



Einführung in die Interoperabilität von Metadaten und Metadatenformaten

Sarah Hartmann (DNB)
Stefanie Rühle (SUB Göttingen)



KIM

Teil 3 Metadatenprofile

Stefanie Rühle (SUB Göttingen)



Metadatenprofil – was ist das?

- “... a declaration specifying which metadata term an organization, information provider, or user community uses in its metadata ... identifies the source of metadata terms used ...”

<ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14855-00-2003-Nov.pdf>

- “... a type of metadata schema ...”

<http://www.ariadne.ac.uk/issue25/app-profiles/>



Metadatenprofil – was ist das

- “In the metadata community, the term application profile has been applied to describe the tailoring of standards for specific applications.”

<http://dublincore.org/documents/singapore-framework/>



Metadatenprofil

- Ein Metadatenprofil befindet sich in einem Spannungsfeld zwischen:
 - Den Anforderungen der Nutzer, die die Anwendung nutzen sollen
 - Den Eigenschaften der Objekte, die beschrieben werden sollen
 - Den Fähigkeiten des Systems, das für Erfassen, Speichern und Bereitstellung der Metadaten zuständig ist.



Metadatenprofil

Dokumentieren = für andere Menschen festhalten:

- **Für wen?**
 - Zielgruppe der Metadatenanwendung
- **Wozu?**
 - Aufgaben, Nutzerszenarien, die mit den Metadaten erfüllt werden müssen
- **Was?**
 - Objekte/Ressourcen, die mit den Metadaten beschrieben werden sollen
- **Womit?**
 - Metadaterme, die für die Beschreibung der Objekte/Ressourcen verwendet werden müssen/können,
 - Standards, die diesen Termen zugrunde liegen
 - Regeln, die für die Verwendung der Terme gelten.



Metadatenprofil – CWA 14855

Beschreibung der Terme unter Verwendung von Attributen zur

- **Identifizierung**
 - Term URI
 - Name
 - Label
 - Defined By
- **Definition**
 - Definition
 - Comments
 - Type of term
- **Beschreibung der Beziehungen**
 - Refines
 - Refined by
 - Encoding Scheme For
 - Has Encoding Scheme
 - Similar To
- **Beschreibung der Einschränkungen**
 - Obligation
 - Condition
 - Datatype
 - Occurrence



Dublin Core Collections Application Profile

Title [dc:title]

Property URI	http://purl.org/dc/elements/1.1/title		
Qualified Name for Property	dc:title		
Defined By	The Dublin Core Metadata Element Set, v1.1 http://purl.org/dc/elements/1.1/		
Type of Term	Property		
Subproperty Of	[n/a]		
Source Label	Title		
Label in this DCAP	[n/a]		
Source Definition	A name given to the resource.		
Usage in this DCAP	The name of the collection.		
Comments for this DCAP	<p>A value string must be provided.</p> <p>To record titles in multiple languages, use multiple statements and in each statement associate the value string with the appropriate language tag.</p> <p>Where an existing name is used, the value string should preserve the original wording, order and spelling of an existing name. Punctuation need not reflect the usage of the original. Subtitles should be separated from the title by a sequence of space-colon-space, for example:</p> <p style="text-align: center;">Voices from the Dust Bowl: The Charles L. Todd and Robert Sonkin Migrant Worker Collection</p>		
Uses Vocabulary Encoding Scheme	[Not specified]		
	Value URI	Value String	Syntax Encoding Scheme(s)
	Not permitted	Mandatory	Not permitted
Obligation	Mandatory		
Condition	[n/a]		
Minimum Occurrences	1		
Maximum Occurrences	unbounded		

<http://dublincore.org/groups/collections/collection-application-profile/index.shtml>





Metadatenprofil

Dokumentieren = für Maschinen lesbar machen

- Semantic Web Technologien
 - Welche Syntax liegt zugrunde
 - Welche Semantik wird verwendet
- Ressource Description Framework (RDF):
 - „Everything must be identified by URIs“
 - RDF ist eine **Aussage**, die angibt, dass eine bestimmte **Resource** (das **Subjekt**) für eine bestimmte **Eigenschaft** (das **Predikat**) einen bestimmten **Wert** (das **Objekt**) hat.
 - wird in einem Triple ausgedrückt: Subject, Predicate, Object
 - ein Triple ist ein sog. **Statement**, ein Set aus Statements ist ein **Model**



