



KIM
Kompetenzzentrum
Interoperable Metadaten

Metadatenstandards

Teil 2 der Einführung in die Interoperabilität von Metadaten und
Metadatenformaten

Sarah Hartmann (DNB)



Metadatenstandards

- sind standardisierte Beschreibungsschemata
 - Spezifikationen, die Metadaterme detailliert beschreiben
- dienen der Festlegung von
 - Elementen, deren Struktur und deren Bedeutung
- ermöglichen die semantische Interoperabilität zwischen Anwendungen, die diesen Standard verwenden
- Ziel
 - einheitliche, maschinenverarbeitbare Beschreibungen von Ressourcen und der Austausch dieser Beschreibungen



Warum sind Standards so wichtig?

- „gemeinsame Sprache“
- Austausch von Metadaten
 - durch Standard verständlich und interpretierbar
- notwendig für den effektiven Einsatz von Metadaten
 - Datenformate für effektive Datenhaltung
 - konsistente Daten für effektives Retrieval
- Dokumentation
- Transparenz
- Verknüpfungen von Objekten



Standards - Auswahl

BibTeX

IPTC-NAA

MPEG-7

MPEG-21

EAD

FOAF

Dublin Core

MARC 21

MAB

MODS

DDC

METS

LOM

EXIF

MOF

FGDC

XMP

ONIX

SIOC

PREMIS

VRA Core

ID3-Tags

XBMF

RAK

RDA

MAB



Typen von Metadatenstandards

- **Datenstruktur und Semantik**
 - Formate (z. B. DC, PICA, MARC) und Syntax (z. B. XML, RDF)
- **Inhalte**
 - Regelwerke (z. B. RAK-WB, AACR2, RDA)
- **Werte**
 - Normdaten, kontrollierte Vokabulare (z. B. PND, LCSH, DDC)
- **Kommunikation**
 - Schnittstellen, Protokolle (z. B. Z39.50, SRU, OAI)



Typen von Formaten

- **Erfassungsformat**
 - dient der Eingabe von Metadaten z. B. zur Erfassung von Titel- oder Normdatensätzen (z. B. PICA3-Format)
 - sollte möglichst einfach anwendbar sein für den Metadaten-Erfasser
- **Internformat**
 - Speicherformat, das zur Speicherung und Verwaltung von Metadaten in Datenbanken dient (z. B. PICA+)
- **Austauschformat**
 - dient dem Austausch von Metadaten zwischen Partnern oder Datenbanken, d. h. das Format enthält Absprachen zur Übertragung von Daten (z. B. MARC 21)



Dublin Core



- Standard zur Beschreibung aller Arten von Ressourcen
- internationaler Standard
- domain-übergreifender Standard
 - Museen, Bibliotheken, Archiven, Regierungsbehörden, Wirtschaft
- Pflege und Weiterentwicklung durch Dublin Core Metadata Initiative (DCMI)
- Entstehen: 1. Dublin Core Metadata Workshop, März 1995, Dublin, Ohio (USA)
- Ziele
 - Einfachheit der Semantik und der Anwendung
 - Basis für semantische Interoperabilität bereitstellen
- Mappings von umfangreicheren Formaten zu DC
- <http://dublincore.org/>



Dublin Core Metadata Element Set (DCMES)

- einfaches Element Set (DC Simple)
 - 15 Elemente

Contributor	Coverage	Creator	Date
Description	Format	Identifier	Language
Publisher	Relation	Rights	Source
Subject	Title	Type	

- alle optional, alle wiederholbar
- Werte als Strings
- **ISO-Standard (seit 2003)**
- <http://dublincore.org/documents/dces/>



DCMI Metadata Terms (DCMI-Terms)

- **Erweiterung des DC Element Sets**
- **Element Refinements**
 - Spezifizierung der Elemente (z. B. date issued, date created)
- **Encoding Schemes**
 - zusätzliche Informationen zu einem Wert
 - Vocabulary Encoding Schemes
 - kontrolliertes Vokabular (z. B. DDC)
 - Syntax Encoding Schemes
 - standardisierte Zeichenketten (z. B. W3C-DTF für Datumsangaben)
- <http://dublincore.org/documents/dcmi-terms/>



