



Evaluierung und Zertifizierung von Metadatenprofilen

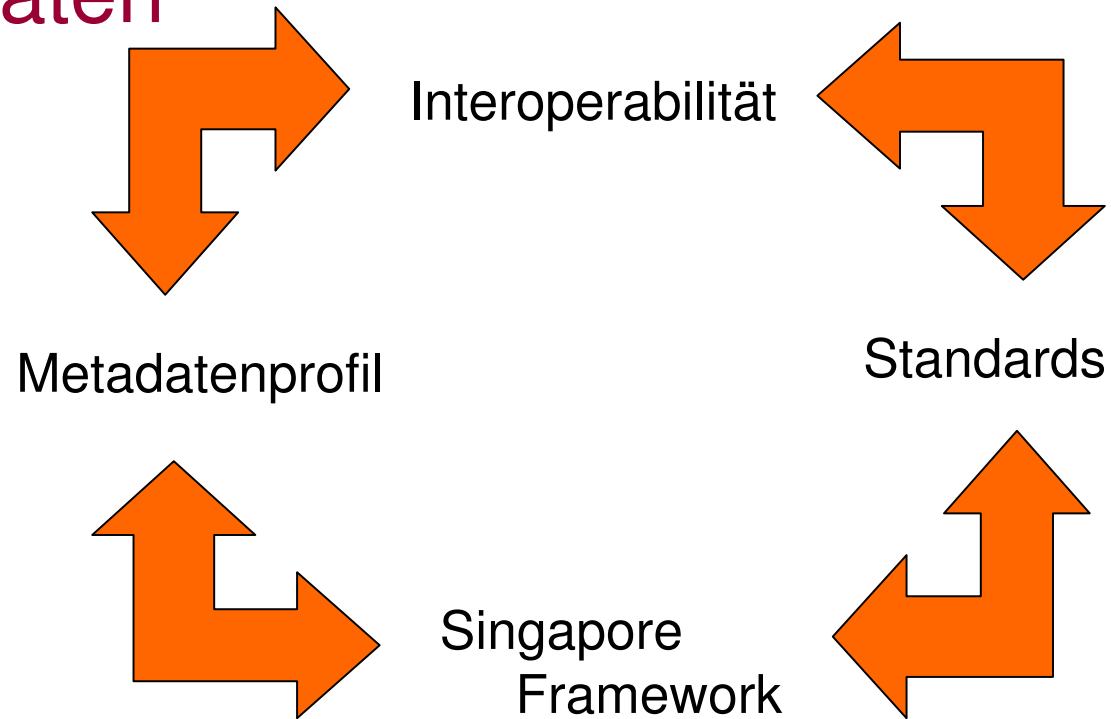
Stefanie Rühle
SUB Göttingen





Überblick

Metadaten





Metadaten – wozu?

- **Objekte beschreiben**
 - strukturiert
 - einheitlich
- **ermöglicht**
 - Finden
 - Identifizieren
 - Selektieren
 - Nutzung

von für den Nutzer relevanten Objekten
- **verbessert den Austausch von Informationen**