URI

- <http://www.ltg.ed.ac.uk/~ht/WhatAreURIs/>
- „URIs are the heart of the web“
- **Identifizieren Ressourcen / Objekte**
- **Ermöglichen den Zugriff auf Beschreibungen einer Ressource / eines Objekts (representation = Metadaten)**
- **Ermöglichen Zugriff auf die Ressource / das Objekt selbst (presentation)**
- **Im Semantic Web gilt: Alles ist eine Ressource = nicht nur Objekte werden mittels URI identifiziert, sondern alle Terme, die im Rahmen von Metadatenstandards und -profilen definiert und anotiert werden.**



Metadatenprofil – Sinn und Zweck

Ein Metadatenprofil steht für ein

- wohl-definiertes
- nachvollziehbares
- wiederverwendbares

Metadatenformat

Und verbessert so

- Nachnutzbarkeit des Formats
- Metadaten Austausch
- Verlinkung



Singapore Framework

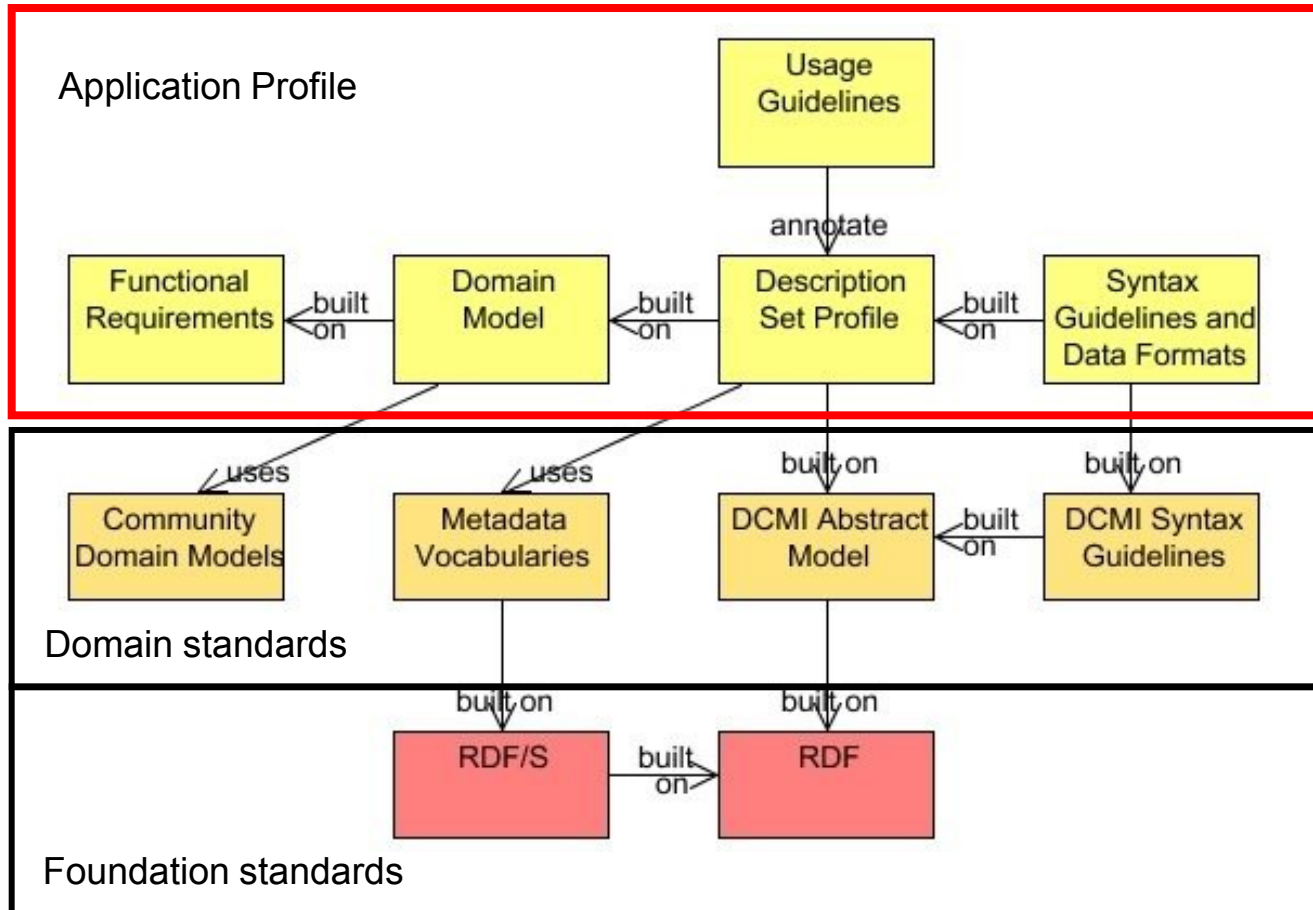
<http://dublincore.org/architecturewiki/SingaporeFramework/>