PICA

- **Verwendung**
 - als Katalogisierungs- bzw. Erfassungsformat und Internformat
 - Erfassungsformat – PICA3
 - Internformat – PICA+
- **OCLC-PICA-Software für zentrale und lokale Bibliotheksdatenbanken**
- **Anwender**
 - Bibliotheksverbände (GBV, HeBIS, SWB), DNB
- **sehr granulares Format**



Maschinelles Austauschformat für Bibliotheken (MAB)

- nationales Austauschformat
 - Titeldaten, Normdaten und Lokaldaten
- 1973 erstellt, 1995 neue Version: MAB2
- MAB XML
- bis Bibliografiejahrgang 2009 von der DNB als Transportformat für ihre Datendienste eingesetzt
- Umstieg auf MARC 21



MARC 21



- **internationales Austauschformat**
 - bibliografische Daten, Normdaten, Besitznachweise
 - aber auch Einsatz als Erfassungs- und Internformat
- **Entstehung**
 - aus UKMARC und CANMARC in den 60er Jahren
- **Pflege und Weiterentwicklung**
 - Network Development and MARC Standards Office (NDMSO)
- **Formatfamilie**
 - UNIMARC, UKMARC, MARC 21, MARCXML
- **Umstieg von MAB2 auf MARC 21 in Deutschland und Österreich**
 - DNB: Titeldaten (seit Dezember 2008), Normdaten sowie Titel- und Bestandsdaten der ZDB (seit Februar 2009)
- <http://www.loc.gov/marc/>



Metadata Encoding and Transmission Standard (METS)



- **Format zur Beschreibung von komplexen digitalen Bibliotheksobjekten**
 - Struktur dieser Objekte
 - deskriptive und administrative Metadaten
- **Einsatzbereich**
 - Verwaltung von digitalen Objekten und Austausch dieser Objekte
- **XML-Dokumentformat**
- **„Containerformat“**
 - Einbindung verschiedener Metadatenstandards
- <http://www.loc.gov/standards/mets/>



Metadata Object Description Schema (MODS)



- Standard für deskriptive Metadaten
- Derivat von MARC 21
- Fokus liegt auf der Beschreibung von digitalen Objekten
 - auch einzelne Komponenten von Objekten beschreibbar
- XML-basiert
- aktuelle Version 3.3 (2008)
 - öffentliches Review von MODS 3.4 (seit April 2010)
- Pflege und Weiterentwicklung
 - Network Development and MARC Standards Office (NDMSO)
- verwandtes Format
 - MADS (Metadata Authority Description Standard)
- <http://www.loc.gov/standards/mods/>



ONIX



- internationaler Metadatenstandard für Produkte des Buchhandels
- Austauschformat für Verlage und Buchhandel
 - z. T. auch von Bibliotheken eingesetzt
- bibliografische und vertriebsrelevante Metadaten
- XML-basiertes Format
- Pflege und Weiterentwicklung
 - ONIX International Steering Committee von EDItEUR, nationale Gruppen
- **Formatfamilie**
 - ONIX for Books, ONIX for Serials, ONIX for Licensing Terms, ONIX for DOI Registration Format, ONIX ISTC Registration Formats
- <http://www.editeur.org/8/ONIX/>



Regeln für die alphabetische Katalogisierung (RAK)

- **Regelwerk für die formale Erschließung**
- **Entstehen**
 - 1977 erstes Rahmenregelwerk für ÖB und WB
 - 1983 Erscheinen der RAK für verschiedene Bibliotheks- und Dokumenttypen
- **RAK-WB, RAK-ÖB, RAK-Musik, RAK-Karten, RAK-Körperschaften, RAK-NBM**
- **Basis**
 - International Standard Bibliographic Description (ISBD)



