB-Formate, nämlich das MAB-Format für Adress- und Bibliotheksdaten (MAB-ADRESS), MAB-Format für Klassifikations- und Notationsdaten (MAB-NOTAT).

Seit dem Erscheinen der 4. Ergänzungslieferung sind die beschlossenen Formaterweiterungen und Formatänderungen bereits Bestandteil des MAB-Formats. Die Anwendung des MAB-Formats wird in der MAB-Dokumentation verbindlich geregelt und auch durch zahlreiche Beispiele verdeutlicht. Bis dato sind vier Ergänzungslieferungen zum Grundwerk erschienen.

Zuständig für die Pflege von MAB2 ist eine Expertengruppe für Datenformate. Die Federführung ist bei der Deutschen Nationalbibliothek. Vertreten sind beinahe alle wichtigen Einrichtungen des deutschen Bibliothekswesens. Anträge auf Änderung oder Erweiterung kann jeder Anwender des MAB-Formats stellen. Protokolle der Sitzungen werden online zur Verfügung gestellt. [79]

## Metadatenstandards - ein Überblick

### 2.2.7 METS - Metadata Encoding and Transmission Standard

METS ist eine Initiative der Digital Library Federation und versucht aufbauend auf der Arbeit von MOA ein XML-Dokumentformat zu entwickeln. Die kodierten Metadaten stehen sowohl für die Verwaltung von Objekten einer digitalen Bibliothek zur Verfügung als auch können die Metadaten für den Austausch verwendet werden. Ein METS-Dokument kann im Kontext des Open Archival Information System (OAIS)-Referenzmodell für die langfristige Erhaltung digitaler Aufzeichnungen entweder als Liefereinheit (Submission Information Package - SIP), Archivierungseinheit (Archival Information Package - AIP), oder als Bereitstellungseinheit (Dissemination Information Package - DIP) verwendet werden.

Ein METS-Dokument besteht, wie in Abbildung 2-1 zu sehen ist, aus sieben Hauptabschnitten: Der Kopfteil (METS Header). Der Kopfteil enthält Metadaten, die das jeweilige METS-Dokument selbst beschreiben, einschließlich der Angaben zum Bearbeiter oder Herausgeber des METS-Dokuments. Weiters gibt es Erschließungsangaben (Descriptive Metadata). Der Abschnitt für die Erschließungsangaben kann sowohl Verweise auf ein externes Dokument, (etwa einen MARC Datensatz in einem OPAC oder ein EAD-Findbuch auf einem WWW-Server), wie auch in das METS-Dokument eingebettete Angaben oder beides enthalten. Es können mehrere externe und interne Erschließungspakete in dem Erschließungsabschnitt benutzt werden. Zusätzlich existieren Verwaltungsangaben (Administrative Metadata).

## Metadatenstandards - ein Überblick

Der Abschnitt für die Verwaltungsangaben liefert Informationen über die Herstellung und Speicherung von Dateien, über Urheberrechte und über die digitalisierte Vorlage. Zusätzlich werden Angaben zur Herkunft der Digitalinformationen erfasst, zum Beispiel über das Verhältnis von Master und Derivaten sowie über Migrationen.

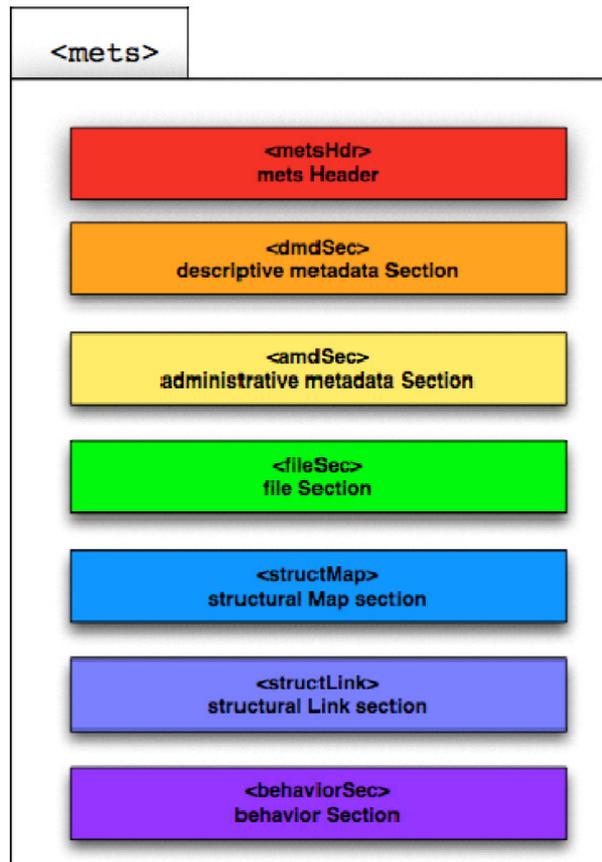


Abbildung 2-1 METS Metadatenstandard

Ähnlich wie bei den Erschließungsangaben können die Metadaten extern oder in das METS-Dokument integriert werden. Des Weiteren gibt es den sogenannten Dateie Abschnitt (File Section). Im Dateienabschnitt werden jene Dateien mit Inhalten, aus denen das digitale Objekt besteht, aufgelistet. Einzelne zusammengehörige Dateien können dabei mit dem Element zusammengefasst werden, um etwa verschiedene Versionen auseinander halten zu können. Die Strukturbeschreibung (Structural Map) ist zentraler Bestandteil eines jeden METS-Dokuments. Sie bildet den inneren Aufbau des digitalen Objektes ab und verknüpft die Elemente der Struktur mit den Dateien, aus denen der Inhalt des digitalen Objektes besteht, sowie auch mit deren Metadaten. Die

## Metadatenstandards - ein Überblick

Strukturverknüpfungen (Structural Links) sind der Abschnitt, der es ermöglicht den Erstellern von METS-Dokumenten das Vorhandensein von Hyperlinks zwischen einzelnen Knoten des im Strukturabschnitt dargestellten hierarchischen Aufbaus des digitalen Objekts zu beschreiben. Diese Funktion ist speziell für die Archivierung von Webseiten gedacht. Das Verhalten (Behavior) des digitalen Objekts kann verwendet werden, um ausführbare Anweisungen für das Verhalten mit den Inhalten in METS-Objekten zu verknüpfen. Jede Verhaltensform hat ein Schnittstellendefinitionselement, welches eine abstrakte Definition eines Satzes von Verhaltensformen in dem jeweiligen Abschnitt enthält. Zusätzlich besitzt jede Verhaltensform ein Mechanismuselement, das ein Modul ausführbaren Codes enthält, mit dem die in der Schnittstellendefinition abstrakt formulierten Verhaltensformen ausgeführt werden können. [80][103]

### 2.2.8 RSWK - Regelwerk für Schlagwortkatalogisierung

Die rechnergestützten Schlagwortkatalogisierungsregeln sind im Bereich von Personennamen, Körperschaftsnamen und Titeln sehr wünschenswert. Eine Arbeit zur alphabetischen Katalogisierung der Deutschen Bibliothek, stammt aus der Zeit vor dem Erscheinen der ersten Auflage der RSWK. Es wurde gewarnt vor einer „Zersplitterung“ der jeweiligen Regelungen, es konnte dies jedoch nicht verhindert werden, da die RSWK sich letztlich auf etablierte deutsche Schlagwortregelwerke und die Ergebnisse von Abstimmungen durch deutsche Bibliotheksverbände und Einzelbibliotheken berufen hat. Dies geschah, um die Akzeptanz der RSWK als neues Einheitsregelwerk zu erhöhen. 20 Jahre später diskutiert die deutschsprachige bibliothekarische Fachwelt endlich ergebnisorientiert über die Angleichung von RAK und RSWK, jetzt allerdings unter Anpassung der RAK an Struktur und Inhalte der Anglo-American Cataloguing Rules (AACR) und dessen Nachfolger. Die im Literaturverweis abgebildete Darstellung ist der Versuch, das Modell eines für die Formal- und Inhaltserschließung gemeinsamen deutschen Regelwerks darzustellen und die noch bestehenden Unterschiede zu erfassen. Siehe dazu [83]

Mit der dritten Auflage der RSWK ist eine Basis geschaffen, auf der das Regelwerk und die Erschließungspraxis der DB (Deutsche Nationalbibliothek) und der Verbände klarer und einheitlicher weitergeführt werden können. In manchen Verbänden wurden Teile auch schon früher angewandt. Sicherlich werden die erforderlichen Umarbeitungen der SWD auch noch eine gewisse Zeit erfordern. Die Anpassungen der RSWK an den Online-Katalog haben aufgezeigt, dass die verbale Erschließung durch die leicht mögliche Anpassung an unterschiedliche Bedürfnisse auch für Wandlungen im Umfeld offen ist. Im Rahmen dieses Berichts können nicht alle Aspekte der Änderungen dargestellt werden. Wichtig erscheint mir jedoch einerseits, dass sich das bibliothekarische Umfeld klar wird, wie die verbale Erschließung in neue Informationsstrukturen integriert werden kann. Hier liegen andererseits noch

## **Metadatenstandards - ein Überblick**

wesentliche Aufgaben der Zukunft. Ein zweites ebenso wichtiges Feld ist, die Änderungen in den RSWK in den OPACs umzusetzen. [81]

Zu der Frage nach einer Bestandserschließung kann zum Kriterium des Umfanges das Kriterium der Qualität hinzukommen. Konkret steht hierbei die Erschließungstiefe und damit die Bestandsdurchdringung der Sacherschließung zur Diskussion. So ist es bei der verbalen Sacherschließung nach RSWK möglich, gewisse Dokumentengruppen, die sich einer engeren, themengenaue Indexierung entziehen, auf lokaler, nicht aber auf der Ebene der Verbundkatalogisierung, nach weiten Aspekten zu beschreiben. Im Einzelnen haben Bibliotheken die Möglichkeit, Literatur (Kinder- und Jugendliteratur, etc.), Musikalien, audiovisuelle Medien oder Altbestände nach Gattung, Strömung, Entstehungsgebiet, Sprache, Zeit und Form durch Schlagwörter zu erschließen. Einen quantitativ bedeutsamen Sonderfall der Bestandsdurchdringung stellt die Erschließung unselbständiger Werke dar, die in der Regel von Bibliotheken aufgrund ihrer Zahl nicht einzeln erschlossen werden können. Ein weiterer sachlicher Zugang erfolgt lediglich über die inhaltliche Beschreibung der in die Kataloge aufgenommenen übergeordneten Sammelwerke. Allerdings lässt es ein modernes Regelwerk wie die RSWK durchaus zu, auch Zeitschriftenbeiträge oder Artikel aus Aufsatzsammlungen mit Schlagworten zu versehen. Den Praxisbeweis einer erfolgreichen verbalen Indexierung unselbständiger Literatur im Rahmen der RSWK-Erschließung erbringen diejenigen Bibliographien, die Bibliotheken auf der Grundlage spezieller Aufträge zur Literatursammlung erstellen. Zu nennen sind hier Landesbibliographien wie etwa die Bayerische oder die Nordrhein-Westfälische Bibliographie (NWBib) oder der Index Theologicus, der von der Tübinger Universitätsbibliothek als Sondersammelgebietsbibliothek der Deutschen Forschungsgemeinschaft (DFG) für Theologie und Religionswissenschaft erstellt wird. [82]

### **2.2.9 PICA - Project of Integrated Catalogue Automation**

OCLC PICA ist ein Anbieter von Bibliothekssoftware in Europa, der aus einer Kooperation der holländischen Pica-Stiftung und dem US-amerikanischen Online Computer Library Center (OCLC) hervorgegangen ist. Die Pica Foundation wurde als Initiative der Königlichen Bibliothek der Niederlande, und anderer Universitätsbibliotheken ins Leben gerufen, um die gemeinsame Katalogisierung und Automatisierung voranzutreiben. Das Kürzel PICA steht dabei für "Project of Integrated Catalogue Automation". 1986 bekam Pica als Stiftungszentrum für Bibliotheksautomatisierung einen eigenen rechtlichen Status, welcher 1997 in eine Stiftung umgewandelt wurde. Im Dezember wurde zusammen mit OCLC die Pica B.V, vergleichbar mit einer deutschen GmbH, gegründet. Durch einen Zusammenschluss mit der für Europa, den Mittleren Osten und Afrika zuständigen Abteilung von OCLC entstand die Organisation OCLC PICA. 2005 besaß OCLC mit 60% die Mehrheit der Gesellschaftsanteile an OCLC PICA, während die Pica-Stiftung 40% hatte. Die Pica-Stiftung wurde von der NBLC-Vereinigung und der SURF-Stiftung getragen. 2005

## **Metadatenstandards - ein Überblick**

übernahm OCLC PICA mit der Sisis Informationssysteme GmbH, welche in Deutschland ist, und der Fretwell-Downing Informatics Group, die in Großbritannien ist, weitere Dienstleister und Softwarehersteller aus dem Bibliothekswesen. 2007 erwarb OCLC die restlichen Geschäftsanteile an OCLC PICA und ist somit nun alleiniger Anteilseigner der Organisation. Ende 2007 werden alle OCLC PICA Niederlassungen in OCLC umbenannt. Durch die Übernahmen in 2005 hat sich das Produktspektrum von OCLC PICA erheblich erweitert. Es umfasst klassische Lokal- und Verbundsysteme und auch Software und Lösungen in den Bereichen Fernleihe, Portale und Linkresolver. OCLC PICA ist zudem für den Vertrieb der OCLC Services und Produkte (WorldCat, CONTENdm, Netlibrary) unter anderem in Europa zuständig.

Im deutschsprachigen Raum bekannt ist OCLC PICA großteils durch die Pica-Software für zentrale- und lokale Bibliotheksdatenbanken unter anderem als OPAC. Die Software wird in Deutschland von der Deutschen Nationalbibliothek und den Bibliotheksverbänden GBV, HeBIS und SWB eingesetzt. Weitere Partner sind u.a. die National Library of Australia, das Fernleihsystem V3 in Großbritannien, der französische Gesamtkatalog SUDOC und der Niederländische Verbundkatalog, der gleichzeitig bei OCLC PICA betrieben wird. OCLC PICA beschäftigt mehr als 250 Personen, von denen ca. 170 in der Entwicklung und dem Support arbeiten. [87]

### **2.2.10 International Standard Organisation (ISO) - Standard 19115**

Der internationale Standard, ISO Standard 19115 der ein geographisches Informationsmetadatensystem ist, wurde 2003 geschaffen. Technische Herausforderungen die Datensätze dem "Federal Geographic Data Committee" (FGDC) kompatibel zu machen sind im Gange. Ein weiteres Ziel ist die Implementierung eines Modells, welches mit XML verwendet werden kann. [3]

Im Mai 2007 war im Newsletter der "National Information Standards Organization" (NISO) zu lesen, dass die ISO/TS 19139:2007 Implementierung ein geographisches Metadaten XML Coding definiert, welches eine Ableitung des XML ISO 19115 Standards ist. Prinzipiell kann der Standard um viele geographische Daten erweitert werden, wie zum Beispiel Karten, Charts und Textdokumente. [25]

### **2.2.11 ISAD(G)- (General) International Standard Archival Description, Second edition**

Auf deutsch übersetzt heißt ISAD Internationale Grundsätze für die archivische Verzeichnung. [88]

Die ISAD(G) ersetzen nationale, regionale oder von einzelnen Archivverwaltungen erarbeitete Regelungen nicht, sondern ergänzen vielmehr oder sollen eine Leitlinie für

## Metadatenstandards - ein Überblick

ihre Formulierung sein. Die Verzeichnungsrichtlinie für die staatlichen Archive in Baden-Württemberg ist ein Beispiel für diese Interpretation. Die ISAD(G), wie sie in der vorliegenden Form ist, ist die Basis für „Encoded Archival Description“ (EAD), die als technisches Werkzeug zur Onlinerepräsentation von Verzeichnungsergebnissen dienen kann. Die Nutzung von EAD als internationales Austauschformat für Online-Suche wird dadurch unterstützt. Desweiteren enthält die vorliegende Ausgabe erstmals auch deutsche Beispiele, um die Anwendung der ISAD(G) in deutschen Archiven zu erleichtern. Sie folgt sozusagen dem Beispiel der englischen, französischen und spanischen Ausgabe. Die Beispiele wurden fast ausschließlich aus online erhältlichen Büchern entnommen. Es sind also keine „konstruierten“ oder künstlich gebildeten Erläuterungen der Regeln der ISAD(G), sondern sie stammen tatsächlich aus der archivischen Praxis. Sie zeigen auf sehr beeindruckende Weise, dass zwischen den Regeln der ISAD(G) und deutscher Verzeichnungspraxis große Ähnlichkeiten bestehen. Die Verzeichnungsinformationen werden in sieben Informationsbereichen zusammengefasst: **Identifikation** beinhaltet unerlässliche Angaben zur Identifikation der Verzeichnungseinheit aufzeigt. **Kontext**, der Angaben über Ursprung und Aufbewahrung der Verzeichnungseinheit beinhaltet. **Inhalt und innere Ordnung**; hier sind Angaben über Betreff und Ordnung der Verzeichnungseinheit zu finden. **Zugangs- und Benutzungsbedingungen** enthalten Angaben über die Verfügbarkeit der Verzeichnungseinheit und **Sachverwandte Unterlagen**, die Angaben über Unterlagen, die eine wichtige Beziehung zur Verzeichnungseinheit aufweisen. **Anmerkungen** dienen für besondere Angaben, die in anderen Bereichen nicht gemacht werden können. Zusätzlich gibt es die **Kontrolle**, die Angaben über den Zeitpunkt, das Verfahren und den Autor der Verzeichnung enthält. Alle 26 von den allgemeinen Regeln abgedeckten Elemente können benutzt werden, aber nur ein Teil von ihnen ist für jede Verzeichnung unbedingt notwendig. Um den internationalen Austausch zu gewährleisten, sind folgende Elemente zwingend zu verwenden, nämlich die Signatur, Titel, Provenienzstelle, Laufzeit, Umfang und Verzeichnungsstufe. Die Beispiele im Text der ISAD(G) illustrieren, wie diese Regeln angewandt werden. Sie geben keine verbindliche Vorschrift, aber die Beispiele sollen die Anwendung der Regeln deutlich werden lassen. Sie geben aber auch keine bestimmte Anwendung vor. Die Beispiele dürfen weder in im Aufbau noch in ihrer Form mit den Regeln verwechselt werden. Um den Kontext deutlicher zu machen, sind in den Beispielen die jeweilige Verzeichnungsstufe bzw. die Provenienzstelle immer mit dazu angegeben. Kommentare sind durch den Begriff „Anmerkung“ gekennzeichnet. Weiterführende Literatur und Beispiele sind im Literaturverweis [89] zu finden.

### 2.2.12 RFC1807- Request for Comments

Dieser RFC definiert das Format für bibliographische Daten, die technische Auswertungen beschreiben. Das Format wird vom der Cornell University Dienst protocol and the Stanford University SIFT System benutzt. Das Original RFC (RFC

## Metadatenstandards - ein Überblick

1357) wurde von D. Cohen, ISI, July 1992 entwickelt. Dieser Standard ist eine Revision von RFC 1357. Neue Felder sind handle, other\_access, keyword, und withdraw. [91]

### Beispiel

```
-----
BIB-VERSION:: CS-TR-v2.1
            ID:: OUKS//CS-TR-91-123
            ENTRY:: January 15, 1992
ORGANIZATION:: Oceanview University, Kansas, Computer Science
            TYPE:: Technical Report
REVISION:: January 5, 1995; FTP access information added
TITLE:: Scientific Communication must be timely
AUTHOR:: Finnegan, James A.
CONTACT:: Prof. J. A. Finnegan, CS Dept, Oceanview Univ,
            Oceanview, KS 54321 Tel: 913-456-7890
            <Finnegan@cs.ouks.edu>
AUTHOR:: Pooh, Winnie The
CONTACT:: 100 Aker Wood
            DATE:: December 1991
            PAGES:: 48
COPYRIGHT:: Copyright for the report (c) 1991, by J. A.
            Finnegan. All rights reserved. Permission is granted
            for any academic use of the report.
HANDLE:: hdl:oceanview.electr/CS-TR-91-123
OTHER_ACCESS:: url:http://electr.oceanview.edu/CS-TR-91-123
OTHER_ACCESS:: url:ftp://electr.oceanview.edu/CS-TR-91-123
RETRIEVAL:: send email to Finnegan@cs.ouks.edu with fax number
KEYWORD:: Scientific Communication
CR-CATEGORY:: D.0
CR-CATEGORY:: C.2.2 Computer Sys Org, Communication nets, Net
            Protocols
SERIES:: Communication
FUNDING:: FAS
CONTRACT:: FAS-91-C-1234
MONITORING:: FNBO
LANGUAGE:: English
NOTES:: This report is the full version of the paper with
            the same title in IEEE Trans ASSP Dec 1976
```

vgl. [90]

### 2.2.13 JUNII und JUNI2

Der junii bzw. junii2 Metadatenstandard ist japanischen Ursprungs. Das Schema ist unter <http://www.nii.ac.jp/irp/info/junii2.xsd> zu finden. Verwendbare gefundene Ressourcen zu diesem Standard sind alle in japanischer Sprache verfasst. Ein Auszug des Codings ist hier zu finden:

```
<xs:complexType name="metaContent">
  <xs:sequence>
    <xs:element ref="title" minOccurs="1" maxOccurs="1"/>
    <xs:element ref="alternative" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="creator" minOccurs="0" maxOccurs="unbounded"/>
```

## Metadatenstandards - ein Überblick

```
<xs:element ref="subject" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="NIISubject" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="NDC" minOccurs="0" maxOccurs="unbounded"/>
...
<xs:element ref="source" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="language" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="relation" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="pmid" minOccurs="0" maxOccurs="1"/>
<xs:element ref="doi" minOccurs="0" maxOccurs="1"/>
<xs:element ref="isVersionOf" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="hasVersion" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="isReplacedBy" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="replaces" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="isRequiredBy" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="requires" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="isPartOf" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="hasPart" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="isReferencedBy" minOccurs="0" maxOccurs="unbounded"/>
<xs:element ref="references" minOccurs="0" maxOccurs="unbounded"/>
...
</xs:sequence>
</xs:complexType>
```

### 2.2.14 ETD-MS -Electronic Theses and Dissertations Metadata Set

Das XML Format für "the Electronic Theses and Dissertations Metadata Set" (ETD-MS) ist unter <http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd> zu finden. [92][93]

Die Qualität einer Universität wird von der Qualität der Studenten und deren intellektuellen Errungenschaften repräsentiert. Diplomarbeiten und Dissertationen spiegeln die Fähigkeit von Institutionen, die die Studenten leiten und bei origineller Arbeit unterstützen. Inzwischen, als digitale Bibliotheken von ETD häufiger geworden sind, beurteilen die Studenten Universitäten nach deren digitalen Bibliothek. Universitäten, welche neue moderne literarische Werkzeuge zur Verfügung stellen wie zB. Multimedia streaming, sind für innovative Studenten attraktiv. Ein ETD Programm ist wie jedes andere Projekt, nämlich dass alle Beteiligten motiviert bis zum Abschluss sind. Bis zu dem Zeitpunkt, wo ETD eine reguläre Tätigkeit im Programm einer Universität ist, stützt sich ETD auf die gemeinsame Arbeit und Aktivitäten von graduierten Studenten, Mentoren, Administratoren, dem Bibliotheksteam, graduierten Leuten und dem IT Team. Der Erfolg eines ETD Programms benötigt alle diese Beteiligten und zusätzlich die höheren administrativen Bürofachkräfte. [94]

Es gibt zwei Haupttypen von ETDs. Die stark bevorzugte Art, welche von Studenten gemacht wird, ist die, dass die Arbeit in elektronischer Form aufbereitet wird (word, LaTeX) und danach übermittelt wird. Typischerweise wird das rohe Format in eine einfach zu erhaltende Form konvertiert, d.h. vom .doc ins .xml Format gewandelt.

## Metadatenstandards - ein Überblick

Diese Form wird normalerweise über eine Netzwerkverbindung mit dazugehörigen Metadaten übermittelt. Einmal übertragen können solche ETD's von interessierten Usern gefunden werden. Die zweite Art der ETD's ist die typische Form des File, das zuvor mit einem Scanner vom Universitätspersonal elektronisch erzeugt wurde. Diese Art der ETD's sind weit nicht so erwünscht als die erste Art, weil der Speicherbedarf sehr hoch ist und keine Voll Text Suche unterstützt wird. Trotzdem kann auf diese Art und Weise Informationsmaterial zugänglich gemacht werden. Der Fokus sollte jedoch auf die erste Möglichkeit (Arbeit in elektronischer Form aufbereitet) gelegt werden. [95]

### 2.2.15 QDC - Qualified Dublin Core

Es gibt mehrere QDC Schemata, zb.

[http://kramerius.nkp.cz/kramerius/schema/oai\\_qdc.xsd](http://kramerius.nkp.cz/kramerius/schema/oai_qdc.xsd) bzw.

<http://epubs.cclrc.ac.uk/xsd/qdc.xsd> oder auch <http://edoc.mpg.de/doc/schema/qdc.xsd>.