„...a framework for designing metadata applications for maximum interoperability and for documenting such applications for maximum reusability.“

- **Standard für die Dokumentation von Metadatenprofilen**
 - Definiert die Komponenten, die für Beschreibung eines Anwendungsprofils notwendig sind
 - Beschreibt wo dieses Profil im Kontext von Domain Modell und Semantic Web steht.
- **Ermöglicht die Kombination von**
 - Verschiedenen Metadatenstandards
 - Metadatenstandards und domainspezifischen Metadatermen**auf der Basis des Dublin Core Abstract Model (DCAM)**
- **Unterstützt die Interoperabilität von Metadatenanwendungen**



Singapore Framework





Singapore Framework – verpflichtende Komponenten

- “Functional requirements“
 - Was für Anforderungne werden sollen unterstützt werden?
 - Einfach Suche, Erweiterte Suche, thematisches Browsing ...
- “Domain model“
 - Was will ich beschreiben?
 - Text, Film, Personen ...
 - Was für Beziehungen bestehen zwischen diesen Objekten?
 - Ist Urheber von, ist Teil von, ...
- “Description Set Profile (DSP)“
 - Was brauche ich, um meine Ziele und Anforderungen zu erreichen?
 - Metadatenterme, Encoding schemas, Regeln ...
 - Spezifiziert die im “Domain model“ genannten Objekte



Singapore Framework – optionale Komponenten

- “Usage guidelines“
 - In welcher Form ist das Metadatenprofil zu verwenden?
 - Was muss bei der Impelementierung beachtet werden?
- “Encoding syntax guidelines“
 - Braucht es eine anwendungsspezifische Syntax?