Anglo-American Cataloguing Rules (AACR)

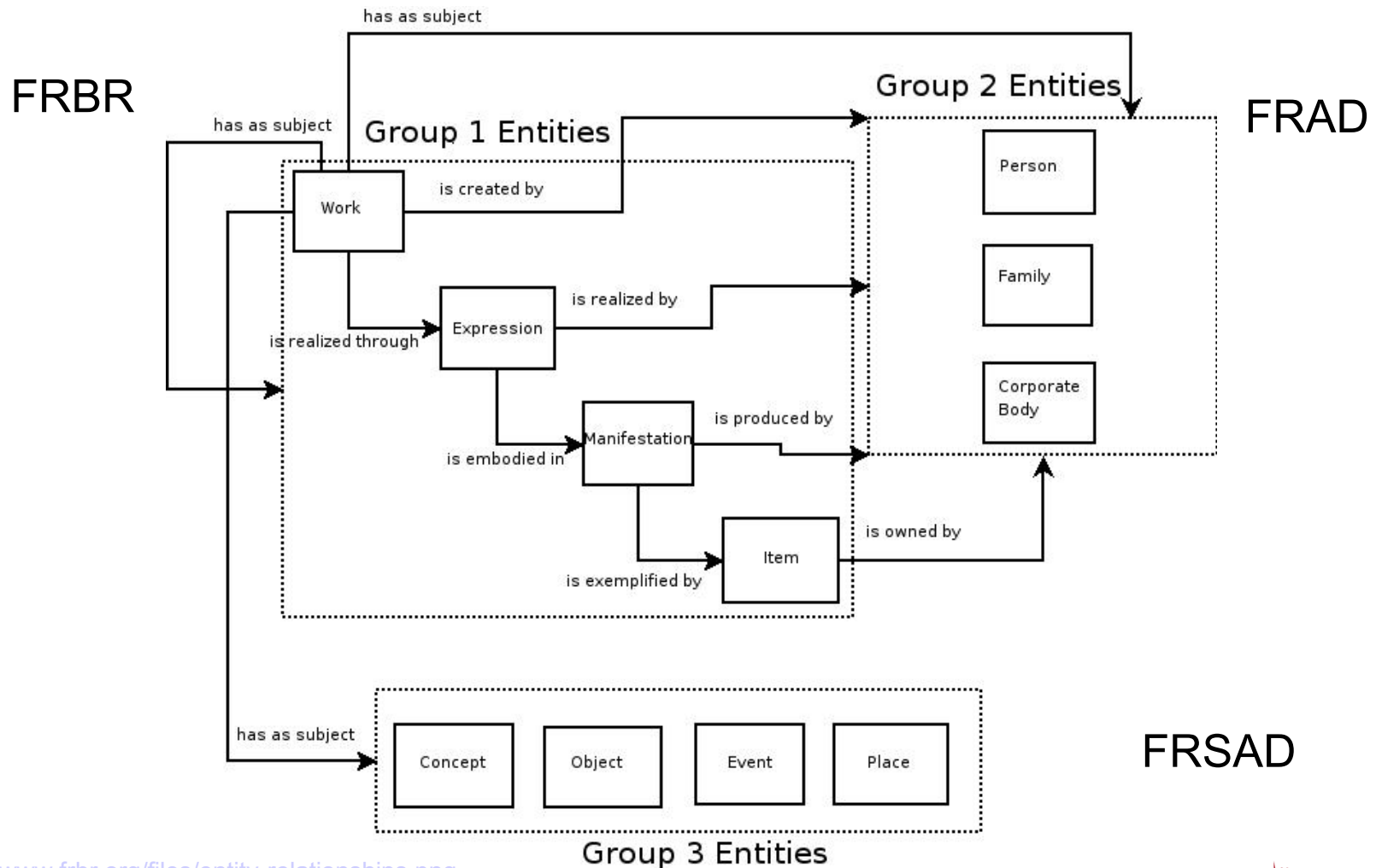
- **Regelwerk für die formale Erschließung**
- **Anwender**
 - in über 30 Ländern weltweit
 - z. B. Bibliotheken in GB, USA, Kanada, LOC, BL
- **Pflege und Weiterentwicklung**
 - Joint Steering Committee (for Revision of AACR)
- **Entstehen**
 - 1. Ausgabe 1967
 - 2. Ausgabe 1978 (AACR2)
 - letzte Revision 2002 (AACR2r), Updates
- <http://www.aacr2.org/>