Mit unqualified DC ist ein kleines und flexibles Schema gegeben, das als kleinster gemeinsamer Nenner komplexerer Systeme aufgefasst werden kann. In Webseiten wird dieses Schema bei den Metatags angewendet, die von robots ausgelesen werden um Suchmaschinen oder andere Verzeichnisse mit Dokumentbeschreibungen zu versorgen. Qualified DC hingegen geht stärker auf die Feinkörnigkeit bestehender Katalog- und Erfassungssysteme ein und versucht durch Erweiterungen des Dublin Core Metadaten Element Sets soviel Information wie möglich zu erhalten. [96]

#### Beispiel:

```
<oai:metadata xmlns:ns=" http://purl.org/dc/elements/1.1/" xmlns:oai=" http://www. -
openarchives.org/OAI/2.0/oai_dc/"
xmlns:ter=" http://purl.org/dc/terms/" xmlns:xsi=" http://www.w3.org/2001/XMLSchema - -
instance">
<ns:title lang="eng">ganz schnell </ns:title >
<ter:alternative lang="" / >
<ter:abstract lang="ger">zusammenfassung </ter:abstract >
<ns:subject >
<ns:ddc >050 </ns:ddc >
</ns:subject >
<ns:subject >
<ns:unclassified >rapid </ns:unclassified >
</ns:subject >
<ns:subject >
<ns:classified >rapid </ns:classified >
<ns:classification xsi:type="oai:MESH" >1256 </ns:classification >
</ns:subject >
<ns:type xsi:type="oai:doc -type">text </ns:type >
<ns:type xsi:type="oai:pub -type">article </ns:type >
<ns:creator type="person">
```

## Metadatenstandards - ein Überblick

```
<ns:firstName >Peter </ns:firstName >
<ns:lastName >Reimer </ns:lastName >
<ns: academicTitle />
<ns: organization />
<ns:email/>
<ns:postal/>
<ns: pndIdentifier />
</ns:creator >
<ns:language >ger </ns:language >
<ns:rights >DPPL </ns:rights >
<ter:created >2004 -12 -13 </ ter:created >
<ter:modified >2004 -12 -13 </ ter:modified >
<ter:valid >2004 -12 -13 </ ter:valid >
<ns:identifier xsi:type = " oai:URN">urn:nbn:de :0009 -1 -1043 </ ns:identifier >
<ns:identifier xsi:type = " oai:URL">http :// www.dipp.nrw.de/ testjournal /archive /104 </ ns: -
identifier >
</oai:metadata >[99]
```

### 2.2.16 EPICUR - Enhancement of Persistent Identifier Services- Comprehensive Method for Unequivocal Resource Identification

URN-Records umfassen folgende Informationen: Einen neu erstellten Metadatensatz mit Informationen über die URN und die zugehörige URL. Es wird zusätzlich eine Metainformation über einen bei dem DataProvider geänderten Record angegeben. Dieser Fall tritt dann ein, wenn die URL aktualisiert wurde. Es können optional hierarchische Zuordnungen (Gesamt- und Teildokument) angegeben werden und die für Teildokumente vergebenen URNs. Optional ist die Abbildung von semantischen Informationen vorgesehen wie zum Beispiel die Angabe einer URN, die für eine neue Dokumentversion vergeben wurde. Das bietet eine separate Schnittstelle zur Implementierung eines aktiven URN-Melde- und Updateprozesses an. [97][102]

Stabile Referenzen können auf digitale Objekte zum Beispiel in Informationssystemen wie Online-Katalogen oder Bibliografien erzeugt werden, wenn ein persistentes Adressierungsschema wie es zum Beispiel mit Uniform Resource Names (URNs) zur Verfügung steht, verwendet wird. Die Dauerhaftigkeit eines solchen Adressierungsschemata kann jedoch nicht allein nur durch deren Anwendung im lokalen Kontext sichergestellt werden. Damit die Anwendung eines persistenten Adressierungsschemata wie den URNs langfristig gewährleistet werden kann, muss eine Infrastruktur mit einer institutionellen Unterstützung geschaffen werden. Nur in diesem Kontext kann sichergestellt werden, dass die Vorteile von URNs als Standard konsequent genutzt werden können wie z. B. als Referenzen, Suchkriterium oder Zugriffsmechanismus auf das Objekt. [98].

## Metadatenstandards - ein Überblick

### 2.2.17 DIDL- Digital Item Declaration Language

Der Namespace diese Standards wird mit "urn:mpeg:mpeg21:2002:02-DIDL-NS" angegeben und ist in diesem Verweis [100] zu finden. Ein Beispiel für Derived DIDL Types ist unten angeführt:

```
<didl:Item>
  <didl:Descriptor>
    <didl:Statement mimeType="text/xml; charset=UTF-8">
      <dii:Identifier xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS">
        urn:isbn:0-395-36341-1</dii:Identifier>
      </didl:Statement>
    </didl:Descriptor>
    ...
  </didl:Item>
```

Im Appendix D ist ein noch ein weiterführendes Beispiel für ein enthaltenes Objekt, welches im base64 Format kodiert ist. [101]

### 2.2.18 Metadaten in HTML

Meta Daten in HTML definieren nicht einen Satz von Eigenschaften, sondern definieren Attribut-Wert Paare, d.h. einer Variable wird genau ein Wert zugeordnet. Die benutzten Browser müssen keine Metadaten Elemente unterstützen. [42] Ein wichtiger Standard für maschinengestützte Metadaten im Internet, welches zum Großteil als HTML Inhalte darstellt, ist unter anderem der häufig genutzte RDF Standard. [43]

### 2.2.19 "Data Dictionary" - Technische Metadaten für digitale Bilder

Das "Data Dictionary" ist ein Satz von Metadaten für gerasterte digitale Bilder, welches dem Benutzer erlaubt digitale Bilder zu entwickeln, zu interpretieren und auszutauschen. Konzipiert wurde das Datenbeschreibungsverzeichnis um Schnittstellen zu Systemen, Diensten und Software zu ermöglichen. Es unterstützt langfristiges Management und lang bestehenden Zugriff. Die ANSI (American National Standard Institute) hat den Standard im Dezember 2006 anerkannt. [3]. Erweitert wurde der Standard durch den "Data Dictionary - Technical Metadata for Digital Still Images" (ANSI/NISO Z39.87-2006). [71] Siehe auch Appendix B für andere Zertifizierungen in diesem Themenbereich von der NISO.

## Metadatenstandards - ein Überblick

### 2.2.20 RDF - Resource Description Framework

Das RDF (Resource Description Framework) ist ein allgemeines Rahmenwerk für den Austausch von Informationen, die über Anwendungen hinweg ohne Datenverlust transferiert werden können. Die ursprüngliche Absicht war Metadaten im Internet darzustellen. Jetzt können aber auch Informationen, welche zwar im Internet identifiziert sind, aber nicht direkt aufgerufen werden können, abgebildet werden. Die Idee von RDF basiert darauf, dass Objekte mittels URIs identifiziert werden und die Informationen von Anwendungen verarbeitet werden können. Das W3C Semantic Web Activity schlug die Idee vor. [3][49]

#### Das RDF Modell

Das RDF Modell beschreibt Ressourcen in Form von einfachen Eigenschaftsbezeichnungen und Werten. RDF stellt einfache Aussagen über Ressourcen, die mit Tripeln beschrieben werden, dar. Der Gegenstand (subject), das Prädikat (predicate) mit Eigenschaft und das Objekt mit einem zugehörigem Wert sind die Bestandteile von RDF, in den so genannten Tripel enthalten sind. RDF modelliert Aussagen in Form von Knoten und Kanten und zwar so, dass ein Knoten ein Subjekt oder ein Objekt darstellt, und eine Kante bildet ein Prädikat ab, welches direkt vom Gegenstandsknoten (Subject) zum Objektknoten führt. RDF stellt eine XML Syntax zur Verfügung, welche das Aufzeichnen und den Austausch für RDF Kurven ermöglicht. Diese sind RDF/XML und MIME RDF+XML. [3][49]

#### RDF Beispiel

Einer einzigen Ressource können mehrere Werte zugeordnet werden. Ein Objekt, welches in mehreren Dokumenten oder Websites verwendet wird, kann auf folgende Art auf mehrere Punkte bzw. Quellen zeigen. [75]

In diesem einfachen und kurzen Beispiel wird die gesammelte Arbeit eines Autors, sortiert spezifiziert.

Beispiel:

```
<RDF xmlns="http://w3.org/TR/1999/PR-rdf-syntax-19990105#">
  <Seq ID="JSPapersByDate">
    <li resource="http://www.dogworld.com/Aug96.doc"/>
    <li resource="http://www.webnuts.net/Jan97.html"/>
    <li resource="http://www.carchat.com/Sept97.html"/>
  </Seq>
  <Seq ID="JSPapersBySubj">
    <li resource="http://www.carchat.com/Sept97.html"/>
```

## Metadatenstandards - ein Überblick

```
<li resource="http://www.dogworld.com/Aug96.doc" />
<li resource="http://www.webnuts.net/Jan97.html" />
</Seq>
</RDF>
```

### 2.2.21 XMP - Extensible Metadaten Plattform

XMP ist die von Adobe entwickelte erweiterbare Metadaten Plattform. Eine Teilmenge von RDF wird in XML dargestellt. Zu beachten ist, dass XMP gültiges RDF ist, aber RDF Merkmale nicht im XMP Format gültig sind. Speziell zu beachten ist, dass "rdf: RDF Elemente" in XMP notwendig sind und nur optional in RDF bestehen müssen. Elemente, die innerhalb von rdf: .....:rdf stehen, sind beschreibende Elemente. Die rdf:ID und RDF:nodeID Attribute werden nicht beachtet. rdf:aboutEach oder rdf:aboutEachPrefix Attribute werden nicht unterstützt. Das gesamte rdf:Description Element wird ignoriert. rdf:ParseType=´Literal´Attribut wird nicht unterstützt. Die Obersten Typen der RDF Knoten werden auch nicht unterstützt. XMP beschreibt genau das XMP Datenmodell (Data Model), das XMP Speichermodell (Storage Model) und das XMP Schemata. [52][53]

Die Firma Adobe stellt für Interessierte das XMP Toolkit zur Verfügung, welches ein Open Source Toolkit API ist. Die Lizenzierung ist so geregelt, dass Adobe die Urheberrechte auf XMP hat und das XMP SDK unter der BSD Lizenz veröffentlicht wurde. Es gibt natürlich Anwendungen von Adobe, welche XMP unterstützen und auch die MS Vista Photo Gallery nützt XMP. Ein sehr bekanntes Schema ist das IPTC4XMP und das IPTC Core Schema für XMP. Dateiformate, die XMP Metadaten einbetten, sind PDF, HTML und AI (Adobe Illustrator), PSD (Adobe Photoshop), Postscript und EPS, SVG/XML, TIFF, JPEG, JPEG2000, PNG, GIF, DNG (Adobe Digital Negative, das ein Format für rohe digital Daten ist) [51] [50].

## Metadatenstandards - ein Überblick

### 2.2.21.1 XMP Daten Modell

Das XMP Metadaten Modell besteht aus einem Satz von Eigenschaften. Die Eigenschaften sind immer als Ressource auf eine bestimmte Entität bezogen. Die Eigenschaften beschreiben eben diese Ressource. Eine Ressource kann eine Datei, z.B. ein PDF Dokument oder ein wichtiger Bestandteil einer Datei sein. Wenn es ein wirklich wichtiger Bestandteil einer Datei ist, wird das durch die Datei Struktur und die Anwendung bestimmt, die die Datei verarbeitet, z.B. könnte das ein Bild eines PDF Dokumentes sein. Irgendeine gegebene Eigenschaft hat einen Namen und einen Wert. Also jede Eigenschaft macht eine Aussage über eine Ressource: "*Eigenschafts\_Name* der Ressource hat eben den *Eigenschafts\_Wert*". [52]

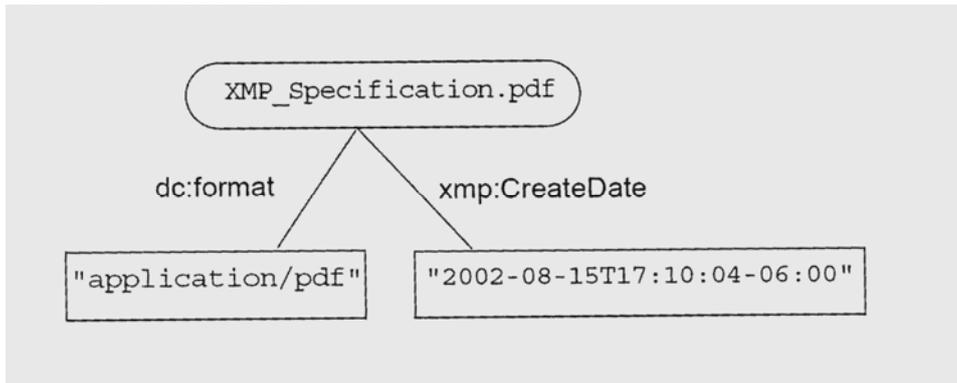
Ein Beispiel wäre: Die Ressource "Moby Dick" (Buch) mit der Eigenschaft: Name: "author" und Wert:"Herman Melville". Das Ergebnis der Auswertung wäre "Der Autor von Moby Dick ist Herman Melville."

XMP Schemata sind typischerweise weniger formal, und werden von der Dokumentation definiert und nicht von maschinenlesbaren Schematadateien. Die Identifikation passiert mit Hilfe des XML Namensraums URI, um Konflikte zwischen den Eigenschaften zu vermeiden. [50-53]

## Metadatenstandards - ein Überblick

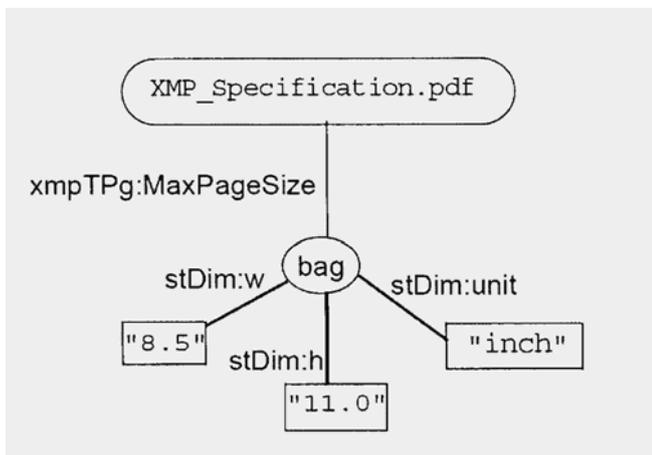
### 2.2.21.2 Eigenschaftswerte

Einfache Typen werden so definiert: Eigenschaftswerte sind einfache Typen, die einzelne Zeichen, Zeichenketten, Booleans, Integer, Entscheidungs- oder reale Nummern Werte haben.



**Abbildung 2-2 Einfache XMP Typen**

Strukturen sind so definiert: Strukturen haben ein oder mehrere Namensfelder und ein Feld kann wiederum aus einer Struktur oder einer Datengruppe (Array) bestehen.



**Abbildung 2-3 XMP Strukturen**

Datengruppen (arrays) sind so definiert: Eine Datengruppe besteht aus einem Satz von Werten. Man könnte sagen, es ist eine Struktur dessen Feldnamen aus Ordnungszahlen bestehen. Die Elemente können einfache Typen, Datengruppen oder Strukturen sein. Zusätzlich sind sie ungeordnet, geordnet oder sind eine alternative Datengruppe.

## Metadatenstandards - ein Überblick

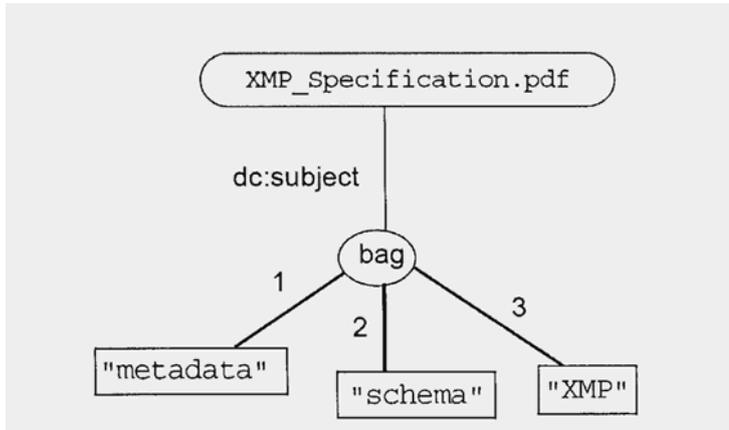


Abbildung 2-4 XMP Arrays

### 2.2.21.3 Eigenschaftsqualitäten

Eigenschaften können "qualifiziert" bzw. erweitert werden indem man zusätzliche Eigenschaften an die eigentlichen Eigenschaften selber hängt, z.B. ein Künstler kann verschiedene Rollen haben:

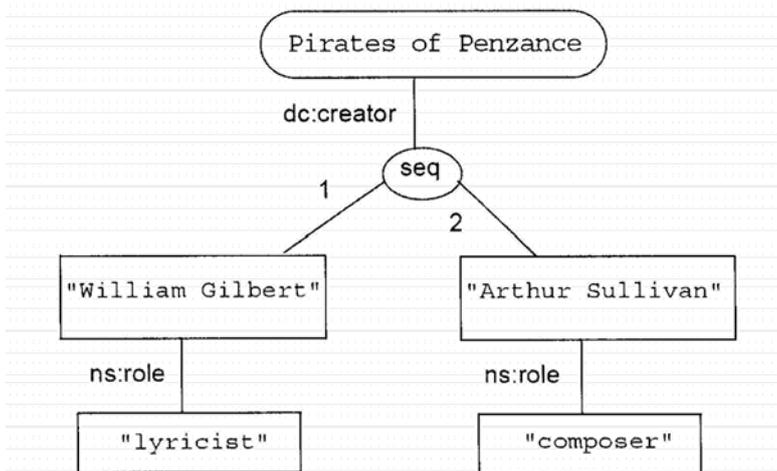


Abbildung 2-5 Eigenschaftsqualitäten

Eine verbreitete Anwendung für „Eigenschaftenqualifizierer“ ist zum Beispiel eine Möglichkeit eine Sprachdatengruppe zu wählen. Der Text wird hier von der benötigten bzw. gewählten Sprache gezogen, z.B. `xml:lang` stimmt überein mit RFC 3066. Standardmäßig benutzt XMP "x-default" und sollte als erstes Element in der Datengruppe des generischen RDF Prozessors als Standardwahl zu finden sein.

## Metadatenstandards - ein Überblick

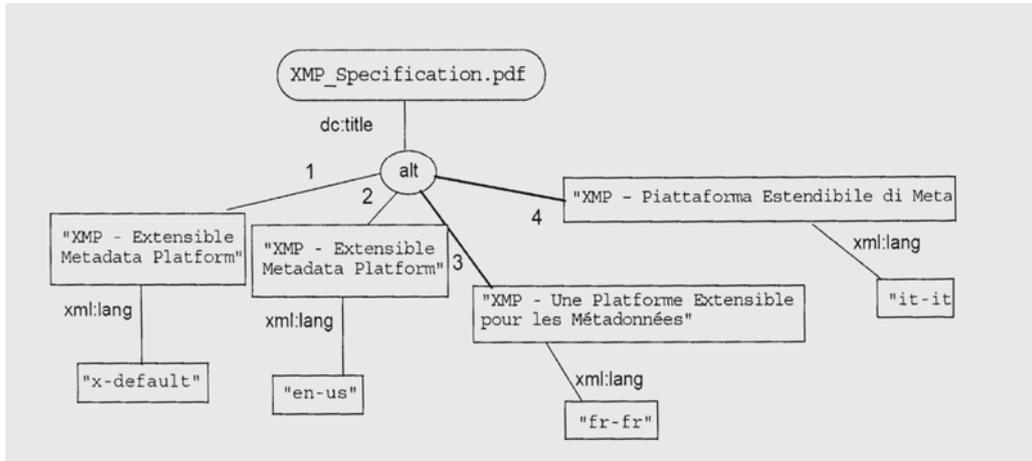


Abbildung 2-6 "Eigenschaftsqualifizierer"

### 2.2.21.4 XMP Speicherungsmodell

Es gibt zwei verschiedene Speicherungsarten: Die interne Speicherung und die externe Speicherung. Die interne Speicherung wird in den Anwendungsdateien eingelagert. Es wird serialisiertes XML benutzt, welches in einen der fünf Unicode Kodierungssätze geschrieben sein muss, und wird in Paketen, welche dann eingebettet werden, aufgeteilt. Wenn die externe Speicherung gewählt wird, so wird ein komplettes, neu gebildetes XML Dokument, welches die Kopfbeschreibung enthält gebildet. Die Dateierweiterung lautet XMP und der MIME Type ist application/rdf+xml. Wichtig ist auch, dass die Anwendung im gleichen Verzeichnis nach einer anderen Zusatzdatei suchen soll, die die Dateiondung .xmp hat.

Der Namensraum URI hat Zeichenketten die mit einer eindeutigen Endung schließen müssen wie zum Beispiel "/" oder "#". Das hat den Grund um Mehrdeutigkeiten zu vermeiden. Zum Beispiel kann während einer Transformierung von XML zu RDF der Namensraum URI mit XML Elementen verbunden werden. Ein Beispiel wie RDF ohne Namensraumteiler Mehrdeutigkeiten erzeugen kann, möchte ich hier darstellen:

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="bogus:example"
    xmlns:ns1="bogus:namespace" xmlns:ns2="bogus:name">
    <ns1:ship>value of ns1:ship</ns1:ship>
    <ns2:spaceship>value of ns2:spaceship</ns2:spaceship>
  </rdf:Description>
</rdf:RDF>
  
```

## Metadatenstandards - ein Überblick

Mehrdeutige RDF Tripel sind die zwei gleichen Prädikate:

Subject: bogus:namespaceship

Prädikat: bogus:namespaceship

Objekt: "value of ns1:ship"

Subjekt: bogus:example

Prädikat: bogus:namespaceship

Objekt: "value of ns2:spaceship"

vgl. [50-53]

### 2.2.21.5 XMP Schema

XMP Metadaten Schemata können mehrere Eigenschaften von einen oder mehreren Schema haben. Die Standard XMP Schemata sind: das Dublin Core Schema, XMP Basis Schema, XMP Rechte Management Schema (Right Management), XMP Medien Management (Media Management), XMP Basis Job Ticket Schema, XMP Paged-Text Schema und das XMP Dynamische Medien Schema. Spezialisierte Schemata sind zum Beispiel das Adobe PDF Schema, das Photoshop Schema, das Camera Raw Schema, das EXIF Schema und das IPTC Schema. [50-53]

### 2.2.22 MPEG-7

Die MPEG (Moving Picture Experts Group) hat den ISO/IEC Standard entwickelt, der einen den wichtigsten Standard für beschreibende Multimedia Inhaltsdaten darstellt. Dieser Standard verwendet XML für das Speichern der Daten und unterstützt die Informationen bis zu einem gewissen Grad auf eine eigene Art und Weise zu interpretieren. MPEG-7 definiert Metadaten Elemente, Strukturen und Beziehungen, um audiovisuelle Objekte zu beschreiben. Diese Objekte können Standbilder, Grafiken, 3D Modelle, Musik, Ton, Sprache, Video und Multimedia Sammlungen sein. [56]. MPEG-7 liefert für diese Aufgaben Werkzeuge für die Beschreibung, selbst, die Definition der Beschreibungssprache und einige System Software-Werkzeuge [55]

#### 2.2.22.1 MPEG-7 Multimedia Beschreibungswerkzeuge

Einerseits gibt es Deskriptoren (descriptors) und andererseits Beschreibungsschemata, die von diesen MPEG-7 Tools erzeugt werden können. [57] Die Deskriptoren definieren die Syntax und die semantischen Eigenschaften jedes Metadatenelements Die Beschreibungsschemata spezifizieren die Struktur und die Beziehung zwischen den Elementen. Im Allgemeinen stellt MPEG-7 eine Reihe von beschreibenden Werkzeugen zur Verfügung, die auf der einen Seite mit generischen und auf der anderen auch mit multimedialen Entitäten umgehen können.

## Metadatenstandards - ein Überblick

Generische Beschreibungswerkzeuge: Generische Beschreibungswerkzeuge werden benützt, wenn Medien wie Ton und Video beschrieben werden sollen.

Inhaltsbeschreibende Werkzeuge: Der Inhalt wird mit der Darstellung der wahrnehmbaren Informationen beschrieben.

Navigation und Zugriff: Spezifiziert die Zusammenfassungen und die Vielfalt der AV Inhalte

Benutzer Interaktion: Stellt eine Beschreibung der Benutzerpräferenzen und die Verlaufsgeschichte betreffend dem Konsum multimedialen Materials der Benutzer.

Inhaltsorganisation: Die Darstellung der Analyse und die Klassifizierung des AV Inhalts.

Inhaltsverwaltung: Die Informationen über die Medienmerkmale, d.h. die Erschaffung und die Verwendung von AV Inhalten.

[58]

### 2.2.22.2 Beschreibende Definitions- Sprache

Mit der Beschreibenden Definitions- Sprache ist es möglich neue Beschreibungsschemata zu erzeugen. Diese Sprache basiert auf der XML Schema Sprache und stellt eine Erweiterung für audiovisuellen Inhalt dar. [58]

### 2.2.22.3 System Werkzeuge (Software)

Die System Tools unterstützen die Speicherung, die Übermittlung, die Synchronisation der Beschreibung mit Inhalt, die Verwaltung und den Schutz von Urheberrechten. [58]

### 2.2.22.4 MPEG-7 Visuell

Die Grundstrukturen und die beschreibenden Elemente (descriptors) decken die wesentlichen visuellen Merkmale ab wie z.B. Farbe, Textur, Zustand, Bewegung, Lokalisierung und die Gesichtserkennung. [58]

### 2.2.22.5 MPEG-7 Audio

Es gibt zwei verschiedene Ebenen auf denen die MPEG-7 Ton Deskriptoren angesiedelt sind. Einerseits gibt es hier beschreibende Daten auf einer unteren und andererseits auf einer oberen Ebene. Auf der unteren Ebene werden allgemein beschreibende Elemente in Ton-Objekten quer durch alle Anwendungen verwendet, zb. bei spektralen-, parametrisierten- und auch zeitlichen Signaleigenschaften. Die Beschreibung auf höherer Ebene ist bei bestimmten Anwendungen spezifisch gehalten. Allgemeine Stimmenerkennung und Inhaltsbeschreibende Werkzeuge stehen auf dieser Stufe ebenso zur Verfügung wie auch instrumentelles Timbre (Klangfarbe) Werkzeuge. Es gibt in diesem Bereich auch gesprochene Inhaltsbeschreibungen, die von Software unterstützt werden, auch Tonsignaturen und melodische

## Metadatenstandards - ein Überblick

Beschreibungswerkzeuge kommen zum Einsatz. In der Literatur wird in diesem Zusammenhang auch von "query-by-humming" gesprochen. [58]

### 2.2.22.6 MPEG Profile

Ein Profil ist eine Teilmenge von Werkzeugen, die im MPEG-7 Standard definiert sind und einige Funktionalitäten für eine oder mehrere Kategorien von Anwendungen bereitstellen. [57]

### 2.2.23 MPEG-21

MPEG-21 stellt ein Gerüst für die Interoperabilität von digitalen multimedialen Objekten zur Verfügung. Es ist ein offenes Rahmenwerk für alle Beteiligten, die Multimediainhalte liefern oder konsumieren. Das Zusammenspiel von Angebot und Nachfrage in der so genannten Kette von digitaler Lieferung und digitalem Konsum ist speziell zu erwähnen. Das Benutzermodell sieht den Benutzer als Entität, welche in den MPEG-21 Umgebungen interagiert und digitale Dinge verwendet. Das Benutzermodell besteht aus Individuen, Konsumenten, Gemeinschaften, Organisationen, Unternehmen, Konsortien, Regierungen und anderen Körperschaften und Initiativen auf der Welt. Die Benutzer werden gezielt durch Beziehungen zu anderen Benutzern mit gewissen Interaktionen identifiziert. Digitale Dinge (Digital Item) sind gekennzeichnet durch den zweiten Teil der MPEG-21 Erklärung, d.h. ein Beschreibung eines digitalen Dinges (digital item declaration, kurz DID) ist ein Dokument, welches den Aufbau, die Struktur und die Organisation eines DID genau bestimmt. Bis zum ersten Quartal 2007 sind 15 Teile als ISO Standards ISO/IEC TR 21000 publiziert worden.

- Part 1: Vision, Technologies and Strategy
- Part 2: Digital Item Declaration (DID)
- Part 3: Digital Item Identification (DII)
- Part 4: Intellectual Property Management and Protection (IPMP)
- Part 5: Rights Expression Language (REL)
- Part 6: Rights Data Dictionary (RDD)
- Part 7: Digital Item Adaption (DIA)
- Part 8: Reference Software
- Part 9: File Format
- Part 10: Digital Item Processing (DIP)
- Part 11: Evaluation Methods for persistent Association Technologies
- Part 12: Test Bed for MPEG-21 Resource Delivery
- Part 13: Scalable Video Coding (in Bearbeitung)
- Part 14: Conformance Testing (in Bearbeitung)
- Part 15: Event Reporting
- Part 16: Binary Format
- Part 17: Fragment Identification of MPEG Resources
- Part 18: Digital Item Streaming (in Bearbeitung) [59]

## Feldstudie

### 3 Feldstudie

Die Feldstudie wurde mittels eines Fragebogens der online auszufüllen war, in Zusammenarbeit mit dem Kompetenzzentrum Interoperable Metadaten (**KIM**) [86] durchgeführt. Die Programmierung erfolgte in HTML Coding nach dem Standard des W3C Konsortiums [72], d.h. jeder Teilnehmer mit gängigem Internet Browser, der diesen Standard beherrscht, konnte an der Feldstudie teilnehmen. Die Teilnehmer wurden mittels unabhängiger Recherche ermittelt und nahmen freiwillig an der Feldstudie teil.

Es wurden über 200 verwertbare Einträge erfasst und in weiterer Folge genau untersucht. Beginnend mit der Beschreibung des Aufbaus werden abschließend die Ergebnisse in Form von Diagrammen dargestellt. Die erhaltenen Daten der Studie wurden mit MS Access in ein triviales Datenbankschema überführt. Die Informationen dort mit SQL Befehlen selektiert bzw. diverse Abfragen erstellt und zuletzt werden die Ergebnisse in MS Excel in Diagrammen dargestellt.

#### 3.1 Aufbau

Die Feldstudie wurde in 4 Fragen unterteilt, die in diesem Kapitel beschrieben werden. Die Umfrage ermittelt die Art der Einrichtung bzw. des Unternehmens der teilnehmenden Institutionen. Die Befragten konnten angeben, ob sie in einem kommerziellen Unternehmen, einer Behörde, einer öffentlichen Einrichtung oder einer Non-Profit-Organisation beschäftigt sind. Zudem gab es für Teilnehmer, deren Institution keiner dieser Angaben entsprach, die Möglichkeit eigene Angaben zu machen. Zu der Art der Einrichtung wurde ermittelt, welcher Branche/Domäne die teilnehmenden Institutionen angehören.

##### 3.1.1 Ausgewählte Fragen

###### 3.1.1.1 Frage 1

a) Sie arbeiten in

- einem kommerziellen Unternehmen
- einer Behörde/Öffentlichen Einrichtung
- einer Non-Profit-Organisation
- Andere (bitte angeben)

## Feldstudie

b) In welcher Branche/Domäne ist ihre Organisation tätig? (Mehrfachnennung ist möglich)

- Information
- Medien
- IT
- Industrie
- Kultur
- Verwaltung
- Gesundheitswesen/Pharma
- Forschung
- Bildung
- Andere
- Welche?

### 3.1.1.2 Frage 2

a) Werden in Ihrer Organisation oder Firma für die Beschreibung von Objekten strukturierte Informationen - Metadaten - verwendet (z. B. um Suche oder Browsen zu unterstützen)?

- Ja
- Nein

b) Wenn ja, für welche Objekte und mit Hilfe welcher Standards (z. B. Dublin Core, ONIX, hausinterne Standards)? (Mehrfachnennung ist möglich)

- Produkte
- Standards
- Waren
- Standards

## **Feldstudie**

- Personen
- Standards
- Bücher
- Standards
- Filme
- Standards
- Webseiten
- Standards
- Museumsstücke
- Standards
- Musik
- Standards
- Andere
- Welche ?
- Standards
- Welche ?
- Standards

### **3.1.1.3 Frage 3**

a) Hat Ihre Organisation oder Firma den Bedarf, strukturierte Informationen (Metadaten) aus unterschiedlichen (Datenbank-) systemen in bestandsübergreifende Systeme zusammenzuführen?

- Ja
- Nein

## Feldstudie

b) Wenn ja, im Rahmen welcher Anwendungsszenarien? (Mehrfachnennung ist möglich)

- Zusammenspielen organisationsinterner Datenbestände
- Gemeinsame Suche über verschiedene organisationsinterne Datenbanken
- Datenaustausch mit externen Partnern