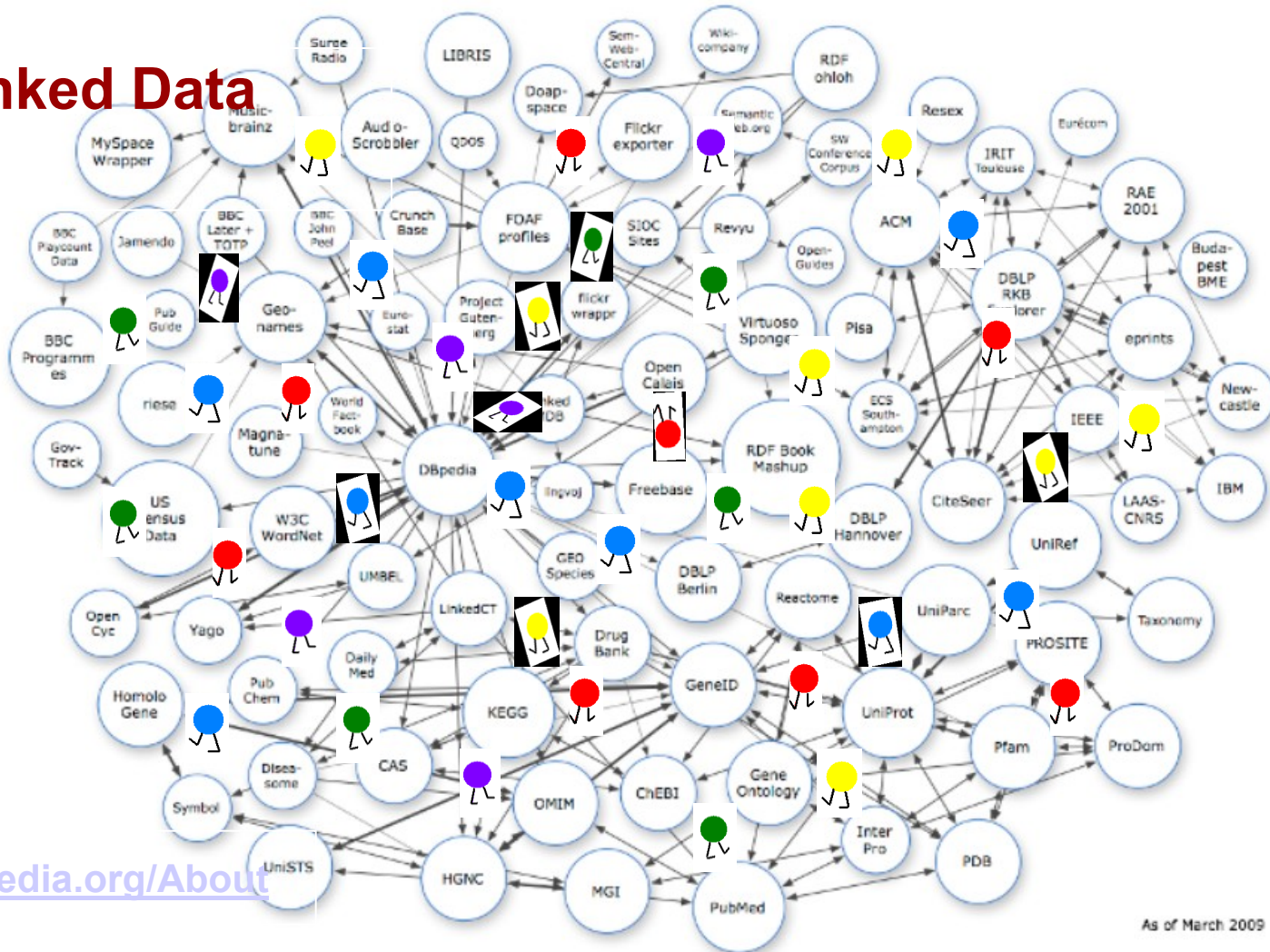
Was ist ein “Description Set Profile” (DSP)?

Und wozu brauche ich das überhaupt?



Unser Ziel ... das semantische Netz

Linked Data



<http://dbpedia.org/About>

As of March 2009



Göttingen, 11. Mai 2010



Metadatenprofil

Dokumentieren = für Maschinen lesbar machen

- Semantic Web Technologien
 - Welche Syntax liegt zugrunde
 - Welche Semantik wird verwendet
- Ressource Description Framework (RDF):
 - „Everything must be identified by URIs“
 - RDF ist eine **Aussage**, die angibt, dass eine bestimmte **“resource”** (das **Subjekt**) für eine bestimmte **“property”** (das **Prädikat**) einen bestimmten **“value”** (das **Objekt**) hat.
 - wird in einem Triple ausgedrückt: subject, predicate, object
 - ein Triple ist ein sog. **Statement**, ein Set aus Statements ist ein **Model**



“menschenslesbar“

Label	LiLa AP name of the term
URI	Specifies the URI of the given term
Term name	(Dublin Core) name of the term; abbreviation for the term URI
Defined by	Name and URI of the metadata vocabulary of the given term
Type of term	Specifies the type of the term according to the Dublin Core Metadata Initiative Abstract Model (DCAM) as class, property, syntax encoding scheme, vocabulary encoding scheme)
Refines	Specifies the property which is refined by the given term
Definition in LiLa AP	Specifies the definition of the given term in the LiLa AP context which will be based on the DC definitions in most cases
Comments and best practice examples	Recommendations on how to use this term
Encoding schemes	Name and URI of the syntax encoding scheme and/or the vocabulary encoding scheme, if applicable
Obligation	Specifies whether filling in this term is mandatory (m), recommended (r), or optional (o)
Repeatable	Specifies whether the term can be repeated
<i>Pica field for mapping</i>	<i>Specifies the Pica library catalogue field which corresponds to the LiLa AP field</i>
<i>LOM field for mapping</i>	<i>Specifies the LOM field which corresponds to the LiLa AP field</i>