- neuer internationaler Erschließungsstandard zur Erschließung aller Objekttypen
- Nutzerbedürfnisse stehen im Vordergrund
- community-übergreifender Standard
 - Bibliotheken, Museen, Archive
- basierend auf FRBR, FRAD
- unabhängig von Metadatenformaten
- Fokus auf Relationen, Rollen, kontrolliertem Vokabular
- Veröffentlichung
 - 1. Release des „RDA-Toolkit“ im Juni 2010
- Mappings von RDA zu Dublin Core, ISBD und MARC 21
- <http://www.rdatoolkit.org/>



Exkurs: FRBR, FRAD, FRSAD



<http://www.frbr.org/files/entity-relationships.png>



Göttingen, 11. Mai 2010



Normdaten

- **Vorteil**
 - nicht alle Metadaten, die zu einem Objekt gehören, müssen im Datensatz, der diese Entität beschreibt, aufgeführt werden
 - Rationalisierung und Zusammenführung von Metadaten
- **Typen von Normdaten (nach FRSAD)**
 - Personen, Körperschaften, Familien, WEMI, Objekt, Geografikum, Sachbegriff, Zeit
- **Datenelemente**
 - Identifier, Namensformen (Ansetzungsform, alternative Benennungen, Synonyme, Äquivalente), Attribute, Relationen



Normdateien

- **nationale Normdateien**
 - PND, GKD, SWD, (ZDB)
 - GND
- **internationale Normdateien**
 - z. B. LC Authorities, RAMEAU
- **Mappings und Crosskonkordanzen**
 - VIAF
 - MACS
 - CrissCross



Nationale Normdateien

- **Personennamendatei (PND)**
 - ca. 3,3 Mio. Datensätze
- **Schlagwortnormdatei (SWD)**
 - ca. 880.000 Datensätze
- **Gemeinsame Körperschaftsdatei (GKD)**
 - ca. 1,3 Mio. Datensätze
- **Zukunft**
 - Zusammenführung zu einer gemeinsamen Normdatei (GND)
- **Zeitschriftendatenbank (ZDB)**
 - ca. 1,5 Mio. Datensätze
- **Zugriffsmöglichkeiten**
 - Datenbankzugriff zum PICA-Iltis-System, SRU- und OAI-Schnittstelle, DNB-Portal, Linked Data Prototyp



Personennamendatei (PND)

- **überregionale Normdatei für Personennamen**
 - Personennamen aus allen Jahrhunderten, aus allen Fach- und Wissensgebieten, aller Sprachen aus allen Staaten
- **Verwendung**
 - formale und inhaltliche Erschließung von Ressourcen
- **Pflege und Weiterentwicklung**
 - kooperativ: alle deutschen und österreichischen Bibliotheksverbände, Schweizerische Nationalbibliothek, Projekte, Archive, Museen
- **zwei Satzarten für Namen (Tn) und Personen (Tp)**



Gemeinsame Körperschaftsdatei (GKD)

- **überregionale Normdatei für deutsche und ausländische Körperschaftsnamen**
- **Entstehen**
 - Aufbau in den 70er Jahren aus den Körperschaftsaufnahmen der Bayerischen Staatsbibliothek, der Staatsbibliothek Preußischer Kulturbesitz und der Deutschen Bibliothek
- **Verwendung**
 - formale Erschließung von Ressourcen
- **Inhalt**
 - Typen: religiöse Körperschaften, musikalische Körperschaften, Kongresse, Firmen, Gebietskörperschaften, Organe von Gebietskörperschaften
 - Ansetzungs- und Verweisungsformen von Körperschaftsnamen (nach RAK-WB), chronologische und hierarchische Relationen, weitere Attribute