- Einbindung von externen Informations-/Datenangeboten
- Andere
- Welche?

### 3.1.1.4 Frage 4

a) Hat Ihre Organisation oder Firma Interesse an einer Unterstützung, um das Zusammenführen von strukturierten Informationen aus verschiedenen Datenbanksystemen zu erleichtern und zu optimieren?

- Ja
- Nein

b) Wenn ja, welche Angebote wären für Sie interessant? (Mehrfachnennung ist möglich)

Informationsmaterial

- Lehrbücher
- eLearning
- Schulungen
- Workshops
- Beratungen
- Andere
- Welche?

## Feldstudie

### 3.1.1.5 Zusatz

a) Ich beantworte diesen Fragebogen im Namen meiner Organisation/Firma/Behörde

- Ja
- Nein

b) Name Ihrer Organisation oder Firma

c) Wenn Sie Interesse an den Ergebnissen dieser Umfrage haben, tragen Sie hier Ihre E-Mail-Adresse ein.

- E-Mail
- Abschicken

## Feldstudie

### 3.2 Teilnehmer

Das folgende Diagramm zeigt die Verteilung der Teilnehmer nach Einteilung in Branchen.

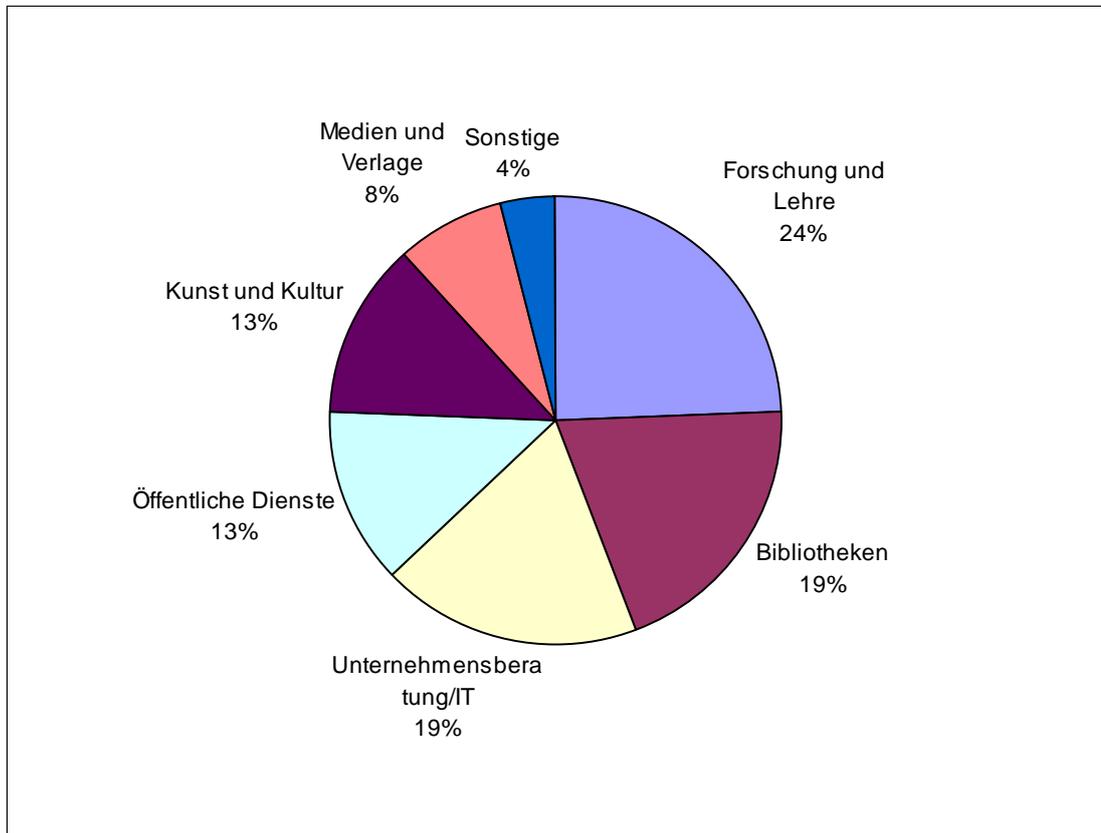


Abbildung 3-1 Teilnehmer der Feldstudie

### 3.3 Ziel

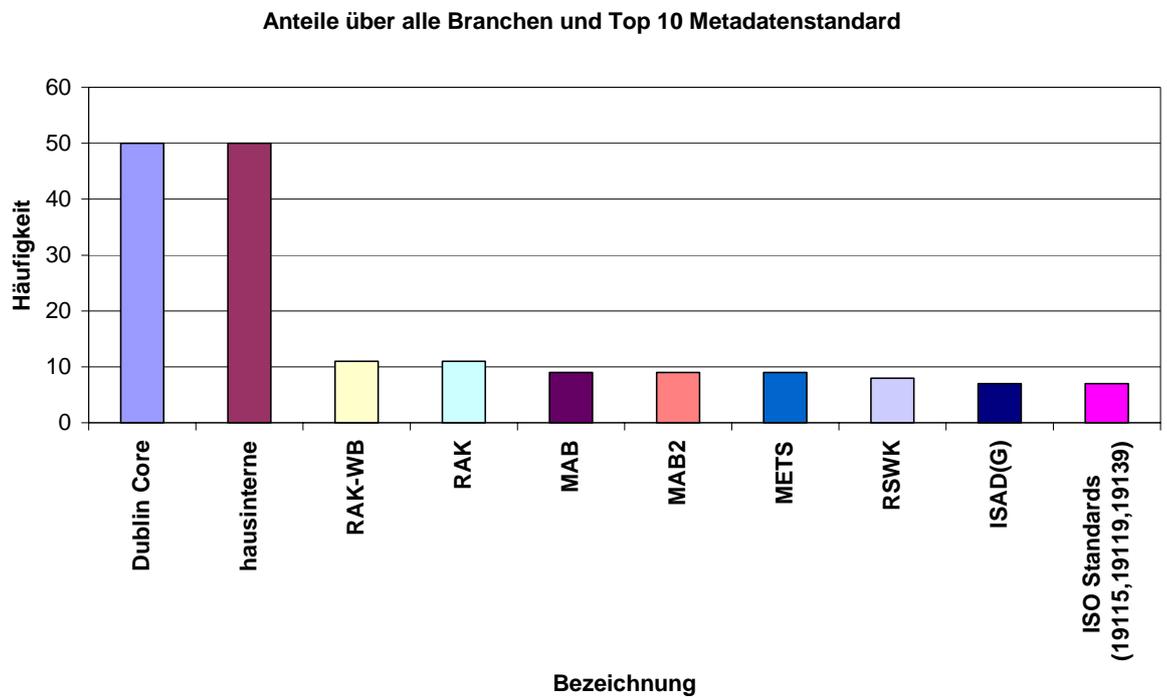
Das Primärziel der Umfrage war herauszufinden, welche Metadatenstandards in welchen Branchen verwendet werden. Ebenso ob die Daten in andere Systeme getauscht werden. Sekundäre Ziele der Umfrage in Rahmen der Diplomarbeit war, das Wissen über den Bedarf an Dienstleistungen wie Fort- und Weiterbildungsmaßnahmen und über die Wünsche von Metadatenanwendern und -entwicklern zu erweitern, um Angebote im Bereich interoperable Metadaten bestmöglich mit den bestehenden Verhältnissen zu erheben.

## Feldstudie

### 3.1 Auswertungen der Feldstudie mit Diagrammen

#### 3.1.1 Auswertung Top 10 Metadatenstandards

Das folgende Diagramm zeigt die Top 10 Metadatenstandards nach Auswertung der Feldstudie über alle Teilnehmer und Branchen.



**Abbildung 3-2** Anteile über alle Branchen und Top 10 Metadatenstandard



## Feldstudie

### 3.1.3.1 Bibliotheken

Das folgende Diagramm zeigt die verwendeten Top 10 Metadatenstandards in Bibliotheken.

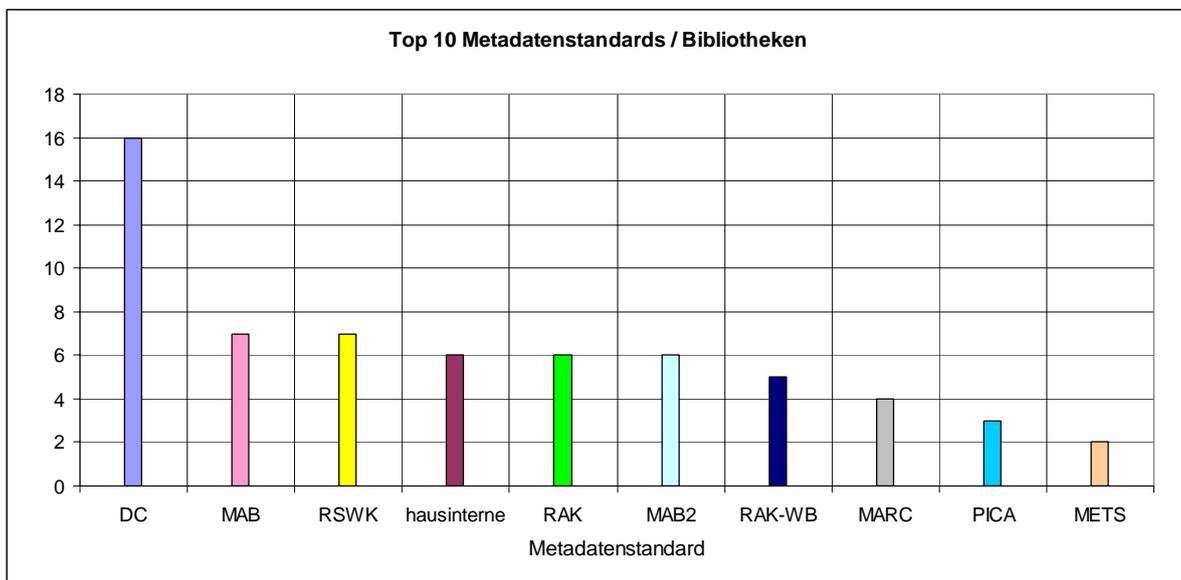


Abbildung 3-4 Top 10 Metadatenstandards / Bibliotheken

### 3.1.3.2 Forschung

Das folgende Diagramm zeigt die verwendeten Top 10 Metadatenstandards in Forschung und Lehre.

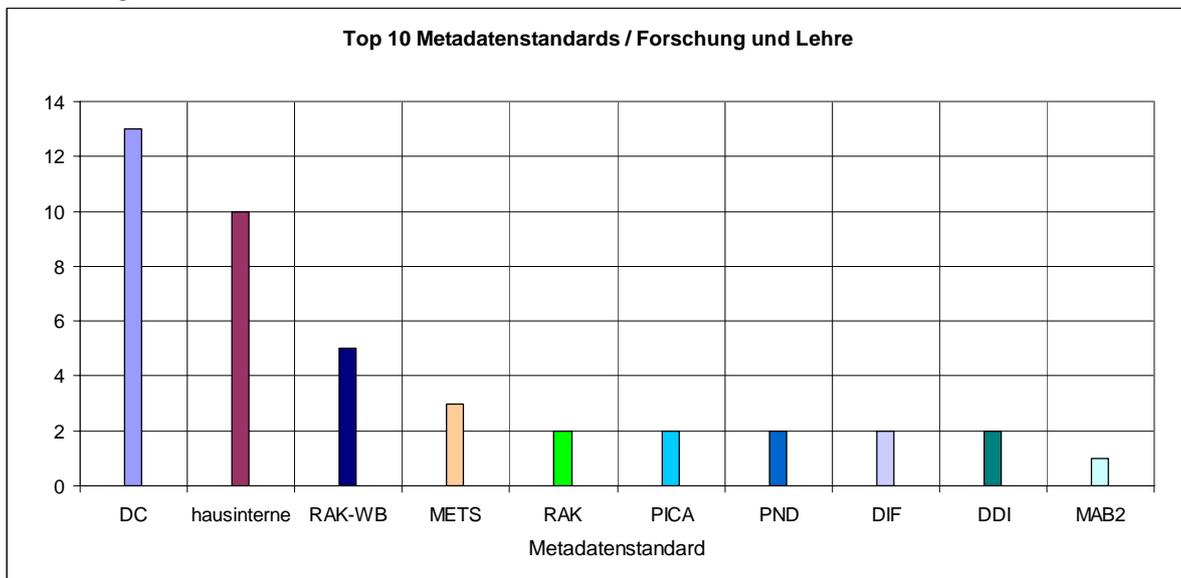


Abbildung 3-5 Top 10 Metadatenstandards / Forschung und Lehre

## Feldstudie

### 3.1.3.3 Medien

Das folgende Diagramm zeigt die verwendeten Metadatenstandards in Medien und Verlage.

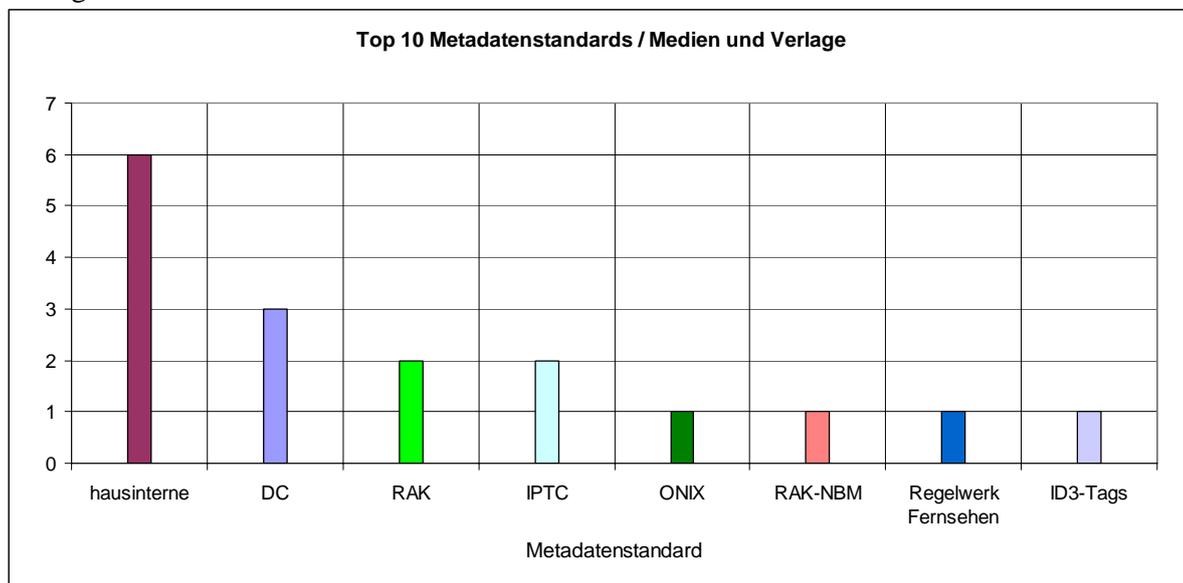


Abbildung 3-6 Top 10 Metadatenstandards / Medien und Verlage

### 3.1.3.4 Verwaltung

Das folgende Diagramm zeigt die verwendeten Top 10 Metadatenstandards im Öffentlichen Dienst.

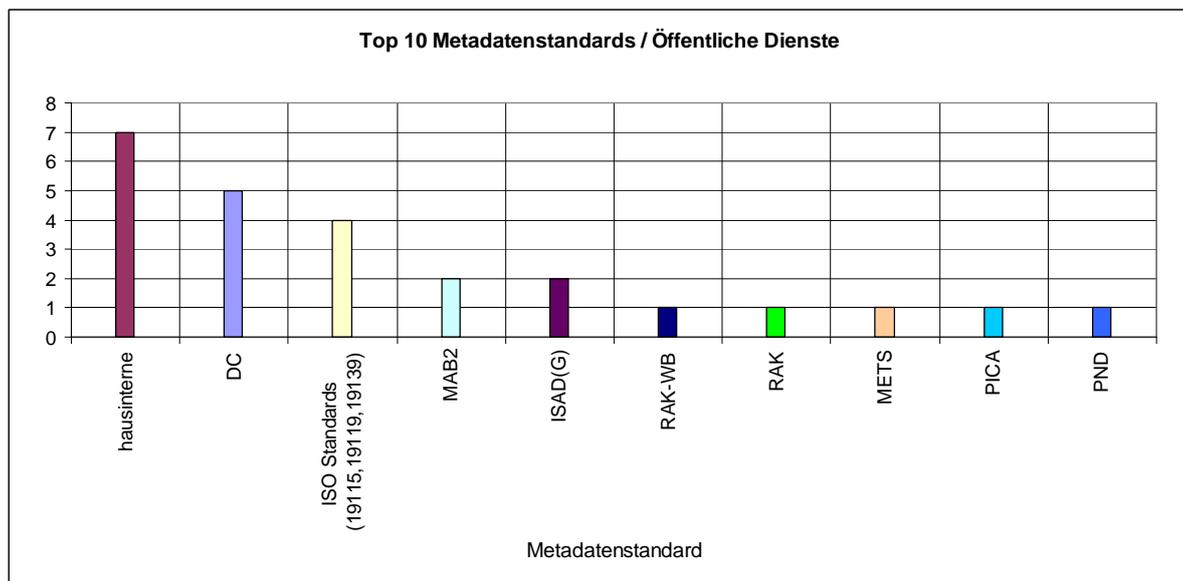


Abbildung 3-7 Top 10 Metadatenstandards / Verwaltung

## Feldstudie

### 3.1.3.5 Kultur

Das folgende Diagramm zeigt die verwendeten Top 10 Metadatenstandards im Bereich Kunst und Kultur.

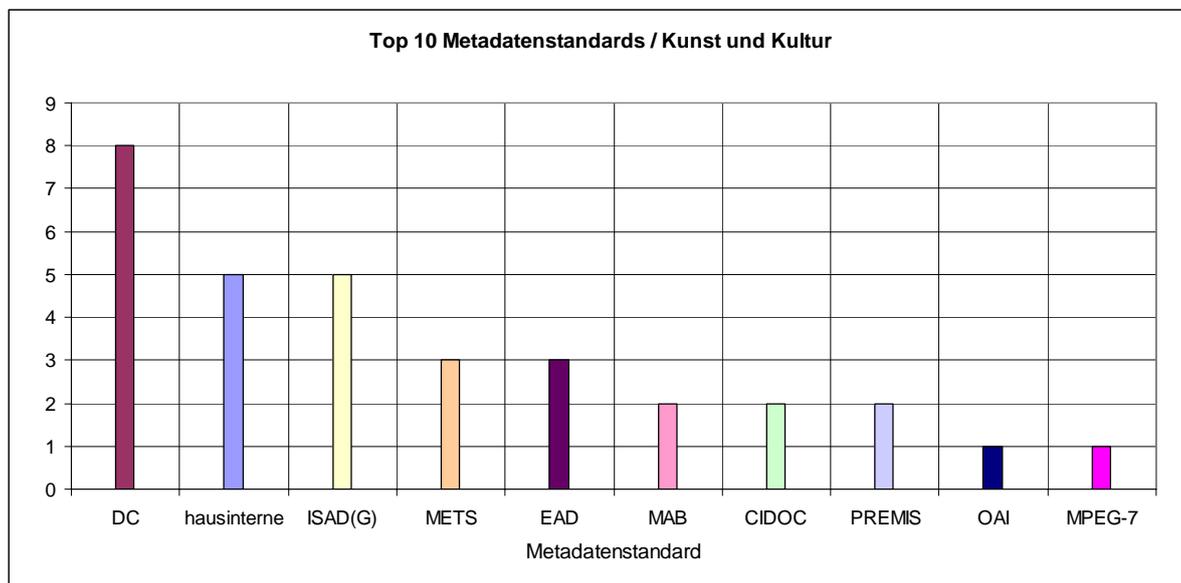


Abbildung 3-8 Top 10 Metadatenstandards / Kultur

### 3.1.3.6 IT

Das folgende Diagramm zeigt die verwendeten Top 10 Metadatenstandards im Bereich Unternehmensberatung und IT-Dienstleistungen.

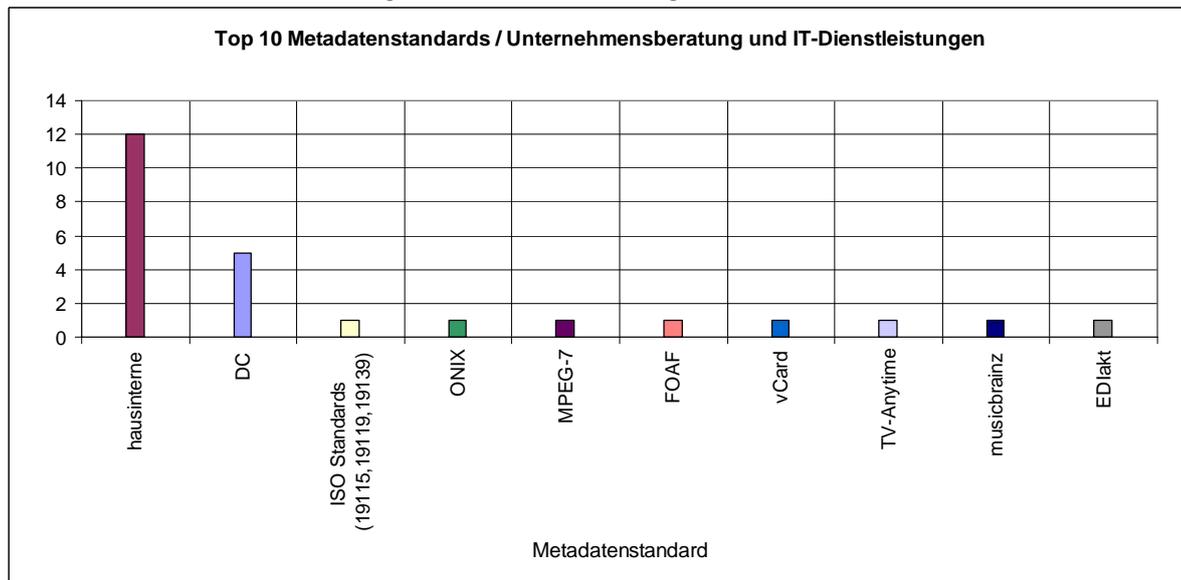


Abbildung 3-9 Top 10 Metadatenstandards / IT

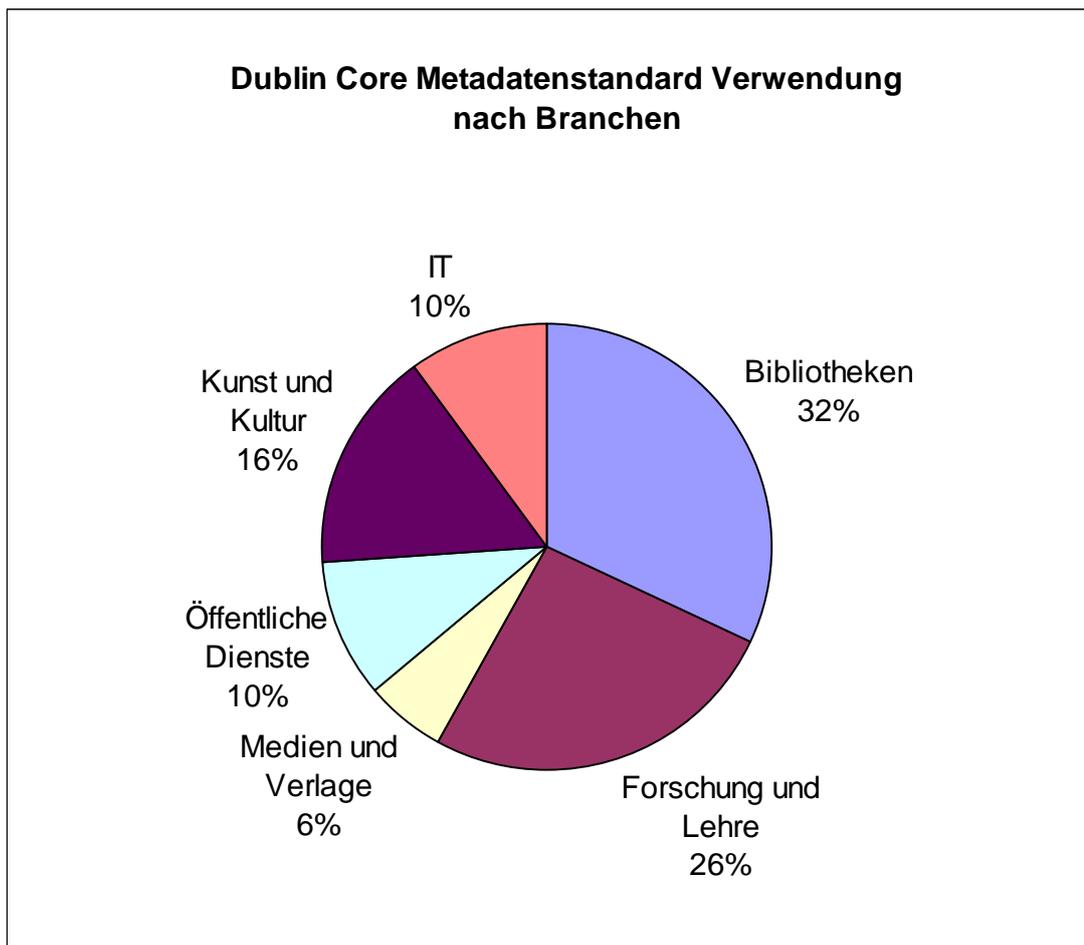
## Feldstudie

### 3.1.4 Auswertung nach Metadatenstandards

Die Auswertung nach Metadatenstandards gibt einen Überblick welche Standards wie häufig in welchen Branchen verwendet werden.

#### 3.1.4.1 Dublin Core

Das folgende Diagramm zeigt die Häufigkeit der Anwendung des Dublin Core Standards in verschiedenen Branchen

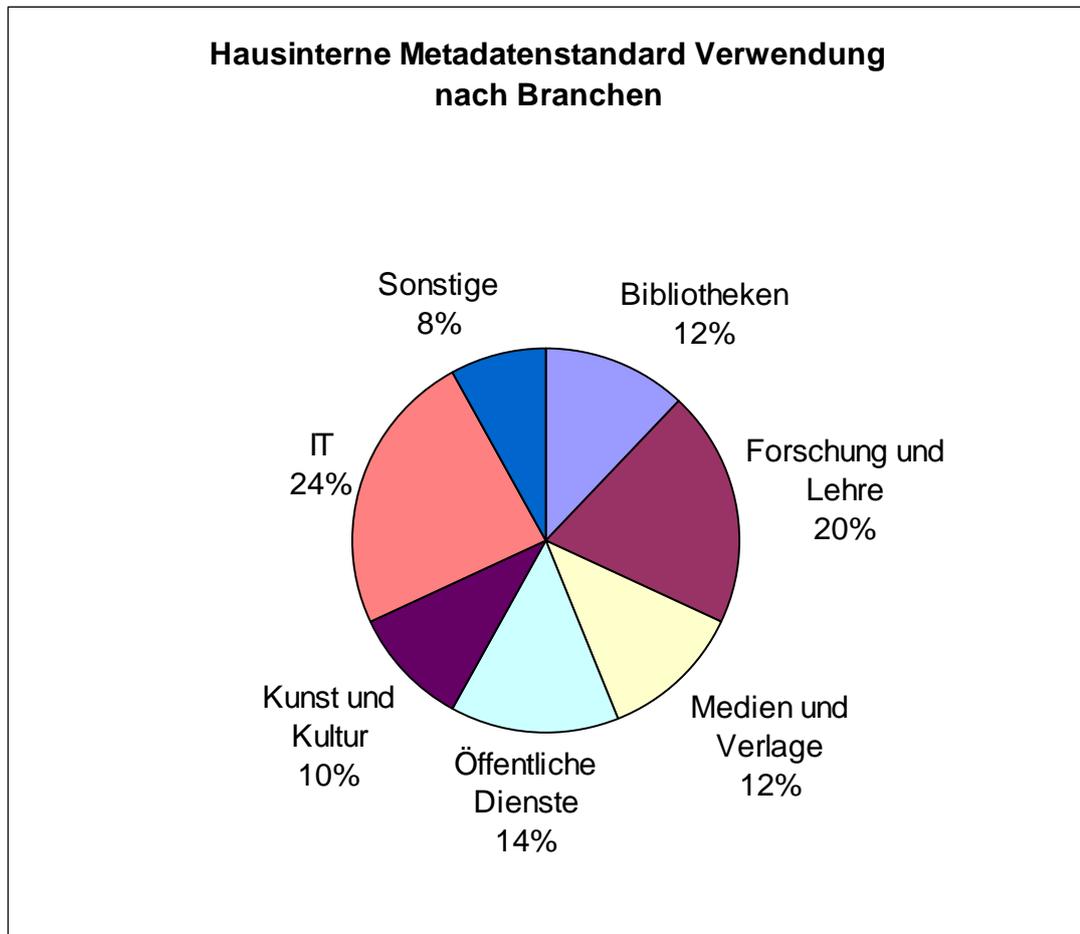


**Abbildung 3-10 Dublin Core Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.2 Proprietäre Standards

Das folgende Diagramm zeigt die Verteilung hausinternen Standards in verschiedenen Branchen

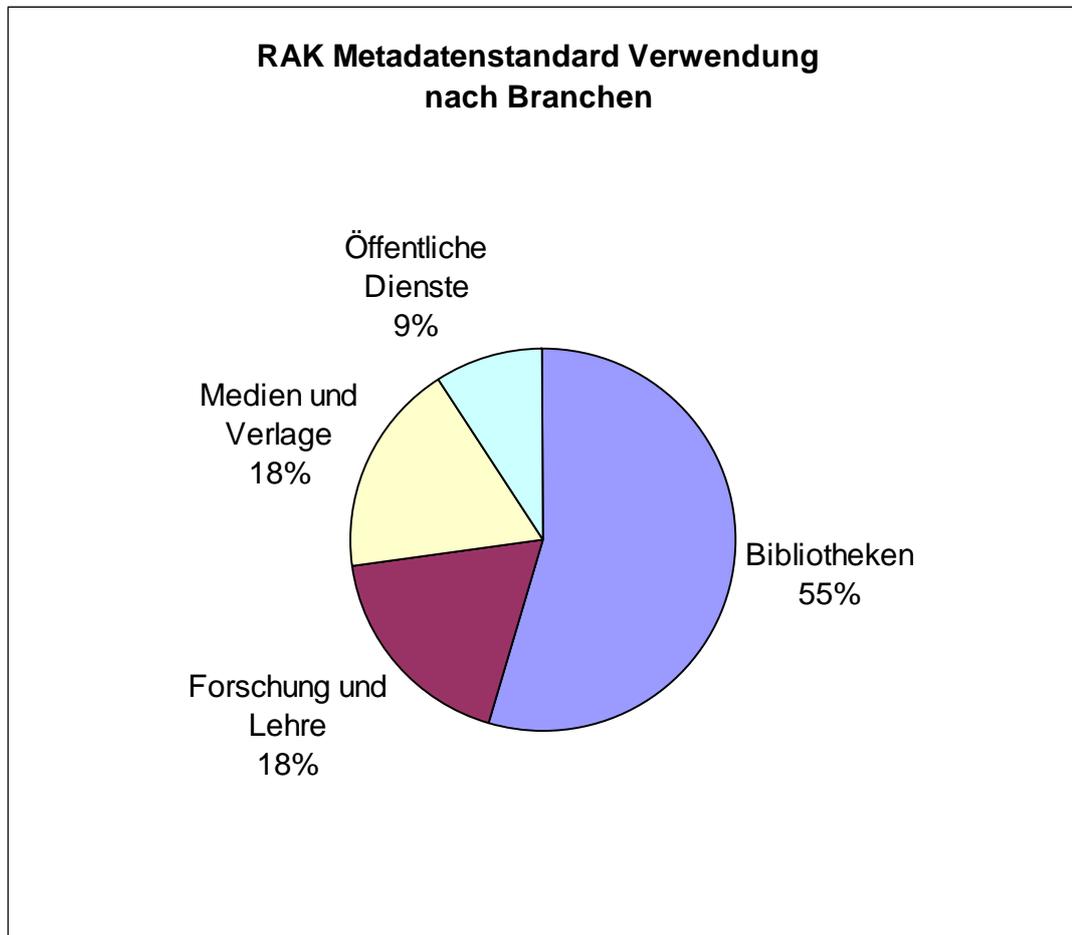


**Abbildung 3-11 Hausinterne Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.3 RAK

Das folgende Diagramm zeigt die Verteilung des RAK Standards in Branchen

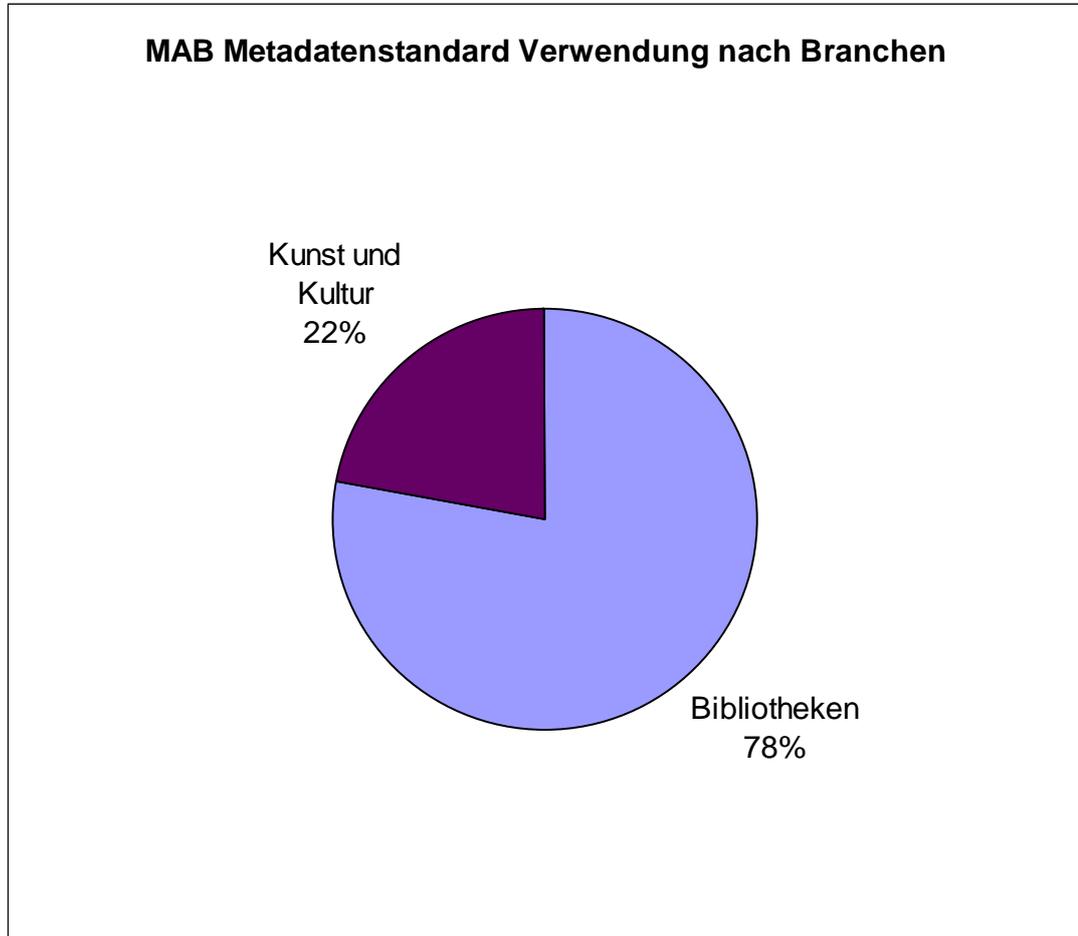


**Abbildung 3-12 RAK Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.4 MAB

Das folgende Diagramm zeigt die Verteilung des MAB Standards in Branchen

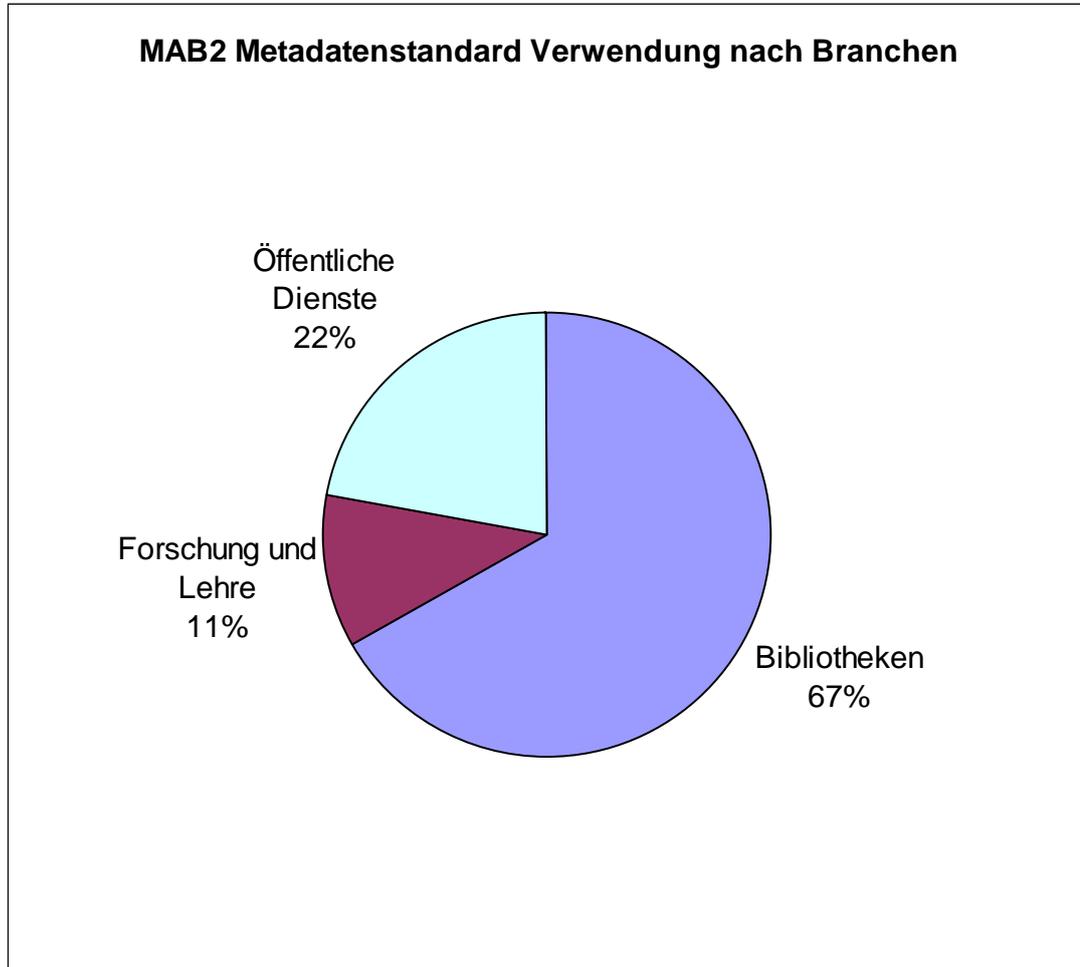


**Abbildung 3-13 MAB Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.5 MAB2

Das folgende Diagramm zeigt die Verteilung des MAB2 Standards in Branchen

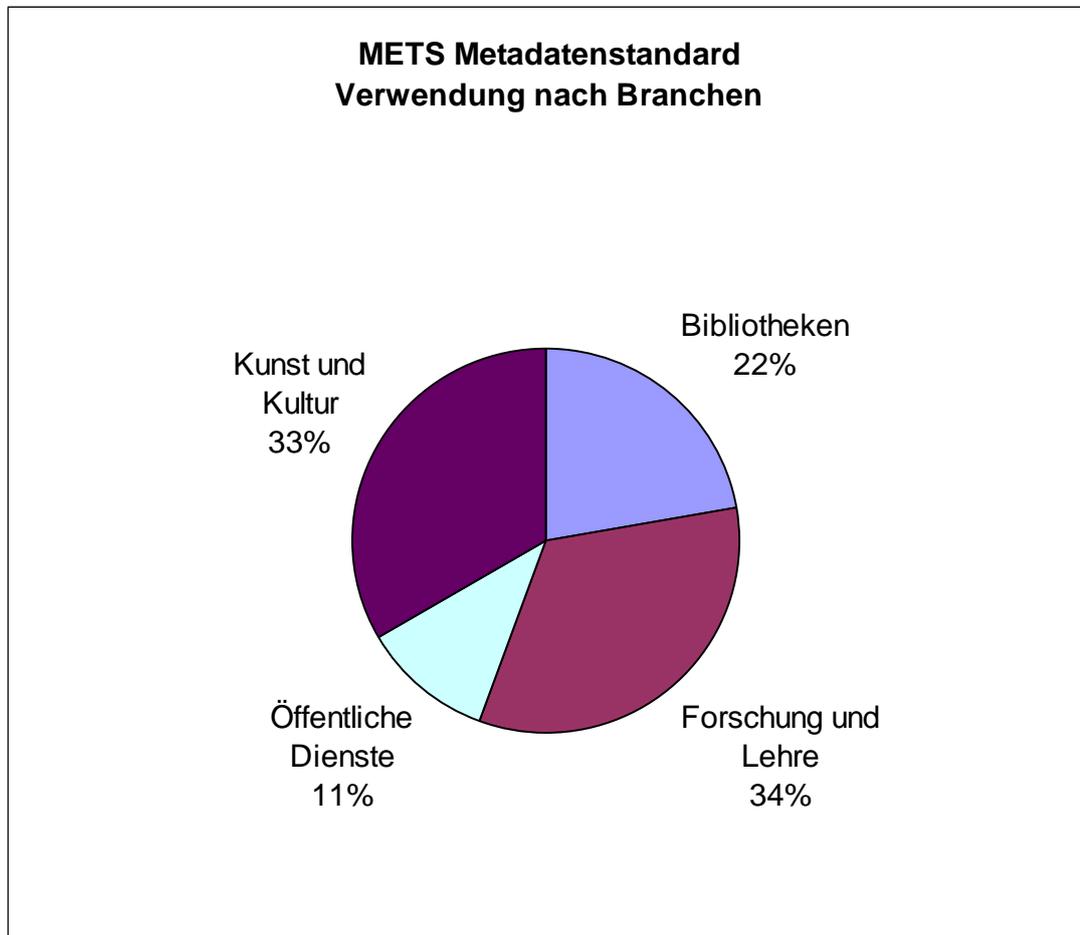


**Abbildung 3-14 MAB2 Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.6 METS

Das folgende Diagramm zeigt die Verteilung des METS Standards in Branchen

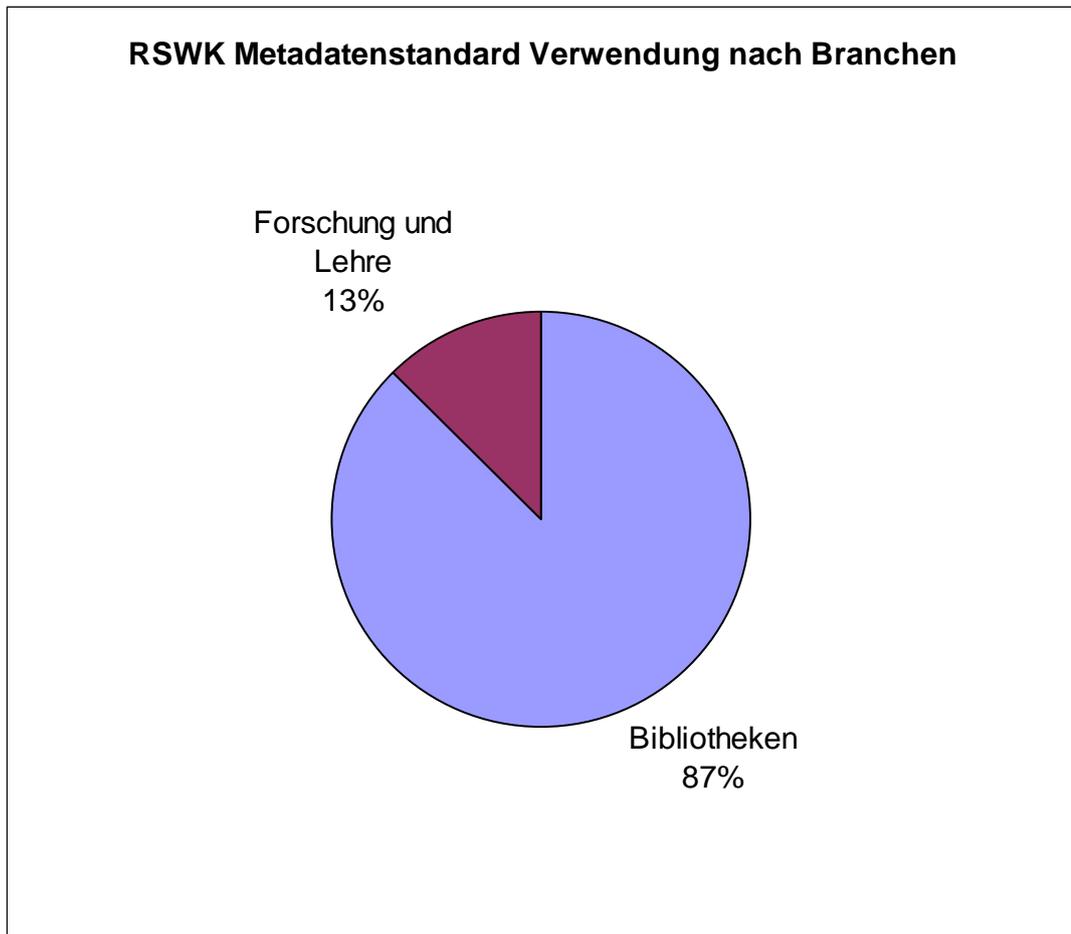


**Abbildung 3-15 METS Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.7 RSWK

Das folgende Diagramm zeigt die Häufigkeit der Anwendung des RSWK Standards in Branchen

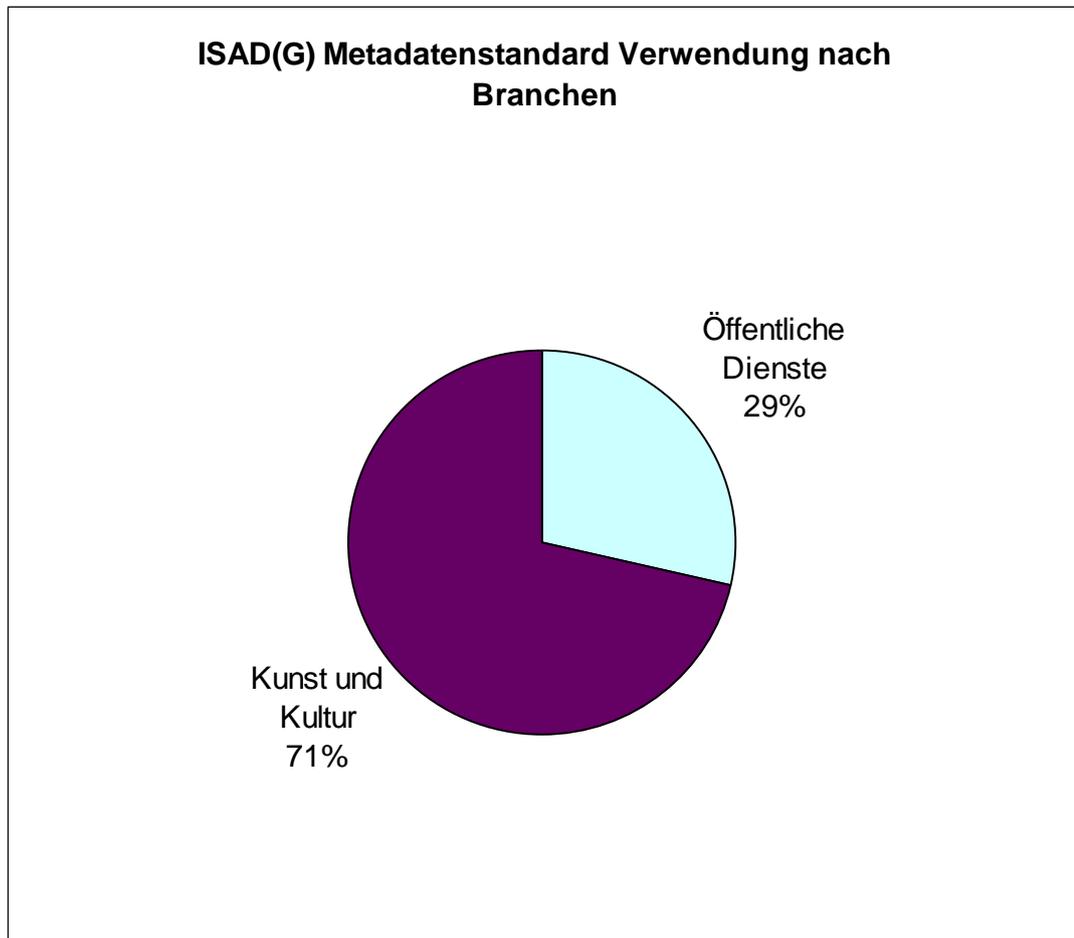


**Abbildung 3-16 RSWK Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.8 ISAD

Das folgende Diagramm zeigt die Verteilung des ISAD Standards in den Branchen Öffentliche Dienste und Kunst & Kultur.

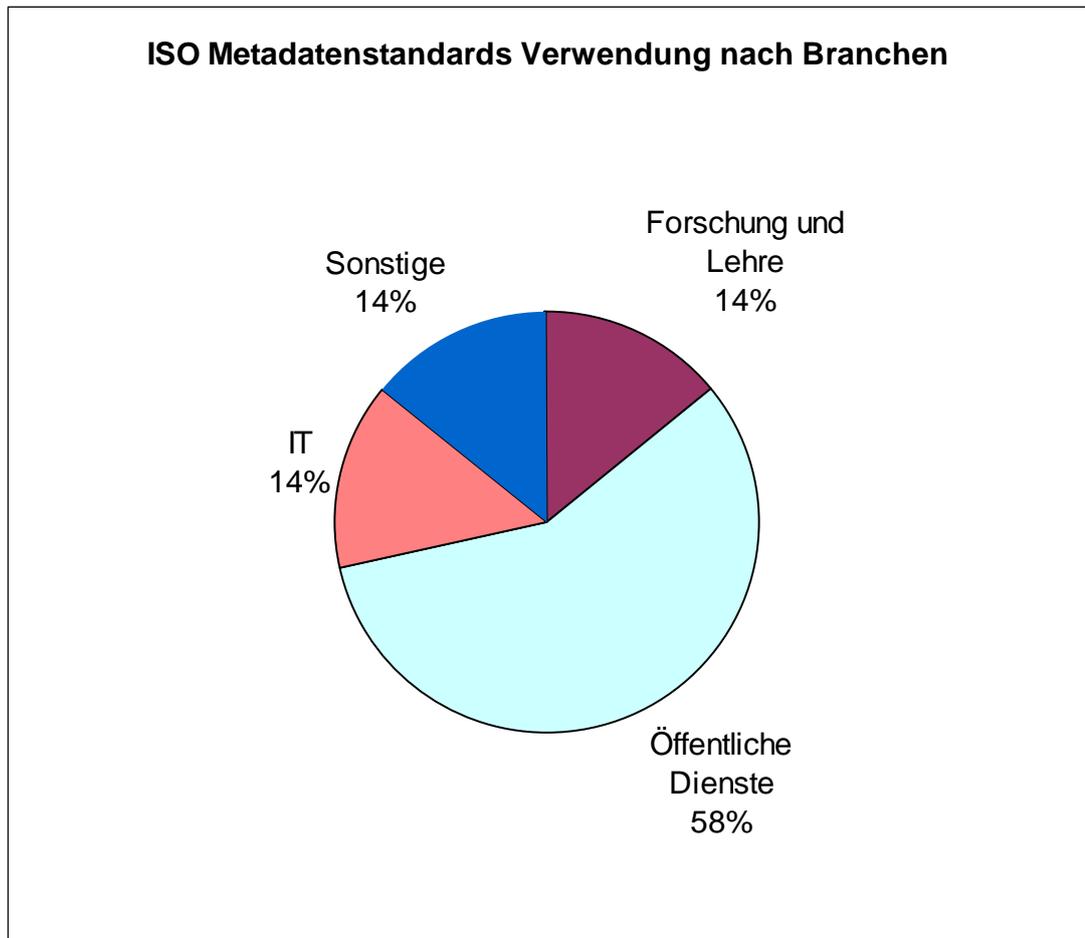


**Abbildung 3-17 ISAD(G) Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.4.9 ISO Metadatenstandards (19115, 19119, 19139)

Das folgende Diagramm zeigt die Häufigkeit der Anwendung der ISO Standards in Branchen, die dieses Format benutzen.

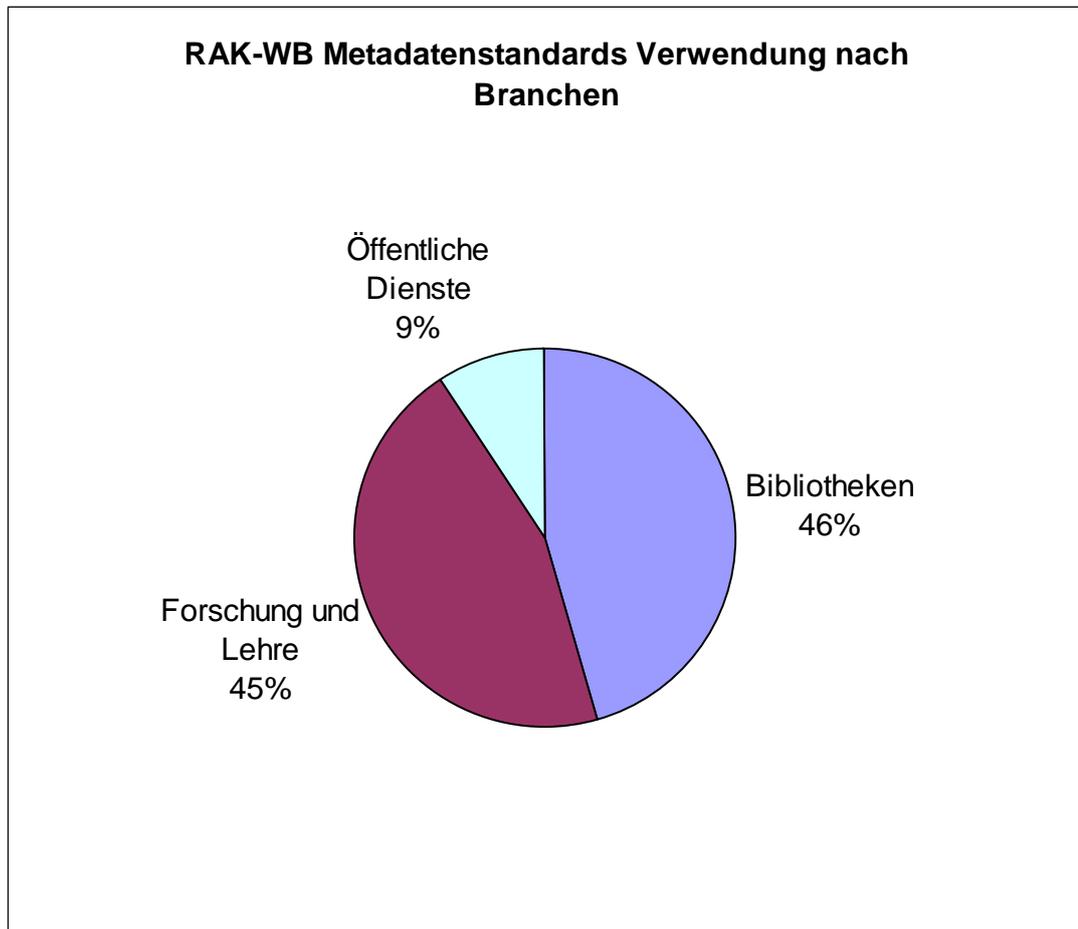


**Abbildung 3-18 ISO Metadatenstandards (19115,19119,19139) Verwendung**

## Feldstudie

### 3.1.4.10 RAK-WB

Das folgende Diagramm zeigt die Verteilung des RAK-WB Standards in den Branchen, die den Standard anwenden.



**Abbildung 3-19 RAK-WB Metadatenstandard Verwendung nach Branchen**

## Feldstudie

### 3.1.5 Auswertung: Zusammenführung von Systemen

Die Feldstudie ermittelt den Bedarf der Institutionen nach Zusammenführung von Metadaten aus unterschiedlichen Systemen in bestandsübergreifende Systeme. Die Teilnehmer gaben an, ob oder ob kein Bedarf in Ihrer Institution existiert. Institutionen, in denen Bedarf vorhanden ist, konnten angeben, im Rahmen welcher Anwendungsszenarien die Metadaten zusammengeführt werden. Zu den vier vorgegebenen Anwendungsszenarien gab es unter anderem die Möglichkeit, verschiedene Anwendungsszenarien zu nennen. Es waren Mehrfachantworten möglich. Die folgende Abbildung gibt eine Übersicht über die Auswahl der Anwendungsszenarien: [104]

#### Frage 3

a) Hat Ihre Organisation oder Firma den Bedarf, strukturierte Informationen (Metadaten) aus unterschiedlichen (Datenbank-) systemen in bestandsübergreifende Systeme zusammenzuführen?

- Ja  
 Nein

b) Wenn ja, im Rahmen welcher Anwendungsszenarien?  
(Mehrfachnennung ist möglich)

- Zusammenspielen organisationsinterner Datenbestände  
 Gemeinsame Suche über verschiedene organisationsinterne Datenbanken  
 Datenaustausch mit externen Partnern  
 Einbindung von externen Informations-/Datenangeboten  
 Andere

Welche?

#### Frage 4

a) Hat Ihre Organisation oder Firma Interesse an Unterstützung, um das Zusammenführen von strukturierten Informationen aus verschiedenen Datenbanksystemen zu erleichtern und zu optimieren?

- Ja  
 Nein

b) Wenn ja, welche Angebote wären für Sie interessant?  
(Mehrfachnennung ist möglich)

- Informationsmaterial  
 Lehrbücher  
 eLearning  
 Schulungen  
 Workshops  
 Beratungen  
 Andere

Welche?

---

### Abbildung 3-20 Auswahl Anwendungsszenarien

## Feldstudie

Die Auswertung ergibt, dass die meisten teilnehmenden Institutionen einen Bedarf an der Zusammenführung von Metadaten in bestandsübergreifende Systeme haben. 84% der Befragten bejahten den Bedarf, 15% gaben an, dass kein Bedarf vorliegt und 1% machten keine Angaben. [104]

Bedarf an Zusammenführung von Strukturierten Informationen

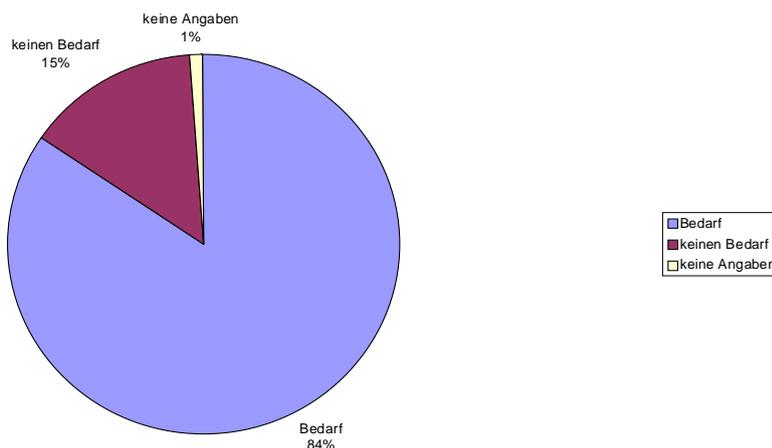
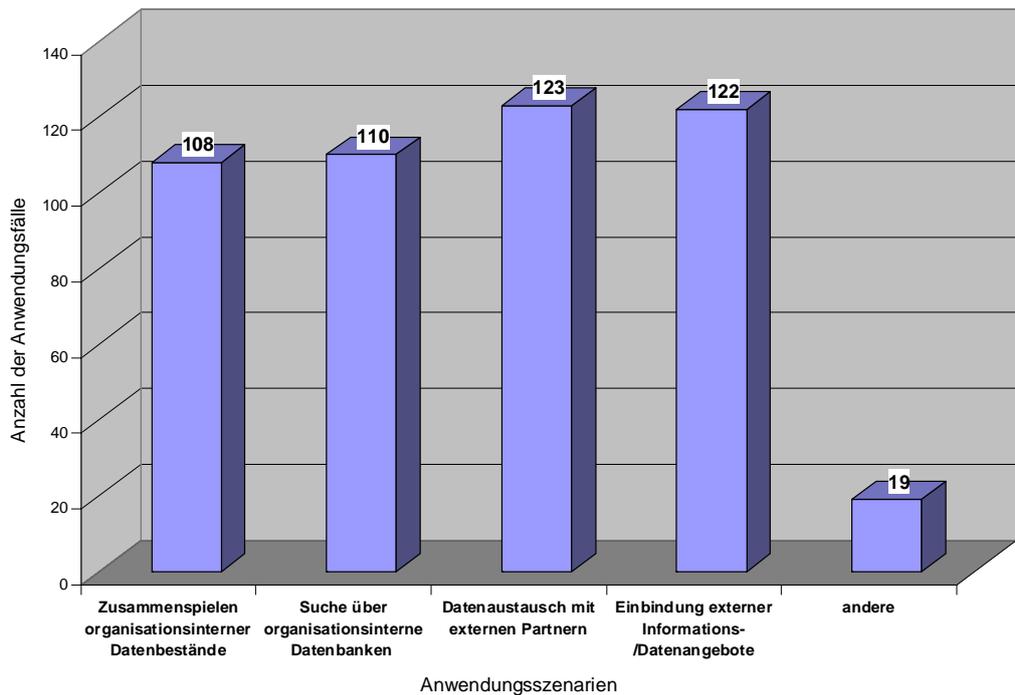


Abbildung 3-21 Bedarf an Zusammenführung von Metadaten

Von den Teilnehmern, die angaben, dass in ihrer Institution Bedarf an der Zusammenführung von Metadaten vorliegt, sind 10% zum Zeitpunkt der Umfrage nicht Nutzer von Metadaten. 87% der Teilnehmer, die Bedarf an der Zusammenführung in bestandsübergreifende Systeme angeben, sind Nutzer von Metadaten. 3% derjenigen, die Bedarf angaben, machten keine Angaben über die Verwendung von Metadaten. Bei Anwendungsszenarien bei der Integration von Metadaten werden insgesamt 482 potentielle Anwendungsfälle, die sich auf 5 Anwendungsszenarien verteilen, von 172 Teilnehmern der Umfrage genannt. Bei den Anwendungsfällen handelt es sich um potentielle Fälle, da im Rahmen der Umfrage nach dem Bedarf an Maßnahmen der Zusammenführung von Metadaten gefragt wurde, und nicht nur nach tatsächlich ausgeführten Maßnahmen. Bei der Auswertung der Mehrfachantworten wurde jede gegebene Antwort gezählt, so dass bei diesen Fragen zwischen 1 und 5 Antworten pro Teilnehmer vorliegen können. Die prozentualen Angaben über die Verteilung der Anwendungen auf die einzelnen Anwendungsszenarien wurde in Abhängigkeit aller von den Teilnehmern angegebenen (potentiellen) Anwendungsfällen berechnet (482), so dass die Anteile der Anwendungsszenarien in der Gesamtheit 100% ergeben. [104]

## Feldstudie

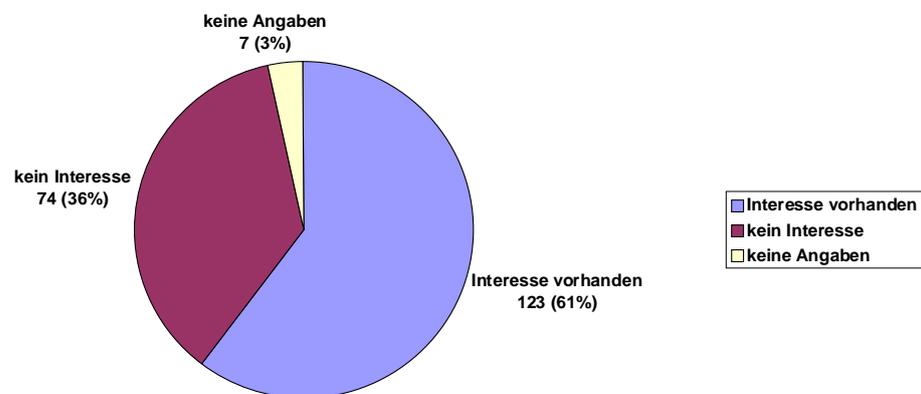


**Abbildung 3-22 Anwendungsszenarien der Zusammenführung von Metadaten**

Die Verteilung auf die unterschiedlichen Anwendungsszenarien ergibt ein relativ gleichmäßiges Bild. Insgesamt 51% der potentiellen Anwendungen, die hier angeführt wurden, beziehen sich auf den Austausch oder die Einbindung von Daten mit externen Partnern. 45 % der angegebenen Szenarien beziehen sich auf organisationsinterne Anwendungen wie das Zusammenspielen und die gemeinsame Suche über verschiedene organisationsinterner Datenbestände. 4% der Anwendungsszenarien lassen sich nicht den vorgegebenen Kategorien zuordnen. Der Datenaustausch mit externen Partnern beträgt 25% des Bedarfs an Anwendungsszenarien, weitere 25% des Bedarfs an Anwendungsszenarien besteht in der Einbindung externer Informationsangebote oder Datenangebote in eigene Anwendungen. 23% des Bedarfs liegt bei der gemeinsamen Suche über verschiedene organisationsinterne Datenbanken und 22% betrifft das Zusammenspielen organisationsinterner Datenbestände. Der Unterstützungsbedarf bei der Zusammenführung von Metadaten, etwa um das Zusammenführen von strukturierten Informationen aus verschiedenen Datenbanksystemen zu erleichtern und zu optimieren, war der erste Teil der Frage nach dem Interesse an unterstützenden Maßnahmen und bezieht sich darauf, bei welchen Teilnehmern Interesse vorliegt und bei welchen nicht. [104]