Label	Subject
URI	http://purl.org/dc/elements/1.1/subject
Term name	dc:subject
Defined by	Dublin Core Metadata Element Set, Version 1.1, http://purl.org/dc/elements/1.1/
Type of term	Property
Refines	-
Definition in LiLa AP	Topic of the described LiLa content.
Comments and best practice examples	The topics of a LiLa content are specified using key words of the Library of Congress Subjects Headings (LCSH) and classifications of the Dewey Decimal Classification System (DDC). It is recommended to use a maximum of five key words. A separate statement should be used for each key word or each subject descriptor.
Encoding schemes	dcterms:LCSH (http://purl.org/dc/terms/LCSH) dcterms:DDC (http://purl.org/dc/terms/DDC)
Obligation	Mandatory
Repeatable	Yes
<i>Pica field for mapping</i>	
<i>LOM field for mapping</i>	<i>general.keywords. For those wishing more specificity of Subject, a category of classification can be used with a purpose of "Subject". classification has elements for description, keywords, and taxonpath(s) that are specific for the purpose.</i>

Table 4: fields of tables and their meanings used to describe the properties.

Quelle: LiLa Application Profile





... und das ist die Sprache, in der die Maschinen „sprechen“

```
<?xml version="1.0" encoding="UTF-8" ?> <rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dcam="http://purl.org/dc/dcam/" > <rdf:Description
  rdf:about="http://dublincore.org/pages/home">
  <dcterms:title>DCMI Home Page</dcterms:title>
  <dcterms:publisher> <rdf:Description
  rdf:about="http://example.org/agents/DCMI"> <rdf:value>Dublin
  Core Metadata Initiative</rdf:value> </rdf:Description>
  </dcterms:publisher> <dcterms:subject
  rdf:parseType="Resource"> <dcam:memberOf
  rdf:resource="http://purl.org/dc/terms/LCSH" />
  <rdf:value>Metadata</rdf:value> </dcterms:subject>
  </rdf:Description> </rdf:RDF>
```



... oder dies?

```
<?xml version="1.0" encoding="UTF-8" ?> <dcds:descriptionSet
  xmlns:dcds="http://purl.org/dc/xmlns/2008/09/01/dc-ds-xml/">
  <dcds:description dcds:resourceURI="http://dublincore.org/pages/home">
  <dcds:statement dcds:propertyURI="http://purl.org/dc/terms/title">
  <dcds:literalValueString>DCMI Home Page</dcds:literalValueString>
  </dcds:statement> <dcds:statement
  dcds:propertyURI="http://purl.org/dc/terms/publisher"
  dcds:valueURI="http://example.org/agents/DCMI">
  <dcds:valueString>Dublin Core Metadata Initiative</dcds:valueString>
  </dcds:statement> <!-- vocabulary encoding scheme URI -->
  <dcds:statement dcds:propertyURI="http://purl.org/dc/terms/subject"
  dcds:vesURI="http://purl.org/dc/terms/LCSH">
  <dcds:valueString>Metadata</dcds:valueString> </dcds:statement>
  </dcds:description> </dcds:descriptionSet>
```



Es geht darum ...

Mensch

- Objekte zu beschreiben
 - z. B. eine Homepage
- Dabei gilt: Jedes Objekt hat eine Vielzahl an Eigenschaften, die für die Beschreibung herangezogen werden können
 - z. B. den Titel einer Homepage
- Ich beschreibe ein Objekt, indem ich den verschiedenen Eigenschaften, die ein Objekt hat, einen Wert zuweise
 - z. B. DCMI Home Page ...
- Jedes Objekt gehört außerdem zu einer Klasse ähnlicher Objekte.
 - Ein Buch gehört zur Klasse der Bücher, eine Person gehört zur Klasse der Personen, ...