Schlagwortnormdatei (SWD)

- **überregionale Normdatei für Schlagwörter**
 - normierter, terminologisch kontrollierter Wortschatz
- **Pflege und Weiterentwicklung**
 - kooperativ: deutsche und österreichische Bibliotheksverbände, Schweizerische Nationalbibliothek, DNB
- **Verwendung**
 - verbale inhaltliche Erschließung von Ressourcen
- **Inhalt**
 - alle Fachgebiete
 - Allgemein- und Individualbegriffe (Produkte, Werktitel, Körperschaften, Kongresse, Bauwerke)
 - Relationen (Oberbegriffe, verwandte Begriffe, Äquivalenzbegriffe)
 - Einordnung in SWD-Systematik und DDC-Notationen



Gemeinsame Normdatei (GND)

- **Ziele des Projektes**
 - Zusammenführung von GKD, SWD, PND, EST-DMA
 - Zusammenführung redundanter Entitäten in GKD und SWD
 - Integration aller Normdatensätze in ein Datenformat, angelehnt an MARC 21
 - gemeinsame Normdatei, um gemeinsam und eindeutig von bibliografischen Daten und der Daten anderer Normdatenanwender (z. B. Archive, Museen) darauf zu verweisen
 - Eignung für das Semantic Web
 - Erweiterung echter Relationen in Normdateien
- **Umstellung geplant 2011**



Zeitschriftendatenbank (ZDB)

- **überregionale Datenbank für periodische Veröffentlichungen (z. B. Zeitschriften, Schriftenreihen)**
 - aus allen Ländern, in allen Sprachen, ohne zeitliche Einschränkung, in gedruckter und elektronischer Form
- **Titeldaten und Bestandsnachweise**
- **Verwendung**
 - Instrument für den Leihverkehr und für Dokumentlieferdienste
 - Nachweis und Auslieferung von Nationallizenzen und elektronischen Zeitschriftenpaketen
- **Redaktion und Weiterentwicklung**
 - Staatsbibliothek zu Berlin
- **Pflege**
 - ca. 4.300 Bibliotheken
- <http://www.zeitschriftendatenbank.de>



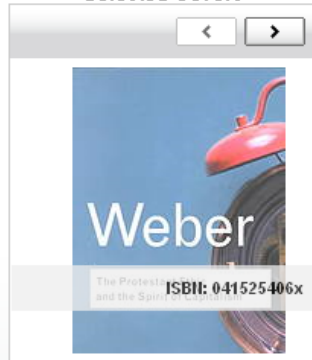
Library of Congress Authorities

- **LC Name Authority File**
 - amerikanische Normdatei für Personen- und Körperschaftsnamen
 - Format: MARC 21 for Authority Data
 - Regelwerk: AACR2
- **Library of Congress Subject Headings**
 - Verwendung: inhaltliche Erschließung z. B. LOC, BL
 - Sachschlagwörter, Geografika
- **Umfang**
 - 265.000 Schlagwörter
 - 5,3 Mio. Personen-, Körperschafts-, Kongress- und geografische Namen
 - 350.000 Serien- und Einheitssachtitel
 - 340.000 Namen/Titel Normdaten
- <http://authorities.loc.gov/>



Virtual International Authority File (VIAF)