## Feldstudie

Der zweite Teil der Frage zielt auf die Wünsche bezüglich Angeboten wie Informationsmaterial, Weiterbildungs- und Fortbildungsmaßnahmen und Beratungsangebote, die zur Unterstützung bei der Integration von Metadaten von Interesse sind. Es standen 6 Angebote zur Auswahl, zudem konnten eigene Angaben zu gewünschten Angeboten gemacht werden. Mehrfachantworten waren möglich. Die meisten Teilnehmer gaben an, Interesse an Angeboten zur Unterstützung der Zusammenführung von Metadaten aus verschiedenen Datenbanksystemen zu haben. Die folgende Abbildung zeigt, dass bei 61% der Befragten Interesse vorhanden ist, 36% der Befragten hat überhaupt kein Interesse daran und 3% macht keine Angaben.

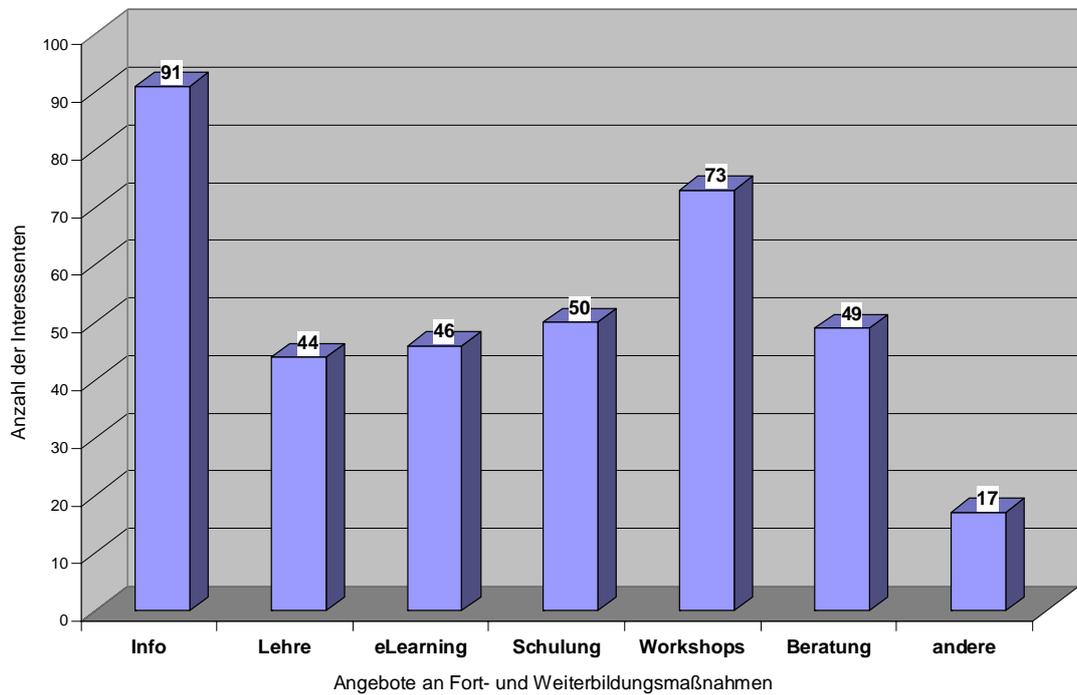


**Abbildung 3-23 Interesse an Angeboten**

Interesse an Angeboten zur Unterstützung bekunden jedoch nicht nur Metadatenanwender. 48% der Teilnehmer, die zum Zeitpunkt der Umfrage angaben, keine Metadaten zur verwenden, bejahen das Interesse an Angeboten.

Auf die Frage nach dem Interesse an Unterstützung bei der Zusammenführung von Metadaten mittels verschiedener Informations- und Lehrangebote, liegen insgesamt 370 Antworten vor. Bei der Auswertung der Mehrfachantworten wurde jede gegebene Antwort gezählt, so dass bei diesen Fragen zwischen 1 und 7 Antworten pro Teilnehmer vorliegen können. Die prozentualen Angaben wurden in Abhängigkeit der Gesamtheit, der von allen Teilnehmern abgegebenen Antworten berechnet, so dass die Anteile der Angebote in der Gesamtheit 100% ergeben.

## Feldstudie



**Abbildung 3-24 Interesse an Diensten**

91 Teilnehmer wünschen sich Informationsmaterialien zum Thema „Zusammenführung von Metadaten“, dies ergibt einen Anteil von 25% der Gesamtheit aller Angaben über gewünschte Angebote (im Folgenden als Gesamtanteil bezeichnet). 73 Teilnehmer haben Interesse an Workshops, die 20% des Gesamtanteils ausmachen. Das Interesse für die restlichen Angebote verteilt sich relativ gleichmäßig: 50 Teilnehmer (13% des Gesamtanteils) wünschen Schulungen, 49 (13% des Gesamtanteils) Beratungsangebote, 46 (12% des Gesamtanteils) E-Learning und 44 Lehrangebote (12% des Gesamtanteils). 17 Teilnehmer (5% des Gesamtanteils) gaben Interesse an weiteren Angeboten an. Hier wurde vorwiegend Interesse an Softwareprogrammen genannt, zum einen zur Unterstützung bei der Konvertierung von Metadatenformaten, zum anderen zur automatischen Generierung und Extraktion von Metadaten. [104]

## Online Analyse

### 4 Online Analyse

#### 4.1 Ziel

Die Online Analyse hat das Ziel zu beschreiben, welche Metadaten-Standards in OAI-PMH Repositories im Internet verwendet werden.

Die Herausforderung besteht darin, ein Java Programm zu implementieren, welches über das WWW automatisiert Abfragen sendet und die Antworten erfasst. Die Abfragen haben eine definierte URL, welche durch ein anderes Java Programm zuvor, mit Hilfe eines öffentlich gegebenen XML Files, generiert werden. Das Ergebnis ist eine Liste der verwendeten Metadatenstandards und dazugehörigen Repositories, welche beschrieben und in Diagrammen dargestellt wird.

#### 4.2 Implementierung

Das Programm wurde mit Hilfe der Entwicklungsumgebung "Eclipse Europe" erstellt. Es ist derzeit ein sehr verbreitetes IDE (Integrated Development Environment) zur Entwicklung von Java Programmen. Das Projekt gliedert sich in mehrere Programme auf, die insgesamt betrachtet den gewünschten Output liefern. Das erste File ist die "ListFriends.xml" Datei. In dieser Datei befinden sich alle Repositories, die eine Liste von OAI-OMH Providern darstellt. Die Liste ist unter <http://www.openarchives.org/Register/BrowseSites> zu erreichen. Bei der Implementierung ist mit einer Routine begonnen worden, die sequentiell alle Repositories durchgeht und die Formate ausliest. Diese Liste wird in XML lokal gespeichert. Die XML Liste wurde in MS Access importiert um dort komplexere Abfragen zu tätigen.

Das Protokoll ist genau beschrieben und hat Möglichkeiten, die zur Verfügung gestellten Daten mittels http anzusteuern und zu übertragen. Die Java-Implementierung des Protokolls: ist unter <https://www.oclc.org/research/software/oai/harvester2.htm> zu finden.

Das Programm, welches das File einliest, das die Liste der Repositories enthält, teilt sich in eine eigene definierte Klasse und dem Hauptprogramm auf. Im *Appendix A* ist das Programm im Detail einzusehen. Die o.a. Klasse stellt einen Container und Methoden für die eingelesenen XML Repositories (File) dar. Wesentliche Stellen im Programm sind die Methode `parseDocument()`, welches das XML File einliest. Die Basis URL wird der neuen Instanz der Klasse `XMLMetadataFormatRead.java` als Input gegeben. Die Instanz der neuen Klasse wird mit der Variable `mdf` implementiert. Die Instanz `mdf` verwendet wiederum eine definierte Klasse `MetadataFormats.java` um die Daten entgegenzunehmen. Wie bereits erwähnt, erledigt die Klasse

## Online Analyse

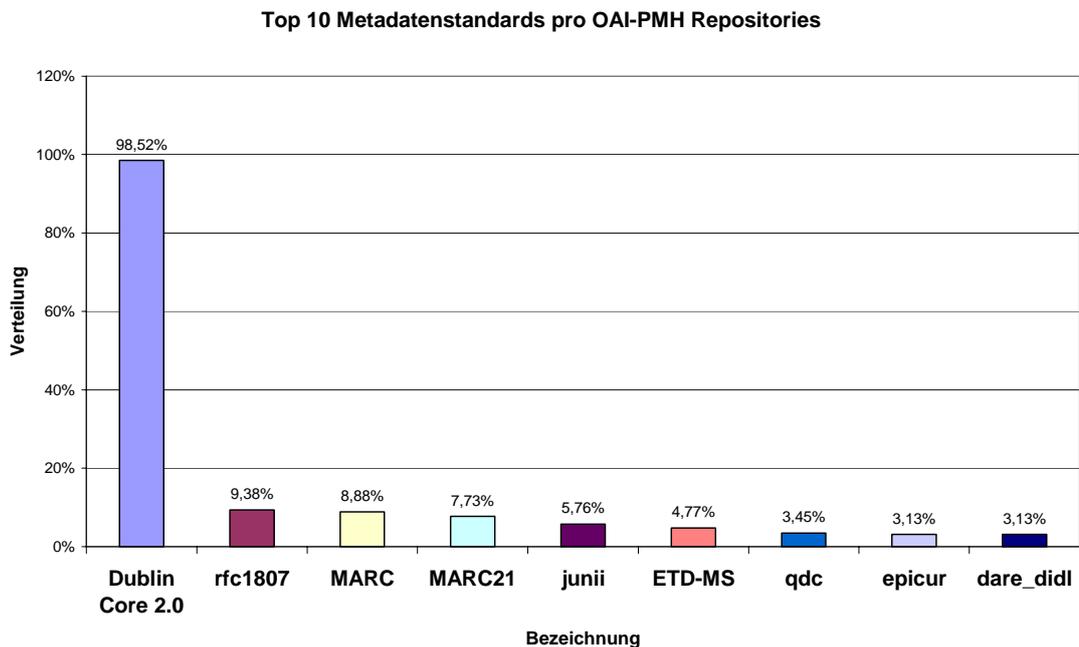
XMLMetadataFormatRead.java das Parsen der per HTTP Stream eingelesenen Metadaten Formate von den Repositories. Der wichtigste Programmpunkt ist wohl die Zeile, wo die übergebene URL geparkt wird. (sp.parse(ss, this);) Das Ergebnis des Programms wird am Ende zum Hauptprogramm durchgereicht und in der eigenen File Klasse mymetaformatfile.java gespeichert. Es wird die Methode mehr\_Inhalt(String) verwendet, um die ermittelten Metadatenformate auf File Ebene zu speichern. Das erzeugte File ist wiederum eine XML Datei, die in mit Hilfe von Access und Excel ausgewertet wird.

### 4.3 Auswertung der Online Analyse

#### 4.3.1.1 TOP 10 Metadatenstandards pro OAI Repository

Das folgende Diagramm zeigt die TOP 10 Metadatenstandards pro OAI Repository.

Die Berechnung erfolgte so, dass jeder eindeutig aufgetretene Metadatenstandard durch die Anzahl aller verschiedener Repositories dividiert wurde.



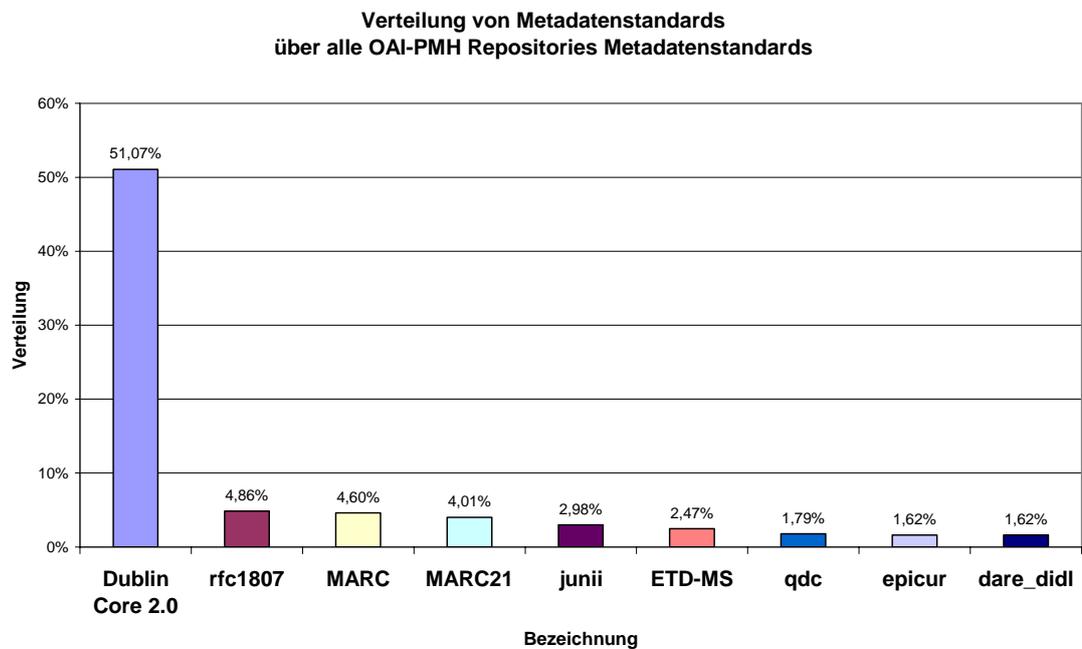
**Abbildung 4-1 Top 10 Metadatenstandards pro OAI Repository**

## Online Analyse

### 4.3.1.2 Metadatenstandards in OAI Repository.

Das folgende Diagramm zeigt die Verteilung von Metadatenstandards in allen OAI Repository.

Die Berechnung erfolgte so, dass jeder eindeutig aufgetretene Metadatenstandard durch die Anzahl aller eindeutigen Kombinationen von Repository und dazugehörigen Metadatenstandards dividiert wurde.

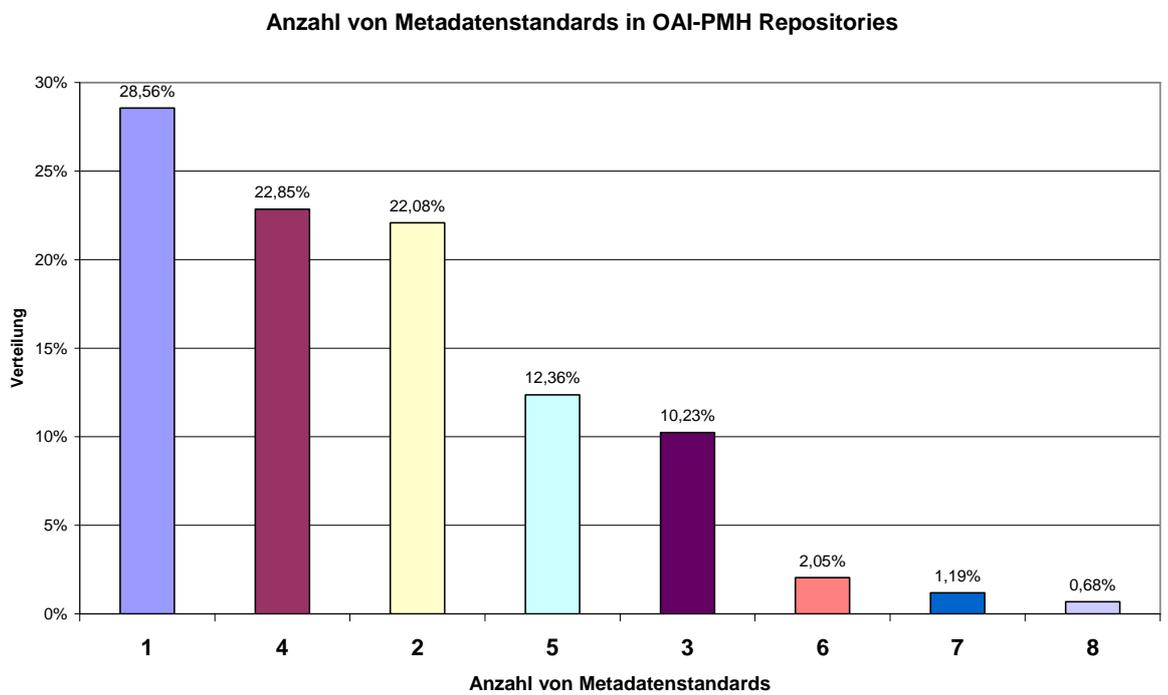


**Abbildung 4-2 Verteilung von Metadatenstandards über alle OAI-PMH Repositories**

## Online Analyse

### 4.3.1.3 Anzahl von Metadatenstandards in OAI Repository.

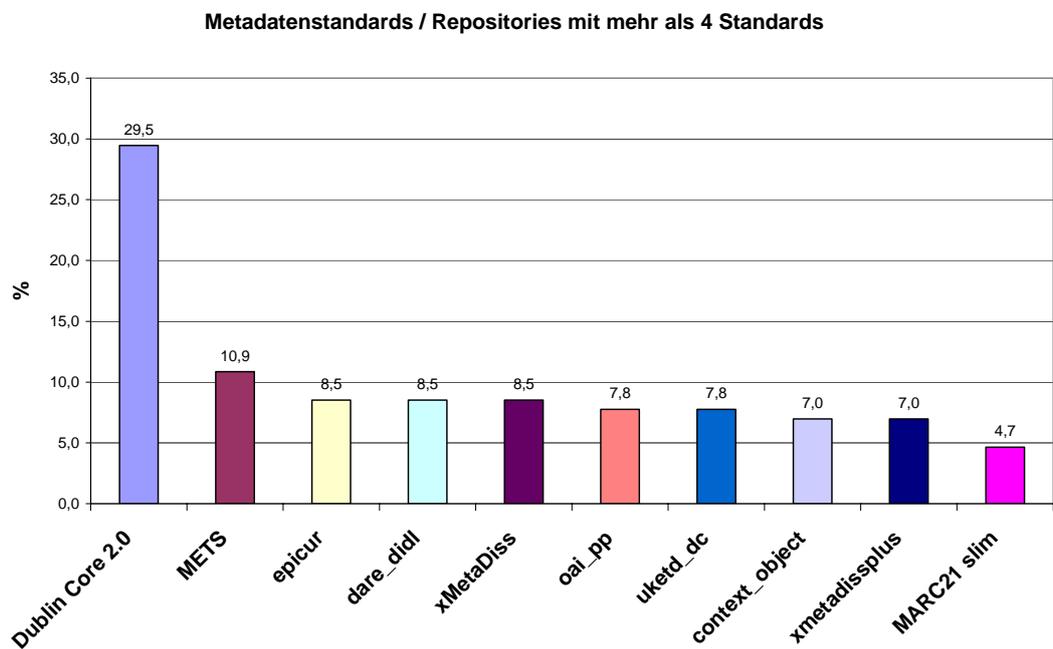
Das folgende Diagramm beschreibt wie viele Metadatenstandards pro OAI Repository vorkommen.



**Abbildung 4-3 Anzahl von Metadatenstandards in OAI-PMH Repositories**

## Online Analyse

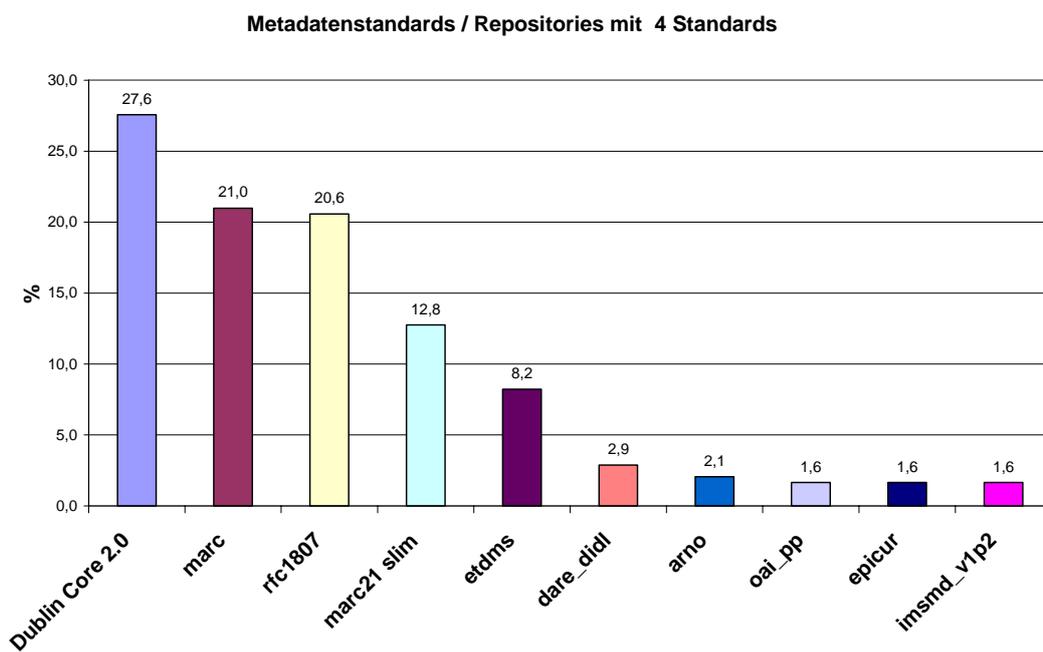
Das folgende Diagramm stellt dar welche Metadatenstandards in OAI Repository vorkommen, die mehr als 4 Standards benutzen.



**Abbildung 4-4 Metadatenstandards pro OAI-PMH Repository mit mehr als 4 Standards**

## Online Analyse

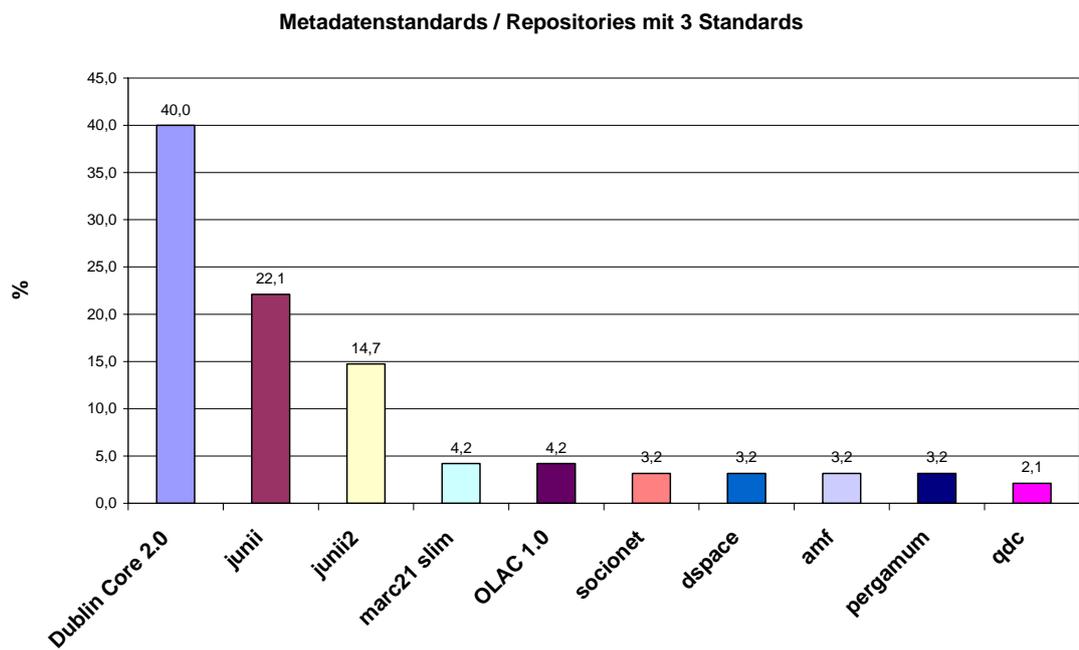
Das folgende Diagramm stellt dar welche Metadatenstandards in OAI Repository vorkommen, die 4 Standards benutzen.



**Abbildung 4-5 Metadatenstandards pro OAI-PMH Repository mit 4 Standards**

## Online Analyse

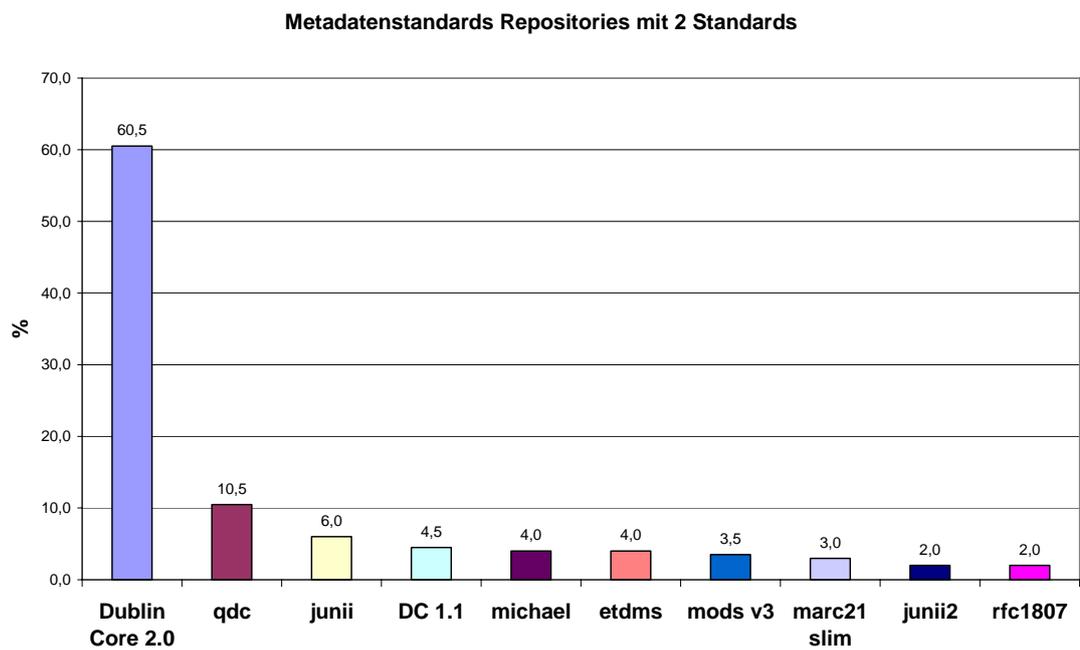
Das folgende Diagramm stellt dar welche Metadatenstandards in OAI Repository vorkommen, die genau 3 Standards benutzen.



**Abbildung 4-6 Metadatenstandards / Repositories mit 3 Standards**

## Online Analyse

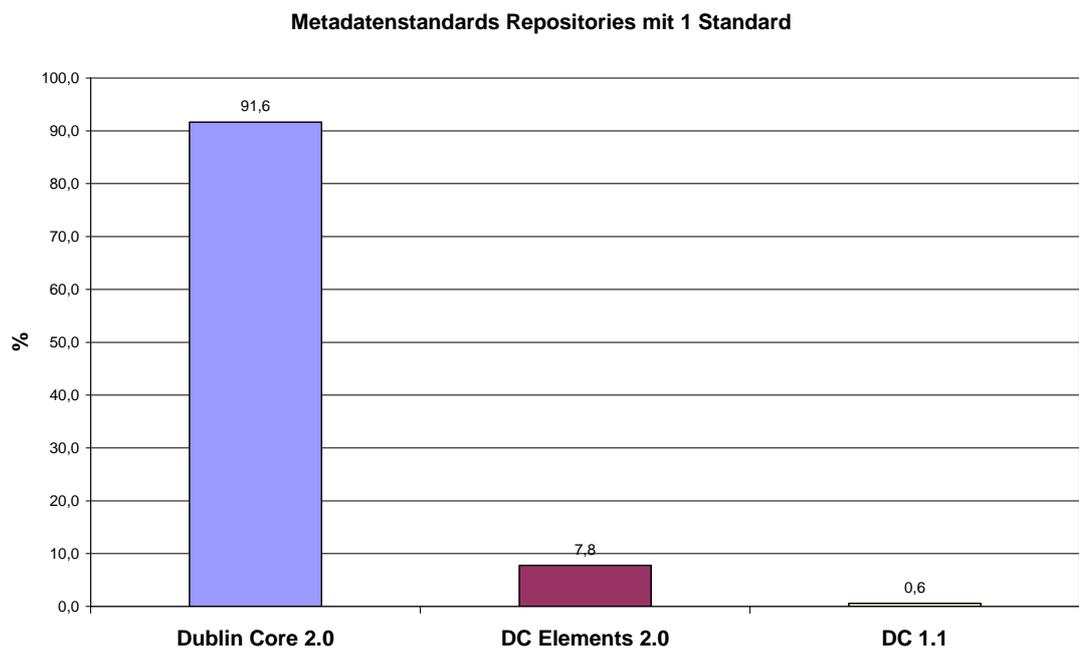
Das folgende Diagramm beschreibt welche Metadatenstandards in OAI Repository vorkommen, die genau 2 Standards benutzen.



**Abbildung 4-7 6 Metadatenstandards / Repositories mit 2 Standards**

## Online Analyse

Das folgende Diagramm stellt dar, welche Metadatenstandards in OAI Repository vorkommen, die genau einen Standard benutzen.



**Abbildung 4-8 6 Metadatenstandards / Repositories mit einem Standard**

## 5 Vergleich Online Analyse und Feldstudie

Das folgende Diagramm stellt dar welche Top10 Metadatenstandards in der Feldstudie benutzt werden. Anschließend wird es mit dem darauf folgendem Diagramm der Online Analyse verglichen.

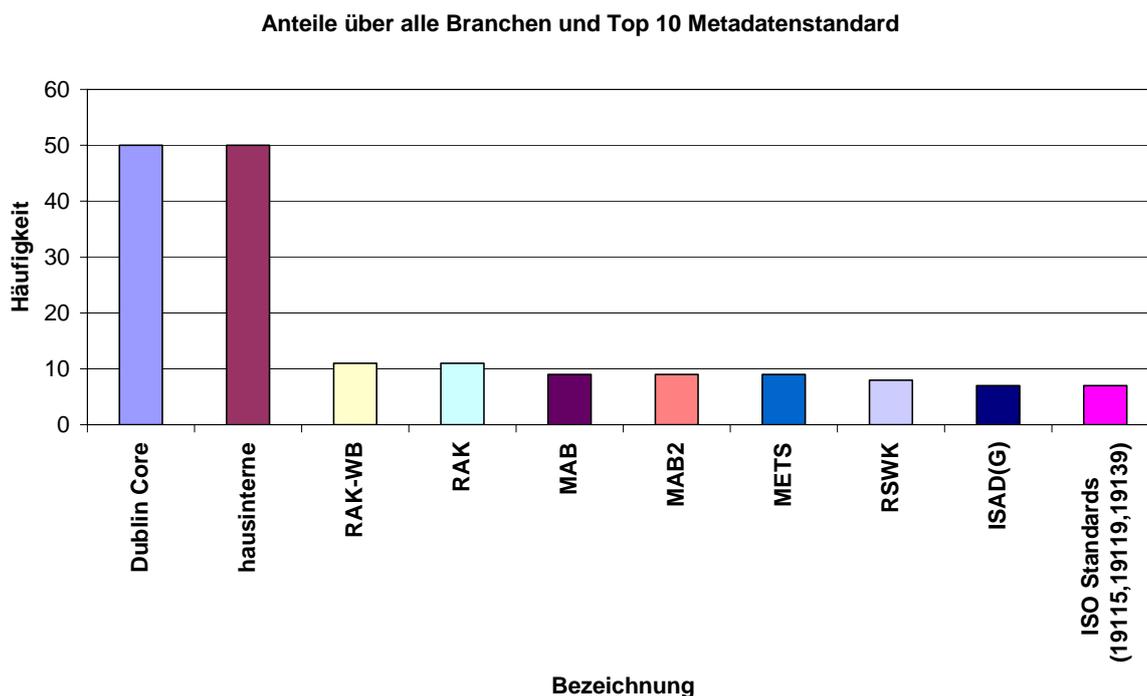


Abbildung 5-1 Anteile über alle Branchen und Top 10 Metadatenstandard

## Vergleich Online Analyse und Feldstudie

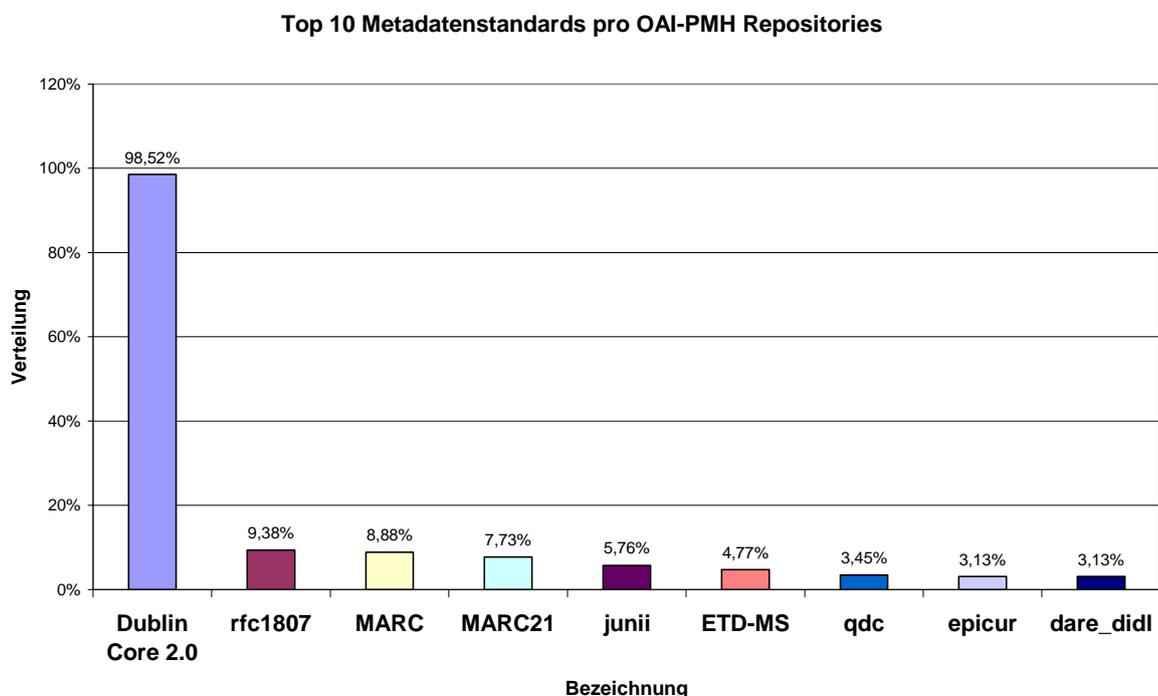
Eine Beschreibungen der Abkürzungen der am meisten verwendeten Metadatenstandards finden Sie in untenstehender Tabelle.

Dublin Core 2.0	Dublin Core 2.0
hausinterne	Proprietäre Metadatenstandards
RAK-WB	Regeln für alphabetische Katalogisierung im wissenschaftlichen Bereich
RAK	Regeln für alphabetische Katalogisierung
MAB	Maschinelles Austauschformat für Bibliotheken
MAB2	Maschinelles Austauschformat für Bibliotheken 2
METS	Metadata Encoding and Transmission Standard
RSWK	Regelwerk für Schlagwortkatalogisierung
ISAD(G)	(General) International Standard Archival Description, Second edition
ISO Standards	Metadatenstandards der Internationalen Standard Organisation

**Tabelle 5-1 2 Erklärung zu den Metadatenstandards im Diagramm**

## Vergleich Online Analyse und Feldstudie

Das unten folgende Diagramm stellt dar welche Top10 Metadatenstandards in OAI Repositories online verwendet werden.



**Abbildung 5-2 Top 10 Metadatenstandards pro OAI-PMH Repository**

Dublin Core 2.0	Dublin Core 2.0
Rfc1807	Request for Comments 1807
MARC / MARC 21	Machine Readable Cataloging Record
JUNII	Japanisches Metadatenformat
QDC	Qualified Dublin Core
EPICUR	Enhancement of Persistent Identifier Services- Comprehensive Method for Unequivocal Resource Identification