Maschine

```
<dcds:description  
dcds:resourceURI="http://dublincore.org/  
pages/home">
```

```
<dcds:statement  
dcds:propertyURI="http://purl.org/dc/ter  
ms/title">
```

```
<dcds:literalValueString>DCMI Home  
Page</dcds:literalValueString>  
</dcds:statement>
```

```
<dcds:statement  
dcds:propertyURI="http://purl.org/dc/ter  
ms/publisher"  
dcds:valueURI="http://example.org/agent  
s/DCMI"> <dcds:valueString>Dublin Core  
Metadata Initiative</dcds:valueString>  
</dcds:statement>
```

```
</dcds:description>
```




Es geht darum ...

Mensch

- Für die Verwendung der Werte, kann ich Regeln aufstellen
 - In dem Feld subject dürfen nur Begriffe aus dem LCSH verwendet werden.
 - Datumsangaben in dem Feld date müssen den W3C Vorgaben entsprechen.

- Indem ich ein Objekt mit einer Eigenschaft und einem Wert beschreibe, mache ich eine Aussage über das Objekt

Maschine

Vocabulary Encoding Scheme:

```
<dcds:statement  
dcds:propertyURI="http://purl.org/dc/terms/subject"  
dcds:vesURI="http://purl.org/dc/terms/LCSH">  
<dcds:valueString>Metadata</dcds:valueString>  
</dcds:statement>
```

Syntax Encoding Scheme:

```
<dcds:statement  
dcds:propertyURI="http://purl.org/dc/terms/date">  
<dcds:literalValueString  
dcds:sesURI="http://www.w3.org/2001/XMLSchema#date">2005-05-05</dcds:literalValueString>  
</dcds:statement>
```

```
<dcds:statement  
dcds:propertyURI="http://purl.org/dc/terms/publisher"  
dcds:valueURI="http://example.org/agents/DCMI">  
<dcds:valueString>Dublin Core Metadata  
Initiative</dcds:valueString> </dcds:statement>
```



Description Set Profile (DSP)

- Eine Metadatenanwendung kann dazu dienen, mehr als eine Ressource zu beschreiben
 - Dokumente, Fotos, Personen ...
- DSP fordert für jeden Ressourcentyp eine eigenes Set von Metadatenelementen = „Description Template“
 - Ein Set für Dokumente, eins für Fotos, eins für Personen, ...
- DSP erlaubt, Metadatenterme aus unterschiedlichen Standards zu verwenden
 - z. B. Für Dokumente Dublin Core Terms, für Fotos MIX, für Personen FOAF
- Voraussetzung: Die Terme müssen DCAM-kompatibel sein
 - Zuweisung von class, property, Syntax Encoding scheme, vocabulary encoding scheme
 - Korrekte URI
- Beispiel: Scholarly Works Application Profile (SWAP)
 - <http://dublincore.org/scholarwiki/SWAPDSP>



Metadatenprofil und Interoperabilität

- Metadatenprofile verbessern die Interoperabilität von Anwendungen, indem sie dokumentieren, wozu was womit beschrieben wird und machen die Metadaten damit
 - für Menschen verständlich
 - für Maschinen lesbar
- Metadatenprofile dokumentieren die lokalen Gegebenheiten einer Anwendung
- Um Metadatenbestände, denen unterschiedliche Metadatenprofile zugrundeliegen, gemeinsam anzubieten (z. B. Europeana) braucht es ein Mapping.



Mapping

Datenbestände aus unabhängigen, heterogenen Systemen müssen auf ein gemeinsames Datenformat gemappt werden.

Auf der Grundlage des kleinsten gemeinsamen Nenners

- Mapping ist das Konvertieren der Feldinhalte eines Datenformats aus einer Struktur in eine andere.
- Mapping funktioniert nur, wenn Metadatenformate interoperabel sind.



Schwierigkeiten beim Mapping

- Unterschiedliche Zeichencodes (Latin1, UTF ...)
- Unterschiedliche Codes für z.B. Sprachen, Länder
- Unterschiedliche Interpunktionsregeln
- Textelemente in einer Sprache müssen in andere Sprache übersetzt werden
- Redundanzen müssen beseitigt werden
- Felder müssen zerlegt werden
- mehrere Felder müssen zusammengefügt werden
- Einzelne Sätze werden in mehrere Sätze zerlegt (Haupt- + Untersätze bei mehrbändigen Werken; Exemplarsätze)
- Mehrere Sätze werden zu einem Datensatz zusammengefasst
- Verknüpfungsstrukturen müssen geändert werden (mehrbändige Werke, Normdatenverknüpfungen)
- ...