Selected Covers



Weber, Max, 1864-1920

Weber, Max

م. فيبر, ماكس, 1864-1920

ובר, מקס, 1864-1920

VIAF ID: 100180950

<http://viaf.org/viaf/100180950>

Preferred Forms

- [100 1a Weber,Max,td 1864-1920](#)
- [100 1a م. فيبر, ماكس, ا 1864-1920](#)
- [100 1a Weber,Max,td 1864-1920](#)
- [100 1a Weber,Max](#)
- [100 1a ובר, מקס, א 1864-1920](#)
- [100 1a Weber,Max,td 1864-1920](#)
- [200 1a Weber,td Max,tf 1864-1920](#)
- [100 1a Weber,Max,td 1864-1920](#)
- [100 1a Weber,Max,td 1864-1920](#)
- [200 1a Weber,td Max,tf 1864-1920](#)
- [100 1a Weber,Max,td 1864-1920](#)
- [100 1a Weber,Max,td \(1864-1920\)](#)



Deutsche Nationalbibliothek

DNB|118629743



Göttingen, 11. Mai 2010

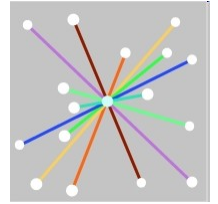


Virtual International Authority File (VIAF)

- **Ziele des VIAF**
 - Abgleich und Verlinkung von Normdateien mit maschinellen Verfahren (aktuell 19 Normdateien)
 - Nutzung der Normierungsergebnisse zur Verbesserung der weltweiten Recherche
 - Möglichkeit der Suche von Namensformen in der eigenen Sprache und der vom Benutzer bevorzugten Schrift
 - Beibehaltung nationaler bzw. regionaler Ansatzungsformen nebeneinander
- **Anwendung**
 - 13 Nationalbibliotheken
 - Fokus auf Individualnamen, im ersten Schritt Personennamen
- <http://www.viaf.org/>



CrissCross



- **Ziel des DFG-Projektes**
 - Erstellung eines multilingualen, thesaurusbasierten und benutzergerechten Recherchevokabulars zu heterogen erschlossenen Dokumenten
 - Mapping zwischen Sachschlagwörtern der SWD und Notationen der DDC
 - Verknüpfungsrichtung ist von der SWD ausgehend zur DDC
 - Mapping zwischen SWD-Sachschlagwörtern mit ihren Äquivalenten in LCSH und RAMEAU (Fortführung der MACS-Arbeiten)
- **Projektpartner: FH Köln, DNB**
- <http://www.fbi.fh-koeln.de/institut/projekte/CrissCross/index.html>



(weitere) kontrollierte Vokabulare

- **Klassifikationen**
 - z. B. Dewey Decimal Classification (DDC), Universal Decimal Classification (UDC), Library of Congress Classification, Regensburger Verbundklassifikation (RVK)
- **Thesauri**
 - z.B. Standard Thesaurus Wirtschaft (STW), Getty Thesaurus for Geographic Names (TGN)
- **sonstige Vokabulare**
 - z. B. MARC Relator Codes



Dewey Decimal Classification (DDC)

- international weit verbreitete Universalklassifikation
 - Einsatz in Bibliotheken in mehr als 135 Ländern
- von Melvil Dewey 1873 entwickelt, 1876 veröffentlicht
- Ordnung nach Fachgebieten
 - 10 Hauptklassen
 - jede Hauptklasse in 10 Klassen der Ebene 2 (divisions)
 - jede Klasse der Ebene 2 in 10 Klassen der Ebene 3 (subdivisions)
- Pflege und Lizenz
 - OCLC
- DDC Summaries als linked open data unter <http://dewey.info/>
- Übersetzungen in mehr als 30 Sprachen
- <http://www.oclc.org/dewey/>



Dewey Decimal Classification (DDC)

- **Deutsche Übersetzung**
 - der 22. Ausgabe der DDC (seit Oktober 2005)
- **Anwendung**
 - Reihen A, B und H der Deutschen Nationalbibliografie werden mit DDC-Notationen erschlossen (seit Bibliografiejahrgang 2007)
- **Online-Dienst Melvil**
 - z. B. MelvilSearch (verbale Suche, Browsing von Hauptklassen bis zu den speziellsten DDC-Klassen)
- **MelvilSearch-Daten unter Creative Commons BY-NC-ND license (seit Februar 2010)**
- <http://melvil.d-nb.de/>
- <http://ddc-deutsch.de>