DIDL	Digital Item Declaration Language

**Tabelle 5-2 Metadaten-Namespace Abkürzungen**

## Vergleich Online Analyse und Feldstudie

Der Vergleich der Online Analyse und der Feldstudie Bezug nehmend auf das Diagramm der Top 10 verwendeten Metadatenstandards lässt erkennen, dass der Dublin Core Metadatenstandard in beiden Auswertungen der am meisten verwendete Standard ist. Desweiteren ist es der einzige Metadatenstandard, der in beiden Auswertungen im gleichen Format vorkommt. Die Anteile des Dublin Core Standards in der Online Analyse sind mit Abstand größer als der am zweit meisten benutzte Metadatenstandard. Im Unterschied dazu ist die Häufigkeit der verwendeten hausinternen Standards in der Feldstudie relativ groß. Viele Institutionen verwenden sozusagen noch selbst entwickelte Formate, die in Summe betrachtet fast genauso oft eingesetzt werden wie Dublin Core.

Die Feldstudie zeigt auch, dass die Schlagwortkatalogisierung mit Metadatenstandards einen doch recht beachtlichen Anteil ausmacht. Es sind Bestrebungen im Gange diese Standards in Verbindung mit bestehenden Regelwerken zu verschmelzen und zu vereinheitlichen. Welche Standards sich hier entwickeln bzw. ob andere Standards eingesetzt werden, wird sich vermutlich noch zeigen.

Metadatenstandards wie MARC, METS und MAB (in den verschiedenen Versionen) haben kumuliert einen nicht unwesentlichen Marktanteil.

# 6 Diskussion

## 6.1 Zukunft von Metadaten und -Standards

In den Anfängen der Schaffung von Metadatenstandards fokussierten sich die Kräfte auf beschreibende Elemente, die für die Entdeckung, die Identifikation und das Wiederauffinden zuständig waren. Andere Entwicklungen erschufen die administrativen Elemente, wie Rechteverwaltung und Erhaltung der Metadaten. Es gibt nur wenige Standards, die sich auf rein technische Metadaten konzentrieren. Allgemein schreitet die Entwicklung, dass Metadatenstandards in internationale Standards übergehen, voran. Das W3 Konsortium vereinigte Metadaten Aktivitäten mit dem semantischen Web. Das semantische Web erhöht den so genannten Interfunktionsaustausch (Interoperability) für Metadaten mehr als spezifische Schemata. [60].

Die Metadaten sollen die Verwaltung von Aufzeichnungen über räumliche und zeitliche Grenzen hinweg erlauben. Sie fallen bei jedem Bearbeitungsschritt an und protokollieren alle Transaktionen. Es sollen sich selbst beschreibende Aufzeichnungen gewonnen werden, die bei der Archivierung für die Verständlichkeit durch Dritte um übergeordnete Metadaten über die Organisation ergänzt werden könnten. Dabei sollen die zukünftigen Aufzeichnungen bereits an Hand der Funktionen und Kompetenzen bewertet sein, bevor sie entstehen. Denn Bewertung und Erschließung werden als Prozesse gesehen, die spätestens mit der Entstehung der Akten beginnen.

Wichtig sei sicherlich die Aufforderung, alle Metadaten, die um Akten und andere Aufzeichnungen herum entstehen, nicht zu vernachlässigen und sie mit ins Archiv zu übernehmen. Archivische Erschließung schafft zusätzlichen Mehrwert für Dritte über das Explizitmachen impliziter Informationen. Ein deutlich angestiegenes Interesse besteht an Kontexten und Entstehungsbedingungen von Informationen als Erklärungshintergründe. Die bisherigen Standards beschreiben nur die Ergebnisse von Aktionen, nicht aber die Aktionen selbst. Untersuchungen berichten über eine auf Ereignisse bezogene mit sogenannter „event awareness“ ausgestattete Beschreibungsmethode von Informationen. [84]

## 6.2 Kritik

### 6.2.1 Profit

Metadaten sind kostspielig, zeitaufwendig und kompliziert zu implementieren. Um die Metadateninformationen zu generieren, welche eben neue Informationen sind, muss Geld und Zeit aufgewendet werden. Profit kann aus Metadaten in Unternehmen, die

## **Diskussion**

Metadaten erzeugen und anwenden wollen, hinsichtlich der o.a. Argumente nur langfristig und indirekt erzeugt werden.

Die Kosten für Metadaten kann durch die Schaffung von geeigneten Standards und Werkzeugen gesenkt werden. Außerdem muss der Zugriff für Benutzer vereinfacht werden. Wenn man die richtigen, geeigneten und passenden Technologien verwendet, wird die Benutzererfahrung und die Qualität des Outputs verbessert. Im Metadatenumfeld ist es wichtig, dass spezifischen Anwendungen im Kontext verwendet werden, d.h. ohne einen gegebenen Anwendungsfall besteht kein Bedarf für Metadaten.

### **6.2.2 Objektivität**

Metadaten sind subjektive Informationen und sind abhängig vom Kontext, in dem sie verwendet oder erzeugt werden. Um die Objektivität zu bewahren, muss ein Administrator andere Metadaten haben, als ein Benutzer, der versucht Informationen zum Thema zu bekommen.

### **6.2.3 Grenzen setzen**

Metadaten können in Ihrem Umfang unendlich werden. Eine Information hat viele Aspekte und Metadaten zu einer Information können sich wiederum auf eine von vielen dieser Aspekte beziehen. Zusätzlich existieren auch Metadaten zu oder über Metadaten. Es muss eine Grenze gezogen werden, wo Metadaten anfangen und wo diese enden. Zusätzlich kann in einem bestimmten Anwendungsfall die Menge der benötigten Metadaten begrenzt werden.

### **6.2.4 Einsatz**

Es wird immer Daten geben, welche sich außerhalb der Hauptinformationsressource befinden und deswegen wird es fast immer nötig sein Metadaten zu verwenden. Sogar wenn Video/Audio/Bild-Daten effizient wiedergewonnen werden können, bleibt trotzdem der Bedarf, zusätzliche Informationen zu den Medien Objekten hinzuzufügen, die nicht Teil des Hauptobjektes sind.

Die Verwendung von Metadaten sei unbrauchbar, hieß es, jedoch sind die meisten Text Suchmaschinen bereits jetzt sehr effizient geworden und auch in Zukunft wird Video/Audio/Bild Wiedergewinnung sehr effizient werden. [70]

## Diskussion

### 6.3 Überlegungen

Die Nachteile, die selbst entwickelte Standards langfristig mit sich bringen könnten, wären die selbst zu erstellenden Crosswalks bzw. Transformationen, um mit anderen Formaten kompatibel zu werden. Diese Probleme der Interoperabilität sind für jeden selbst entwickelten Metadatenstandard vorhanden, hingegen gibt es für offizielle Metadatenstandards häufig Crosswalks. Der Aufwand einer Transformation von einem nicht-hausinternen Metadatenstandard dürfte geringer sein als bei proprietären Formaten.

Die Möglichkeiten den Dublin Core Metadatenstandard zu "qualifizieren", d.h. den individuellen Bedürfnissen anzupassen, ist gegeben. Eine Verwendung von hausinternen Standards sollte meiner Meinung nach vermieden und eine Erweiterung des Dublin Core bevorzugt werden. Die Weiterentwicklung von Metadatenstandards ist, meiner Meinung nach, ein sehr wichtiges Ziel um Informationen und Daten, die im Internet, in Bibliotheken oder in verschiedenen Applikationen entstehen, besser aufbereiten zu können. Wenn Informationen veredelt werden, bedeutet dies einen entstandenen Mehrwert.

Im Gegensatz dazu würde eine Vernachlässigung der Fülle von Informationen ein Datenchaos entstehen lassen, welches beinahe keinen Wert hätte. Die Interoperabilität für Daten und Informationen ist die Grundvoraussetzung, dass überhaupt eine Verbindung zwischen verschiedenen Daten, Inhalten und Benutzern entstehen kann, und dadurch möglicherweise Erkenntnisse wachsen, die zuvor vielleicht nicht denkbar wären. Eine Metadatengenerierung bei der Entstehung von Informationen könnte, meiner Meinung nach, auch ein wesentlicher Schritt sein, die Qualität der Metadaten zu verbessern. Mit Dublin Core als großen Faktor im Metadatenstandardbereich ist eine Basis gegeben auf die man bauen könnte.

### 7 Literatur und Webressourcen

- [1] Metadata Server Einführung in Metadaten, <http://www2.sub.uni-goettingen.de/intrometa.html>
- [2] Metadata Activity Statement, <http://www.w3.org/Metadata/Activity.html>
- [3] Understanding Metadata, NISO Press National Information Standards Organization, ISBN: 1-880124-62-9, 2004
- [4] Metadaten und Persistente Identifikatoren, <http://www8.informatik.uni-erlangen.de/IMMD8/Lectures/DIGLIB/05-metadaten.pdf>
- [5] Introduction to Metadata, Pathways to Digital Information, Online Edition 2.1, [http://www.getty.edu/research/conducting\\_research/standards/intrometadata/setting.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.html)
- [6] Metadata.Net, Metadata tools and services, <http://www.metadata.net/tools.html>
- [7] Klas, Wolfgang, and A.Steh. Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media. McGraw-Hill series on data warehousing and data management. New York; McGraw-Hill, 1988
- [9] Third International Conference on Information Visualisation (IV'99), page 446, Data and Metadata for Finding and Reminding, James Gheel, University of Ulster, Terry Anderson, University of Ulster, 1999
- [10] Paper given by Dr. Warwick Cathro, Assistant Director-General, Services to Libraries Division at the Standards Australia Seminar, "Matching Discovery and Recovery", <http://www.nla.gov.au/nla/staffpaper/cathro3.html>
- [11] Einbindung von Metadaten, [http://e-campus.uibk.ac.at/planet-et-fix/M8/8.3.1\\_Metadaten/40md\\_einbindung.html](http://e-campus.uibk.ac.at/planet-et-fix/M8/8.3.1_Metadaten/40md_einbindung.html)
- [12] Keyword Extraction from the Web for FOAF Metadata, [http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/keyword\\_extraction\\_from\\_the\\_web/](http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/keyword_extraction_from_the_web/)
- [13] Dublin Core Metadata Initiative, <http://dublincore.org/documents/usageguide/#introduction>
- [14] Exchange of Metadata in XML Format According to CWM, [http://help.sap.com/saphelp\\_bw30b/helpdata/en/60/2edd3b8f1b127de1000000a114084/content.htm](http://help.sap.com/saphelp_bw30b/helpdata/en/60/2edd3b8f1b127de1000000a114084/content.htm)

## Literatur und Webressourcen

- [15] Introduction to RDF Metadata, Standardization Efforts at W3C,  
<http://www.w3.org/TR/NOTE-rdf-simple-intro-971113.html>
- [16] Metadata, Mapping between metadata formats, Michael Day,  
<http://www.ukoln.ac.uk/metadata/interoperability/>, 2002
- [17] Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- [18] Dublin Core Metadata Registry, <http://dublincore.org./dcregistry/>