<http://www.allegro-c.de/formate/formneu.htm>



Beispiel MAB zu Dublin Core

MAB	Definition MAB	DC-Element	DC Definition
331 / 310	Hauptsachtitel in Vorlageform oder Mischform / Hauptsachtitel in Ansetzungsform	DC.Title	A name given to the resource
335 / 370	Zusätze zum Hauptsachtitel / Weitere Sachtitel	DC.Title.Alternative	Any form of the title used as a substitute or alternative to the formal title of the resource.
304	Einheitssachtitel	DC.Title.Alternative	Any form of the title used as a substitute or alternative to the formal title of the resource.
376	Normierte Zeitschriftentitel	DC.Title.Alternative	Any form of the title used as a substitute or alternative to the formal title of the resource.
670	Sachtitel in abweichender Orthographie	DC.Title.Alternative	Any form of the title used as a substitute or alternative to the formal title of the resource.
104; 108-196	Name der Personen(en)	DC.Creator	An entity primarily responsible for making the content of the resource.
204; 208-296	Name der Körperschaft(en)	DC.Creator	An entity primarily responsible for making the content of the resource.

Quelle: http://db1-www.sub.uni-goettingen.de/servlets/metaformList1?Table=MAB_DC&Head=MAB_DC



Dumb-Down

- Prinzip der Dublin Core Community: granulare Strukturen lassen sich auf allgemeinere Strukturen mappen
 - dcterms.issued -> dc.date
 - dcterms.alternative -> dc.title
 - dcterms.extent -> dc.format
- Die granularen Terme werden im Metadatenprofil als refinements der allgemeineren Terme beschrieben.



Mapping

- **Vom Besonderen zum Allgemeinen**
 - dcterms.alternative -> dc.title
- **Vom Allgemeinen zum Besonderen**
 - dc.date -> dcterms.issued oder dcterms.modified oder dcterms.created
- **Vom Besonderen zum Besonderen**
 - MAB Kategorie 670 -> dcterms.alternativ
- **Vom Allgemeinen zum Allgemeinen**
 - dc.title -> MAB Kategorie 331 oder 310
- **In allen Fällen gilt: Mapping ist aufwendig und führt zu Informationsverlust**



Zusammenfassung

- **Metadatenprofile** dokumentieren, **was wozu womit für wen** mit Metadaten beschrieben wird
 - „menschenslesbar“
 - „maschinenslesbar“
- **Metadatenprofile erleichtern**
 - die Nachnutzbarkeit
 - den Austausch von Metadaten
 - die Verlinkung zwischen Konzepten
- **Das Singapore Framework** ist ein Standard, der beschreibt, wie ein Metadatenprofil aufgebaut ist. Komponenten eines Metadatenprofils sind:
 - Functional requirements
 - Domain model
 - Description Set Profile (DSP)
 - Usage Guidelines
 - Encoding syntax guidelines



Zusammenfassung

- Das **DSP** fordert für jede Art von Ressourcen, die in einer Anwendung beschrieben werden, ein eigenes Beschreibungstemplate.
- Dabei gilt: Im DSP können Terme aus verschiedenen Standards zusammengeführt werden vorausgesetzt sie sind DCAM-kompatibel.
- Um unterschiedliche Formate in einem gemeinsamen Format anzubieten braucht es ein **Mapping**
- Mapping hat fast immer Informationsverlust zur Folge



Übung 3

- Erstellen Sie ein Mapping für Ihre Metadaterme zu dem folgenden Format
 - dc.title
 - dc.contributor
 - dc.date
 - dc.identifer
 - dc.type
- Für das Format gilt:
 - Alle Terms sind verpflichtend
 - Das folgende kontrollierte Vokabular für dc.type ist ebenfalls verpflichtend:
 - Buch
 - Internet
 - Audio und Video
 - Sonstige



Vielen Dank



Göttingen, 11. Mai 2010