Standard-Thesaurus-Wirtschaft (STW)

- **Terminologie**
 - Ökonomie und verwandte Themen (z.B. Sozialwissenschaften)
 - Fach-, Allgemein- und geografische Begriffe
- **Entstehen**
 - Mitte der 90er Jahre
- **Pflege und Weiterentwicklung**
 - durch die Deutsche Zentralbibliothek der Wirtschaftswissenschaften (ZBW)
- **Umfang**
 - ca. 6000 Deskriptoren
 - ca. 20.000 Nicht-Deskriptoren
 - deutsch und englisch
- **Creative Commons-Lizenz**
- **SKOS-Repräsentation**
- <http://zbw.eu/stw>



Getty Thesaurus of Geographic Names (TGN)

- **Terminologie**
 - Geografika (politische und physische Entitäten)
 - Fokus auf Orte, die im Bereich Kunst und Architektur von Bedeutung sind
- **Pflege und Weiterentwicklung**
 - durch Getty
- **© by J. Paul Getty Trust**
- **Umfang**
 - ca. 895.000 Deskriptoren (multilingual)
 - über 1 Millionen Terme
 - Namen, Koordinaten, Datumsangaben, Ortstypen, Erläuterungen, Quellenangaben
- <http://www.getty.edu/>



Schnittstellen

- **Interoperabilität**
 - wird gewährleistet durch definierte Metadaten-Schemata, durch Konkordanzen zwischen Schemata / Modellen
 - und durch gemeinsame Transferprotokolle
- **dienen der Kommunikation**
- **z. B. Z39.50, SRU/SRW, OAI, SOAP, SOA, SPARQL**



Z39.50

- Schnittstelle bzw. Protokoll zur Kommunikation zwischen heterogenen Bibliothekssystemen
 - technische Realisierung von Retrieval in verteilten Systemen
- Abfrage externer Datenquellen und ggf. Übernahme der strukturierten Daten
- Anwendungen z. B. in Portalen
- Entwicklung und Verwaltung
 - LOC



SRU / SRW

- Search/Retrieve via URL (SRU)
- Search/Retrieve Webservice (SRW)
- standardisiertes Webservice-Protokoll
- Weiterentwicklung von Z39.50
- Anwendung
 - Suchabfragen von Datenbanken
 - gezielte Suche mittels Suchindizes und Suchbegriffen (boolesche Suche)
 - Datenübernahme von Treffern in die eigene Anwendung
 - unterstützte Format z.B. der DNB: MARC XML, MAB-XML, OAI-DC
- Retrievalsprache CQL (Common / Contextual Query Language)
- <http://www.loc.gov/standards/sru/>



Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

- offene Schnittstelle zum Austausch von XML-Daten
- Abfragen eines OAI-Repository mit einem einfachen Web-Browser



[http://services.d-nb.de/oai/repository?
verb=**ListRecords**
&from=**2008-06-09T08:11:11Z**
&until=**2008-06-10T15:13:13Z**
&set=**ddb:publikationen:online**
&metadataPrefix=**onix**](http://services.d-nb.de/oai/repository?verb=ListRecords&from=2008-06-09T08:11:11Z&until=2008-06-10T15:13:13Z&set=ddb:publikationen:online&metadataPrefix=onix)

Quelle: urn:nbn:de:101-2009120311



SPARQL

- Simple Protocol and RDF Query Language
- Anfragesprache auf RDF-Graphen
- Protokoll zur Übertragung von SPARQL-Anfragen und – Ergebnissen
- W3C-Spezifikation (seit 15. Januar 2008)
- mehrere Datenquellen können gleichzeitig abgefragt werden
- optionale Daten bei Abfrage möglich, Filtern und Sortieren möglich

<http://www.w3.org/TR/rdf-sparql-query/>



Göttingen, 11. Mai 2010



Vielen Dank

Fortsetzung folgt!