- [19] ISO/IEC JTC1 SC32 WG2 Development/Maintenance, <http://metadata-standards.org/11179/>
- [20] Dublin Core Metadata Initiative's (DCMI) ,  
<http://dublincore.org/dcregistry/index.html>
- [21] Z39.50 A Primer on the Protocol,  
[http://www.niso.org/standards/resources/Z3950\\_primer.pdf](http://www.niso.org/standards/resources/Z3950_primer.pdf) , page 12 , 2002
- [22] SAP, Metadata Repository,  
[http://help.sap.com/saphelp\\_nw04/helpdata/en/4e/ea683c25e9096de10000000a114084/content.htm](http://help.sap.com/saphelp_nw04/helpdata/en/4e/ea683c25e9096de10000000a114084/content.htm)
- [23] DIGITAL LIBRARIES: Metadata Resources GENERAL RESOURCES AND INDICES, <http://www.ifla.org/II/metadata.htm>
- [24] Metadatenmodelle im Document Engineering, Vorstellung der Modelle,  
<http://www.unibw.de/inf2/getFILE?fid=1287143>, 2004
- [25] NISO Neuigkeiten, <http://www.niso.org/news/newslines/NISONewsline-May2007.htm>
- [26] Open Archives, Registered Data Providers,  
<http://www.openarchives.org/Register/BrowseSites>
- [27] The Cedars Project, <http://www.leeds.ac.uk/cedars/metadata.html>, 2001
- [28] DCMI Metadata Terms, <http://dublincore.org/documents/dcmi-terms/#H1>, 2006
- [29] Metadata Glossary,  
<http://dublincore.org/groups/tools/dctools2006/2MetadataGlossary.pdf>
- [30] Dublin Core Metadata Initiative, <http://dublincore.org/documents/usageguide/>
- [32] Encoding Guidelines, <http://dublincore.org/resources/expressions/>

## Literatur und Webressourcen

- [33] Dublin Core Metadata Element Set, Version 1.1,  
<http://dublincore.org/documents/dces/> , 2007
- [34] DCMI Metadata Terms, <http://dublincore.org/documents/dcmi-terms/> , 2006
- [35] Using Dublin Core - The Elements,  
<http://dublincore.org/documents/usageguide/elements.shtml#identifier> , 2006
- [36] Using Dublin Core - Dublin Core Qualifiers,  
<http://dublincore.org/documents/usageguide/qualifiers.shtml> , 2007
- [37] Network Development and MARC Standards Office, Library of Congress,  
<http://www.loc.gov/marc/faq.html> , 2006
- [38] MARC 21 Reference Materials, <http://www.loc.gov/marc/umb/um11to12.html> ,  
2003
- [39] Data Dictionary - Technical Metadata for Digital Still Images,  
<http://www.niso.org/standards/resources/Z39-87-2006.pdf> , page 47
- [40] EXIF.org, Specifications , <http://www.exif.org/specifications.html> , 2003
- [41] Draft Standard for Learning Object Metadata,  
[http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf) , page 4-10  
, 2002
- [41] Metdata Made Simpler, Using Metadata,  
[http://www.niso.org/news/Metadata\\_simpler.pdf](http://www.niso.org/news/Metadata_simpler.pdf) , 2001
- [42] Embedding Accessibility Role and State Metadata in HTML Documents,  
Examples of the technique,  
<http://www.w3.org/WAI/PF/adaptable/HTML4/embedding-20061212.html#examples> , 2006
- [43] Metadata Activity Statement , Simple explanation of concepts ,  
<http://www.w3.org/Metadata/Activity> , 2002
- [44] SAP Class JCO.MetaData ,  
[http://help.sap.com/javadocs/NW04S/current/jc/com/sap/mw/jco/JCO.MetaData.html#method\\_summary](http://help.sap.com/javadocs/NW04S/current/jc/com/sap/mw/jco/JCO.MetaData.html#method_summary)
- [49] Expressing Dublin Core metadata using the Resource Description Framework (RDF) , <http://dublincore.org/documents/2006/05/29/dc-rdf/>
- [50] Extensible Metadata Platform (XMP) ,

## Literatur und Webressourcen

- <http://www.adobe.com/products/xmp/standards.html> , 2007
- [51] Extensible Metadata Platform (XMP) ,  
<http://www.adobe.com/products/xmp/overview.html> , 2007
- [52] Adobe® XMP Toolkit,  
[http://www.adobe.com/devnet/xmp/pdfs/xmp\\_sdk\\_overview.pdf](http://www.adobe.com/devnet/xmp/pdfs/xmp_sdk_overview.pdf) , 2007
- [53] XMP Specification,  
[http://www.adobe.com/devnet/xmp/pdfs/xmp\\_specification.pdf](http://www.adobe.com/devnet/xmp/pdfs/xmp_specification.pdf) , page 37 - 88 ,  
2005
- [54] Adobe XMP Develop Center, Eclipse Project,  
<http://www.adobe.com/devnet/xmp/sdk/eula.html> , 2007
- [55] Suche auf ISO.org nach MPG-7 , <http://www.iso.org/iso/search.htm?qt=mpeg-7&searchSubmit=Search&sort=rel&type=simple&published=true>
- [56] ISO/IEC TR 15938-8:2002/Amd 2:2006, Extraction and use of MPEG-7 perceptual 3D shape descriptor , 2006
- [57] ISO/IEC 15938-9:2005 , Information technology -- Multimedia content description interface -- Part 9: Profiles and levels , 2005
- [58] MPEG-7 Overview, <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm> , 2004
- [59] MPEG-21 Overview v.5, <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm> , 2002
- [69] Semantic Web - an Evolution for the Future, <http://www.w3.org/2002/Talks/WS-HongKong-IH/OverviewPrint.html>
- [70] Revisions to DCMI Metadata Terms, <http://www.dublincore.org/usage/public-comment/2007/07/dcterms-changes/>
- [71] NISO Metadata for Images in XML Schema, <http://www.loc.gov/standards/mix/> ,  
2007
- [72] XHTML™ 1.0 The Extensible HyperText Markup Language (Second Edition),  
<http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd> , 2002
- [73] Lagoze, Carl, Lynch, Clifford A. Daniel, Ron, Jr., The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata, Cornell University, 1996

## Literatur und Webressourcen

- [74] The Dublin Core: A Simple Content Description Model for Electronic Resources by Stuart Weibel, <http://www.asis.org/Bulletin/Oct-97/weibel.htm>
- [75] O. Lassila and R. R. Swick. "Resource Description Framework (RDF) model and syntax specification". W3C Working Draft WD-rdf-syntax-19981008., <http://www.w3.org/TR/WD-rdf-syntax>
- [76] Stefan Fangmeier, Einsatz und Funktion ausgewählter Normdateien am Beispiel des Bibliotheksverbundes Niedersachsen /Sachsen-Anhalt / Thüringen, Fachhochschule Köln Fachbereich Bibliotheks- und Informationswesen, Köln, 1995
- [78] Internationale Grundsätze für die archivische Verzeichnung, übers. und bearb. von Rainer Brüning und Werner Heegewaldt, Marburg 1994
- [77] Gerald Maier, Das BAM-Portal – Gemeinsames Internetportal für Bibliotheken, Archive und Museen (<<http://www.bam-portal.de>>), Neue Wege zu digitalen Kulturgut-Informationen
- [79] Formate und Schnittstellen, <http://www.d-nb.de/standardisierung/formate/mab.htm>
- [80] METS: Überblick und Anleitung, [http://www.loc.gov/standards/mets/METSOverview.v2\\_de.html](http://www.loc.gov/standards/mets/METSOverview.v2_de.html)
- [81] Friedrich Geißelmann, Zur dritten Auflage der RSWK, BIBLIOTHEKSDIENST 33. Jg. (1999)
- [82] Holger Flachmann, Zur Effizienz bibliothekarischer Inhaltserschließung: Allgemeine Probleme und die Regeln für den Schlagwortkatalog (RSWK), BIBLIOTHEKSDIENST 38. Jg. (2004)1
- [83] Hans-Peter Wessel , RSWK und RAK – die Geschichte einer unendlichen Annäherung?, BIBLIOTHEKSDIENST 39. Jg. (2005),
- [84] Angelika Menne-Haritz, Meta-Daten Konferenz in Wolfheze (NL), <http://www.archive.nrw.de/archivar/2000-04/A29.htm>
- [85] Martin Iordanidis, Anwendbarkeit von XML Schema für Daten und Metadaten im Bereich digitaler Bibliotheken, Magisterarbeit, 2003
- [86] Umfrage zur Metadatenutzung, <http://kim-forum.org/>
- [87] OCLC PICA, [http://de.wikipedia.org/wiki/OCLC\\_PICA](http://de.wikipedia.org/wiki/OCLC_PICA)

## Literatur und Webressourcen

- [88] ICA, International Council of Archives, <http://www.ica.org/en/node/30000>
- [89] Rainer Brüning, Werner Heegewaldt, Nils Brübach: ISAD(G) - Internationale Grundsätze für die archivische Verzeichnung, Zweite, überarbeitete Ausgabe, Veröffentlichungen der Archivschule Marburg Institut für Archivwissenschaft, (<http://www.ica.org/en/node/30175>)
- [90] A Format for Bibliographic Records,  
<http://www.ietf.org/rfc/rfc1807.txt?number=1807>
- [91] Lasher, R. and Cohen, D. 1995 A Format for Bibliographic Records. RFC. RFC Editor.
- [92] ETD-ms: an Interoperability Metadata Standard for Electronic Theses and Dissertations-- version 1.00,  
[http://www.saber.ula.ve/documentos/tesis/doc/NDLTD\\_ETD-ms.htm](http://www.saber.ula.ve/documentos/tesis/doc/NDLTD_ETD-ms.htm)
- [93] ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations -- version 1.00, revision 2,  
<http://www.ndltd.org/standards/metadata/current.html>
- [94] Ana Pavani and Joseph M.Moxley, The Guide for Electronic Theses and Dissertations, Why ETD's?, page 43, 2001
- [95] Edward Fox, What are ETDs?, <http://etdguide.org/content/1.1.htm>
- [96] Thomas Krämer, Interoperabilität von Metadatenstandards und Dokumentenretrieval, 2004
- [97] Kathrin Schroeder, XML schema on the basis of the XEPICUR transfer syntax,  
<http://www.persistent-identifier.de/xepicur/version1.0/xepicur.xsd>
- [98] Schröder, Kathrin, Rahmenbedingungen zur persistenten Addressierung digitaler Objekte, 7. InetBib-Tagung vom 12. bis 14. November 2003 in Frankfurt am Main, 2004
- [99] Dr. Peter Reimer, Jochen Schirrwagen, Dr. Wolfram, Horstmann, Technische Dokumentation des DiPPPublikationssystems, 2006
- [100] Schema for Derived DIDL Types ,  
[http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-21\\_schema\\_files/did/didl.xsd](http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-21_schema_files/did/didl.xsd)
- [101] Jeroen Bekaert, Patrick Hochstenbach, Herbert Van de Sompel, Using MPEG-21 DIDL to Represent Complex Digital Objects in the Los Alamos National

## Appendix A

Laboratory Digital Library, D-Lib Magazine , November 2003

- [102] Schroeder, Kathrin<sup>1</sup>, Persistent Identification for the Permanent Referencing of Digital Resources - The Activities of the EPICUR Project Enhanced Uniform Resource Name URN Management at Die Deutsche Bibliothek, Haworth Press, 2006  
(<http://www.ingentaconnect.com/content/haworth/sl/2006/00000049/00000003/art00005>)
- [103] METS Metadata Encoding & Transmission Standard,  
<http://www.loc.gov/standards/mets/>
- [104] Mirjam Keßler, Umfrage zur Nutzung von Metadaten, KIM: Kompetenzzentrum für interoperable Metadaten, 2007

## Appendix A

**Das Java Programm**, welches das File einliest, das die Liste der Repositories enthält, teilt sich in die Klasse ListFriend.java und dem Hauptprogramm, XMLRepositoryURLsRead.java, auf.

Es folgt der Programm-Code mit kurzer Beschreibung am Ende des jeweiligen Programmes:

```
public class ListFriend {
    private String baseURL;
    private String id;
    public ListFriend(){
    }
    public ListFriend(String baseURL, String id) {
        this.baseURL = baseURL;
        this.id = id;
    }
    public ListFriend(String baseURL) {
        this.baseURL = baseURL;
    }
}
```

## Appendix A

```
public String getbaseURL() {
    return baseURL;
}
public void setid(String id) {
    this.id = id;
}
public String getid() {
    return id;
}

public void setbaseURL(String baseURL) {
    this.baseURL = baseURL;
}

public String toString() {
    StringBuffer sb = new StringBuffer();
    sb.append(getbaseURL());
    return sb.toString();
}
}
```

Die o.a. Klasse stellt einen Container und Methoden für die eingelesenen XML Repositories (File) dar.

### Das Hauptprogramm

```
import java.io.IOException;
import java.io.InputStreamReader;
import java.net.URL;

import java.util.ArrayList;
import java.util.Iterator;
import java.util.List;

import javax.xml.parsers.ParserConfigurationException;
import javax.xml.parsers.SAXParser;
```

## Appendix A

```
import javax.xml.parsers.SAXParserFactory;
import javax.xml.transform.TransformerException;

import org.xml.sax.Attributes;
import org.xml.sax.SAXException;
import org.xml.sax.helpers.DefaultHandler;

import ORG.oclc.oai.harvester2.app.RawWrite;
import ORG.oclc.oai.harvester2.verb.ListMetadataFormats;
import ORG.oclc.oai.harvester2.verb.ListMetadataFormats_pub;

public class XMLRepositoryURLsRead extends DefaultHandler {

    List myEmpIs;

    private String tempVal;

    private ListFriend tempEmp;

    public XMLRepositoryURLsRead() {
        myEmpIs = new ArrayList();
    }

    public void runExample() {
        parseDocument();
        try {
            printData();
        } catch (IOException e) {
            e.printStackTrace();
        } catch (ParserConfigurationException e) {
            e.printStackTrace();
        } catch (SAXException e) {
            e.printStackTrace();
        } catch (TransformerException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }
}
```

## Appendix A

```
    }
}

private void parseDocument() {
    SAXParserFactory spf = SAXParserFactory.newInstance();
    try {
        SAXParser sp = spf.newSAXParser();
        sp.parse("ListFriends.xml", this);
    } catch (SAXException se) {
        se.printStackTrace();
    } catch (ParserConfigurationException pce) {
        pce.printStackTrace();
    } catch (IOException ie) {
        ie.printStackTrace();
    }
}

private void printData() throws IOException,
ParserConfigurationException,
    SAXException, TransformerException {
    String h = null;
    URL myURL;
    int nextchar;
    int cc = 0;
    mymetaformatfile mf = new mymetaformatfile();
    ListMetadataFormats_pub ml = new
ListMetadataFormats_pub();
    XMLMetadataFormatRead mdf;

    mf.fileinit();

    mf.mehr_Inhalt("<?xml version=\"1.0\" encoding=\"UTF-
8\"?>");
    mf.mehr_Inhalt("<MetadataFormatList>");
        Iterator it = myEmpls.iterator();
    while ((it.hasNext())) {
```

## Appendix A

```
cc = cc + 1;
System.out.print(cc);
nextchar = 0;
h = "";
h = ml.getRequestURL((it.next().toString()), null);
mf.mehr_Inhalt("\n" + "<Repository> <url>" + h +
"</url>");

if (cc <= 800) {
    try {
        myURL = new URL(h);
        mdf = new XMLMetadataFormatRead();

mf.mehr_Inhalt(mdf.runExample(h));

        System.out.println(h);
        * InputStreamReader in = new
        *
        * while ((nextchar = in.read()) != -1)
        * (char)nextchar; mf.doAppend();
        * System.out.print((char)nextchar); }
        * System.out.println("CLOSE-----
        *
        */
    } catch (Exception e) {

        e.printStackTrace();

        mf.mehr_Inhalt("<MFStandard>Error</MFStandard><mfscout></mfscou
        nt></Repository>");

        System.out.println("ERROR!!!-----
        -----");
    }
}
```

## Appendix A

```
        }
        else
mf.mehr_Inhalt("<MFStandard></MFStandard><mfscout></mfscout></Repository>");
        }
        mf.mehr_Inhalt("</MetadataFormatList>");
        mf.fileclose();
    }
    public void startElement(String uri, String localName, String
qName,
        Attributes attributes) throws SAXException {
        // reset
        tempVal = "";
        if (qName.equalsIgnoreCase("baseURL")) {
            // create a new instance of employee
            tempEmp = new ListFriend();
            tempEmp.setbaseURL(attributes.getValue("baseURL"));
//            tempEmp.setid(attributes.getValue("id"));

        }
    }

    public void characters(char[] ch, int start, int length)
        throws SAXException {
        tempVal = new String(ch, start, length);
    }

    public void endElement(String uri, String localName, String
qName)
        throws SAXException {

        if (qName.equalsIgnoreCase("baseURL")) {
            // add it to the list
            myEmps.add(tempEmp);
            tempEmp.setbaseURL(tempVal);
//        } else if (qName.equalsIgnoreCase("id")) {
//            tempEmp.setid(tempVal);
        }
    }
}
```

## Appendix A

```
    }  
}  
  
public static void main(String[] args) {  
    XMLRepositoryURLsRead spe = new XMLRepositoryURLsRead();  
    spe.runExample();  
}  
}
```

Wesentliche Stellen im Programm sind die Methode `parseDocument()`, welches das XML File einliest. Die Basis URL wird der neuen Instanz der Klasse `XMLMetadataFormatRead.java` als Input gegeben. Die Instanz der neuen Klasse wird mit der Variable `mdf` implementiert. Die Instanz `mdf` verwendet wiederum eine definierte Klasse `MetadataFormats.java` um die Daten entgegenzunehmen.

```
public class MetadataFormats {  
  
    private String Prefix;  
  
    public MetadataFormats(){  
  
    }  
  
    public MetadataFormats(String Prefix) {  
        this.Prefix = Prefix;  
    }  
  
    public String getPrefix() {  
        return Prefix;  
    }  
}
```

## Appendix A

```
public void setPrefix(String Prefix) {
    this.Prefix = Prefix;
}

public String toString() {
    StringBuffer sb = new StringBuffer();
    sb.append(getPrefix());
    return sb.toString();
}
}
```

Wie oben erwähnt, erledigt die Klasse XMLMetadataFormatRead.java das Parsen der per HTTP Stream eingelesenen Metadaten Formate von den Repositories.

```
import java.io.IOException;
import java.io.InputStreamReader;
import java.net.URL;

import java.util.ArrayList;
import java.util.Iterator;
import java.util.LinkedList;
import java.util.List;

import javax.xml.parsers.ParserConfigurationException;
import javax.xml.parsers.SAXParser;
import javax.xml.parsers.SAXParserFactory;
import javax.xml.transform.TransformerException;

import org.xml.sax.Attributes;
import org.xml.sax.SAXException;
import org.xml.sax.helpers.DefaultHandler;

import ORG.oclc.oai.harvester2.app.RawWrite;
import ORG.oclc.oai.harvester2.verb.ListMetadataFormats;
import ORG.oclc.oai.harvester2.verb.ListMetadataFormats_pub;
```

## Appendix A

```
public class XMLMetadataFormatRead extends DefaultHandler{

    List myEmpIs;

    private String tempVal;
    public String ss = "";
    private String rmf = "";
    public String mya = "";

    private MetadataFormats tempEmp;

    public XMLMetadataFormatRead(){
        myEmpIs = new ArrayList();
    }

    public String runExample(String ss) {
        this.ss = ss;
        parseDocument();
        try {
            rmf = printData();
        } catch (IOException e) {
            e.printStackTrace();
        } catch (ParserConfigurationException e) {
            e.printStackTrace();
        } catch (SAXException e) {
            e.printStackTrace();
        } catch (TransformerException e) {
            e.printStackTrace();
        }
        return rmf;
    }

    private void parseDocument() {
```

## Appendix A

```
SAXParserFactory spf = SAXParserFactory.newInstance();
try {
    SAXParser sp = spf.newSAXParser();
    sp.parse(ss, this);
} catch (SAXException se) {
    se.printStackTrace();
} catch (ParserConfigurationException pce) {
    pce.printStackTrace();
} catch (IOException ie) {
    ie.printStackTrace();
}
}

private String printData() throws IOException,
ParserConfigurationException, SAXException, TransformerException{
    String md = "" , h ;
    int i = 0;
    String[] ay = new String[i];
    int nextchar;

    System.out.println("Anzahl gefundener Metastandards " +
myEmpls.size() + "'.");

    Iterator it = myEmpls.iterator();
    while(it.hasNext()) {
        h = "";
        h = it.next().toString();

        md = "<MFStandard>"+ h + "</MFStandard>" + "\n" + md
+ "\n" ;

    }

    md = ("<mfscount>" + myEmpls.size() + "</mfscount>" + md
+ "</Repository>");
```

## Appendix A

```
        return md;
    }

    //Event Handlers
    public void startElement(String uri, String localName, String
qName, Attributes attributes) throws SAXException {
        //reset
        tempVal = "";
        if(qName.equalsIgnoreCase("metadataNamespace")) {
            //create a new instance of employee
            tempEmp = new MetadataFormats();

            tempEmp.setPrefix(attributes.getValue("metadataNamespace"));
            //ListMetadataFormats

        }
    }

    public void characters(char[] ch, int start, int length) throws
SAXException {
        tempVal = new String(ch,start,length);
    }

    public void endElement(String uri, String localName, String
qName) throws SAXException {

        if(qName.equalsIgnoreCase("metadataNamespace")) {
            myEmps.add(tempEmp);
            tempEmp.setPrefix(tempVal);
        }
    }
}
```

## Appendix A

Der wichtigste Programmpunkt ist wohl die Zeile, wo die übergebene URL geparkt wird.

```
sp.parse(ss, this);
```

Das Ergebnis des Programms wird am Ende zum Hauptprogramm durchgereicht und in der eigenen File Klasse mymetaformatfile.java gespeichert. Es wird die Methode mehr\_Inhalt(String) verwendet, um die ermittelten Metadatenformate auf File Ebene zu speichern.

```
import java.io.BufferedWriter;
import java.io.FileWriter;
import java.io.IOException;

public class mymetaformatfile {

    public BufferedWriter myfout ;
    public String tw ;

    public void fileinit() {

        try {

            String fileName = "c:\\\\prefix.xml";
            //mehr_Inhalt(c);
            myfout = new BufferedWriter(new FileWriter(fileName,
true));

            //      mehr_Inhalt("<?xml version="1.0" encoding="UTF-
8"?>");
        } catch (IOException e) {

            System.out.println("IOException:");
            e.printStackTrace();

        }

    }

}
```

## Appendix A

```
}  
    public void mehr_Inhalt(String tw) {  
        this.tw = tw;  
        try {  
  
            myfout.write(tw);  
  
        } catch (IOException e) {  
            // TODO Auto-generated catch block  
            e.printStackTrace();  
        }  
    }  
  
    public void fileclose() {  
        try {  
            // mehr_Inhalt("</MetadataFormatList>");  
            myfout.close();  
        } catch (IOException e) {  
  
            System.out.println("IOException:");  
            e.printStackTrace();  
  
        }  
  
    }  
  
}
```

Das erzeugte File ist wiederum eine XML Datei, die in mit Hilfe von Access und Excel ausgewertet wird.

Alle notwendigen Files um eine erneute Abfrage zu machen, sind unter der Adresse xyz oder/und auf der beiliegenden CD verfügbar.

## Appendix B

### Appendix B

Eine Auflistung einiger NISO Standards sehen in folgender Tabelle [21]

Tabelle NISO Standards

Information Interchange Format	Z39.2-1994 (R2001)
Library Statistics	Z39.7-1995
International Standard Serial Numbering	Z39.9-1992 (R2001)
Guidelines for Abstracts	Z39.14-1997
Scientific and Technical Reports — Elements, Organization, and Design	Z39.18-1995
Guidelines for the Construction, Format, and Management of Monolingual Thesauri	Z39.19-1993 (R1998)
Criteria for Price Indexes for Printed Library Materials	Z39.20-1999
Proof Corrections	Z39.22-1989
Standard Technical Report Number Format and Creation	Z39.23-1997
Micropublishing Product	Z39.26-1997

## Appendix B

Information	
Information on Microfiche Headers	Z39.32-1996
Printed Information on Spines	Z39.41-1997
Standard Address Number (SAN) for the Publishing Industry	Z39.43-1993 (R2001)
Extended Latin Alphabet Coded Character (ANSEL)	Z39.47-1993 (R1998)
Permanence of Paper for Publications and Documents in Libraries and Archives	Z39.48-1992 (R1997)
Information Retrieval (Z39.50): Application Service Definition and Protocol Specification (Version 3)	Z39.50-1995
Codes for the Representation of Languages for Information Interchange	Z39.53-2001
Serial Item and Contribution Identifier (SICI)	Z39.56-1996
Eye-Legible Information on Microfilm Leaders and Trailers and on Containers of Processed Microfilm on Open Reels	Z39.62-2000
East Asian Character Code for Bibliographic Use	Z39.64-1989 (R1995)
Durable Hardcover Binding for Books	Z39.66-1992 (R1998)
Holdings Statements for	Z39.71-1999

## Appendix B

Bibliographic Items	
Single-Tier Steel Bracket Library Shelving	Z39.73-1994 (R2001)
Guides to Accompany Microform Sets	Z39.74-1996
Data Elements for Binding Library Material	Z39.76-1996
Guidelines for Information About Preservation Products	Z39.77-2001
Library Binding	Z39.78-2000
Environmental Conditions for Exhibiting Library and Archival Materials	Z39.79-2001
Title Pages for Conference Publications	Z39.82-2001
Syntax for the Digital Object Identifier	Z39.84-2000
The Dublin Core Metadata Element Set	Z39.85-2001
Specifications for the Digital Talking Book	Z39.86-2002
Electronic Manuscript Preparation and Markup	ANSI/NISO/ISO 12083
Environmental Guidelines for the Storage of Paper Records	TR01-1995
Guidelines for Indexes and Related Information Retrieval Devices	TR02-1997
Guidelines for Alphabetical Arrangement of Letters and	TR03-1999

## Appendix C

Sorting of Numerals and Other Symbols	
--	--

## Appendix C

### Beispiele für die Definition der 1.1 DCMI Metadaten Begriffe

#### Begriffsname: date

URI: <http://purl.org/dc/elements/1.1/date>

Label: Date

Definition: Ein Zeitpunkt oder eine Zeitspanne, die mit einem Ereignis während der Lebenszeit der Ressource zusammenhängt

Anmerkung: Das Datum kann benutzt werden um temporäre Informationen, die jegliche Granularität haben können, auszudrücken. Empfohlene Praxis ist eine Verwendung eines Kodierungsschemata wie das W3CDTF Profil, der ISO 8601 [W3CDTF].

Hinweise: [W3CDTF] <http://www.w3.org/TR/NOTE-datetime>

Begriffstyp: Element

Status: empfohlen

Ausgabedatum: 1999-07-02

#### Element Verbesserungen/Erweiterungen:

Begriffsbezeichnung: created

URI: <http://purl.org/dc/terms/created>

Label: Created

Definition: Erschaffungsdatum der Ressource

Begriffsart: Element-Verbesserung

Verbessert: <http://purl.org/dc/elements/1.1/date>

Status: empfohlen

Ausgabedatum; 2000-07-11

#### Kodierungsschema

Begriffsbezeichnung: DCMI Typ

URI: <http://purl.org/dc/terms/DCMI Type>

Label: DCMI Typen Vokabular

Definition: Eine Liste von Typen, die benutzt werden, um die Eigenschaften oder das Genre des Inhalts der Ressource zu kategorisieren

Siehe auch: <http://dublincore.org/documents/dcmi-type-vocabulary/>

Begriffsart: Kodierungsschema

Qualifiziert: <http://purl.org/dc/elements/1.1/type>

## Appendix C

Status: registriert [28]

Ein Satz von empfohlenen Kodierungsschemata ist von DCMI verfügbar.

Ausdrucksweise für den "Simple Dublin Core in RDF/XML"  
DCMI Empfehlungen seit 2002-07-31

Richtlinien für das Implementieren von "Dublin Core in XML"  
Es existieren DCMI Empfehlungen seit 2003-04-02

Ausdrucksweise für den "Qualified Dublin Core in RDF/XML"  
Vorgeschlagene DCMI Empfehlungen gibt es seit 2002-05-15

Ausdrucksweise für die qualifizierten Dublin Core Meta- und Verknüpfungselemente in HTML/XHTML, DCMI Empfehlungen bestehen seit 2003-11-30 [32]

Konkret sieht "Das Dublin Core Element Set 1.1." 15 Elemente für die Beschreibung elektronischer Dokumente vor. [33]

Es kann eine Einteilung in drei Gruppen vorgenommen werden:

- **Inhaltsbezogene Elemente (Content)**

Title: Titel des Dokuments

Subject: Thema des Dokumentes

Description: ist eine Beschreibung der Ressource in Textform oder ein Link, der zur Beschreibung führt

Type: Dieses Feld gibt an um welche Art der Ressource es sich hier handelt. Ist es ein Roman, ein Gedicht, eine technische Beschreibung, oder ein Wörterbuch, etc... Es gibt dazu ein Vokabular, welches aus der DCMI entstanden ist (DCMI Type Vocabulary)

Source Verweis zur Quelle, wo diese Ressource hergeleitet wurde.

Relation: Dieses Feld stellt Verbindungen zu anderen Ressourcen, die im Zusammenhang mit dem Inhalt stehen, aber doch einen eigenständigen Inhalt vorweisen können.

Coverage: Unter diesem Begriff werden zeitliche und räumliche Aspekte abgedeckt.

- **Elemente, die sich auf das geistige Eigentum beziehen (Intellectual property).**

Creator: Urheber des Dokumentes

Publisher: Herausgeber des Dokumentes

## Appendix C

Contributor Dieses Feld listet die Personen oder Organisationen, die einen wesentlichen Beitrag zur Ressource geleistet haben.

Rights: Hinter diesem Feld stehen die rechtlichen Bedingungen oder es werden Links eingefügt, die auf Nutzungsbedingungen verweisen.

- **Identifizierende Elemente (Instantiation)**

Date: Ist das Datum, welches nach den Standards ISO 8601 und W3C Datetime empfohlenen wurde.

Format: Physikalische oder digitalisierte Daten zur Quelle wie z.B. Datenformat oder Größe. Ähnlich wie bei Element Type gibt es Empfehlungen, woher das definierte Format entnommen werden sollte.

Identifier: Ist eine eindeutige Identifikation der Ressource innerhalb eines Zusammenhangs wie z.B. URI (Uniform Resource Identifier), DOI (Digital Object Identifier) oder ISBN (International Standard Book Number), etc.

Language: Die Sprache, in der die Inhalte verfasst wurden. Das Feld ist mit dem ISO Code 639 definiert. Es ist ein zweistelliger Code und hat darauf den optionalen Code für den Landescode, nämlich ISO 3166.

DC hat noch ein zusätzliches Merkmal, das jedes Element durch zehn Attribute kennzeichnet. Diese Kennzeichnung ist nach ISO Standard 11179 Information Technology - Specification and Standardization of Data Elements - gegeben.

Die Elemente sind folgende: Name, Identifier, Version, Registration Authority, Language, Definition, Obligation, Datatype, Maximum Occurance, Comment. Sechs von diesen Elementen haben jedoch folgende vorgegebene Werte:

Version: 1.1

Registration Authority: Dublin Core Metadata Initiative

Language: en

Obligation: Optional

Datatype: Character String

Maximum Occurance: Unlimited. [34]

### Qualifier

Die Qualifier, die wie wir bereits gelesen haben, aus 15 Elementen bestehen sind bei vielen Anwendungen nicht ausreichend. Dublin Core bietet die Möglichkeit dieses Schema zu erweitern um feinere Strukturen und eine ausführliche Semantik zu ermöglichen. Diese Erweiterungen werden das "fully qualified Dublin Core" genannt.

Es wird bestimmten Elementen ein vordefiniertes Vokabular zugewiesen. Wenn nun ein Nutzer auf einen "unbekannten" Begriff stößt, sollte er dennoch in der Lage sein, den Qualifier zu ignorieren und das Element so nützen als ob es nicht qualifiziert wäre. Dieses Verfahren nennt man das Dump-down Prinzip. [36]

## Appendix C

Die Qualifier wurden von Arbeitsgruppen der Dublin Core Metadaten Initiative entwickelt. Zurzeit gibt es zwei Klassen von Qualifiern:

### 1. Element Refinement

Ein Qualifier grenzt ein Element ab oder definiert es näher. Unterteilungen für definierte Verbesserungen/Erweiterungen (Refinement) müssen öffentlich zugänglich sein. Zum Beispiel:

Title	alternative
Description	tableofContents
	abstract
Date	created
	valid
	available
Format	issued
	modified
	extent
	medium
Relation	isVesionof
	hasVersion
	isReplacedby
	replaces
	isRequiredby
	Requires
	isPartOf
	hasPart
	isReferencedBy
	References
Coverage	isFormatOf
	HasFormat
	conformsTo
Coverage	Spatial
	temporal

Tabelle 1

[36]

### 2. Encoding Scheme

Die Encoding Scheme Qualifier bestimmen Schemata, welche eine Hilfestellung leisten, um einen Wert eines Elementes zu interpretieren. Angaben welche

## Appendix C

Regelwerke und welches kontrolliertes Vokabular bzw. Notationen verwendet wurden, sind auch enthalten.

Wie bereits erwähnt müssen auch diese Qualifier klar bestimmt und öffentlich zugänglich sein.

<b>Element</b>	<b>Encoding Scheme</b>	<b>Definition</b>
Subject	LCSH	Library of Congress Subjects Headings
	MeSH	Medical Subject Headings
	DDC	Dewey Decimal Classification
	LCC	Library of Congress Classification
	UDC	Universal Decimal Classification
Date	Period	DCMI Period
	W3CDTF	W3C Encoding rules for dates and times
Resource Type	DCMI Type	DCMI Type Vocabulary
Format	IMT	Internet Media Types
Resource Identifier	URI	Uniform Resource Identifiers
Language	ISO639-2	ISO 639-2: Codes for the representation of names and countries
Relation	URI	Uniform Resource Identifiers
Spatial	Point	DCMI Point
	ISO3166	ISO 3166:Codes for the representation of names of countries
	Box	DCMI Box
	TGN	Getty Thesaurus of Geographic Names
Temporal	Period	DCMI Period
	W3CDTF	W3C Encoding rules for dates and times

## Appendix D

Ich möchte nun jeweils ein Beispiel für Austauschformate des Dublins Core Standards anführen. Die Ausarbeitungen basieren auf 3 wesentlichen Standards,

- auf einem HTML-basierenden Format
- einem XML-basierenden Format
- einem RDF-basierenden Format

Das HTML basierende Format sieht im Detail so aus und beschreibt ein fiktives Dokument.

<HTML Coding einfügen>

XML-basierendes Format: Das Beispiel beschreibt Metadaten der Firma UKOLN (UK Office for Library and Information Networking).

<XML coding einfügen>

RDF-basiertes Format. In diesem Fall wird der Dublin Core Standard in Zusammenhang mit RDF verwendet.

<RDF coding einfügen>

Allgemein kann man sagen, dass der Dublin Core Standard sehr einfach gehalten ist, d.h. ein fast jeder Autor kann diesen Standard ohne größeren Aufwand verwenden. Elementareinträge lassen sich von grundsätzlich aus vorhandenen Metadatenbanken oder Medienbeschreibungen anderer standardisierter Beschreibungsschemata befüllen.

Die Werte der Metadatenelemente und deren Attribute werden als Strings abgebildet, somit kann Freitext eingegeben werden.

Man kann als negativen Punkt hinsichtlich des DC Standards erwähnen, dass das zugrunde liegende Datenmodell sehr einfach und trivial gehalten ist, und somit die detaillierte Ausdrucksfähigkeit darunter leidet. [35]

## Appendix D

```
<didl:DIDL xmlns:didl="urn:mpeg:mpeg21:2002:01-DIDL-NS">
  <didl:Container>
    ...
    <!-- Item containing content -->
    <didl:Item>
      ...
      <!-- ObjectType of Item -->
      <didl:Descriptor>
        <didl:Statement mimeType="text/xml; charset=UTF-8">
          <dip:ObjectType xmlns:dip="urn:mpeg:mpeg21:2002:01-DIP-NS">
            urn:my:Argument</dip:ObjectType>
          </didl:Statement>
        </didl:Statement>
      </didl:Descriptor>
    </didl:Item>
  </didl:Container>
</didl:DIDL>
```

## Appendix D

```
</didl:Descriptor>
...
</didl:Item>
...
<!-- Processing Item -->
<didl:Item>
  <!-- Qualification of the Item as Processing Item -->
  <didl:Descriptor>
    <didl:Statement mimeType="text/xml; charset=UTF-8">
      <dii:Type xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS">
        urn:mpeg:mpeg21:2002:01-DIP-NS:PI</dii:Type>
      </didl:Statement>
    </didl:Descriptor>
    <!-- Processing Item identification -->
    <didl:Descriptor>
      <didl:Statement mimeType="text/xml; charset=UTF-8">
        <dii:Identifier xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS">
          urn:bar:a333766936</dii:Identifier>
        </didl:Statement>
      </didl:Descriptor>
      <!-- Actual processing method -->
      <didl:Component>
        <didl:Descriptor>
          <!-- Argument of processing method -->
          <didl:Statement mimeType="text/xml; charset=UTF-8">
            <dip:Argument xmlns:dip="urn:mpeg:mpeg21:2002:01-DIP-NS">
              urn:my:Argument</dip:Argument>
            </didl:Statement>
          </didl:Descriptor>
          <!-- Actual code for processing method-->
          <didl:Resource mimeType="...">
            ...Link to processing code...</didl:Resource>
          </didl:Component>
        </didl:Item>
      </didl:Container>
    </didl:DIDL>[101]
```