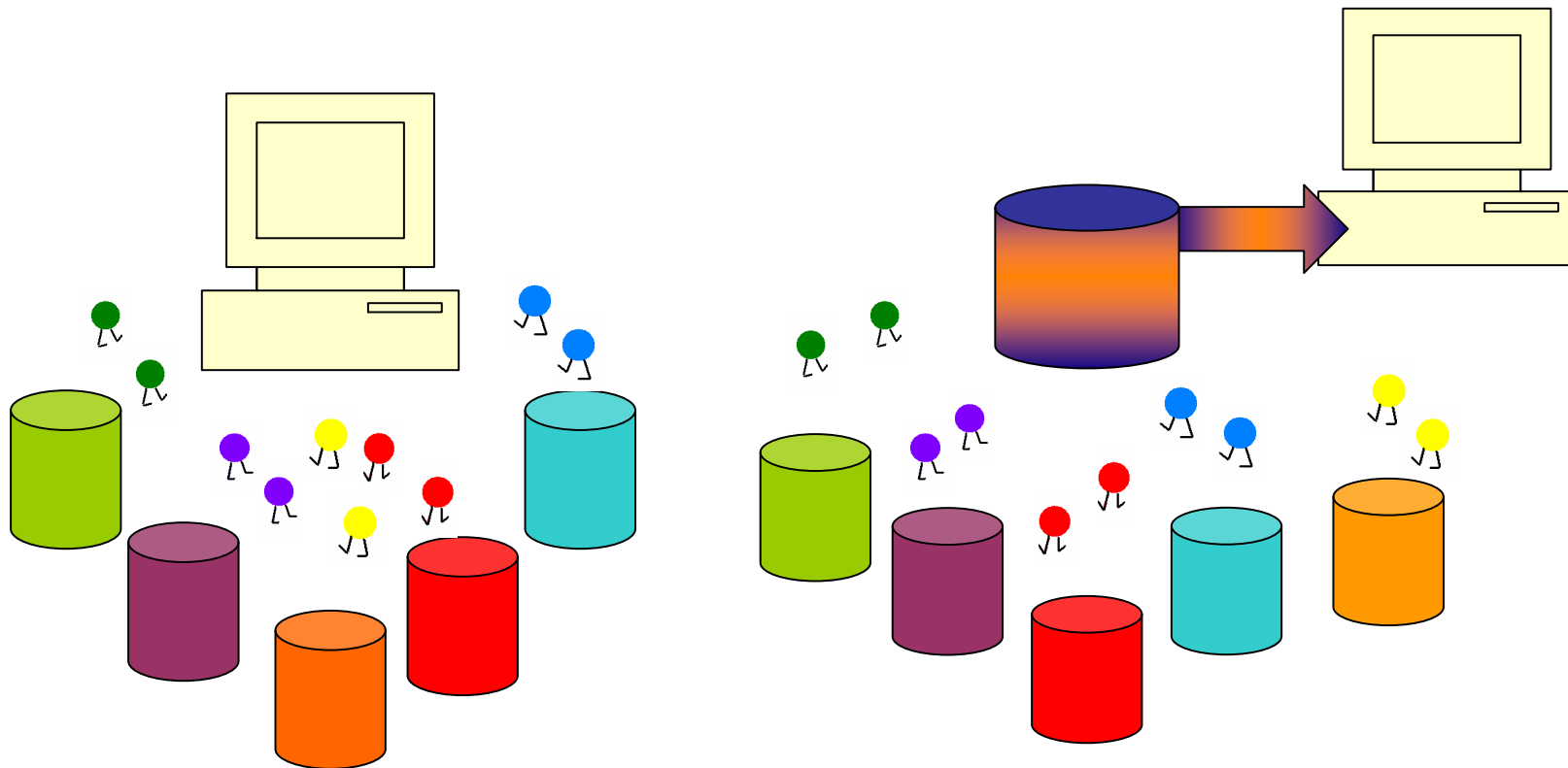
Metadaten – wo?

- **In Organisationen**
 - Für die Bestandsaufnahme und Organisation von Informationsobjekten
- **Zwischen Organisationen**
 - Um den Austausch und die Nutzbarkeit von Informationen zu erleichtern
- **In Datenportalen**
 - Um Informationen zu suchen, zu finden und zu nutzen



Metadaten lernen “laufen” ...

Suchmaschinen





Interoperabilität

“ ... the ability of two or more systems or components to exchange information and to use the information that has been exchanged.” (<http://en.wikipedia.org/wiki/Interoperability>)

Syntaktische Interoperabilität

- “If a system is capable of communicating and exchanging data, it is syntactic interoperable. For communicating data, specified data formats, communication protocols, interfaces of descriptions and the like are fundamental. In general XML or SQL standards provide syntactic interoperability.” (<http://en.wikipedia.org/wiki/Interoperability>)

Semantische Interoperabilität

- “... is the ability of two or more computer systems to exchange information and have the meaning of that information automatically interpreted by the receiving system accurately enough to produce useful results, as defined by the end users of both systems.” (http://en.wikipedia.org/wiki/Semantic_interoperability)



Interoperabilität

Interoperabilität von Metadatenformaten

- Gemeinsamer Nenner = Metadatenstandards
- Dokumentation = Metadatenprofil



Metadatenstandards

- **sorgen für**
 - Konsistenz der Metadaten
- **verbessern**
 - Metadatenutzung
- **erlauben**
 - Austausch von Metadaten
 - Verlinkung von Objekten

anwendungsübergreifend





Metadatenstandards

"The nice thing about standards is that you have so many to choose from ..."

Quelle: Andrew Tanenbaum, *Computer Networks*, 2nd edn., p. 254

Aber:

- Welcher Standard wird in meiner Community genutzt?
- Welcher Standard passt zu meiner Anwendung / in meinem Kontext?
- Welcher Standard entspricht am ehesten meinen Anforderungen?



Interoperabilität

Interoperabilität von Metadatenformaten

- Gemeinsamer Nenner = Metadatenstandards
- Dokumentation = Metadatenprofil



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?**
 - Metadatenterme, 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

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.itg.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

- Gründe für Dokumentation in Form eines Metadatenprofils:
 - Sichtbarkeit
 - Nachhaltigkeit
 - Kosten sparen
- Gründe für die Begutachtung/Zertifizierung von Metadatenprofilen
 - Interoperabilität von Metadatenanwendungen unterstützen
 - Nutzung von Metadaten außerhalb lokaler Anwendungen erleichtern
 - Wiederverwertbarkeit von Metadaten möglich machen
- Evaluierungsrichtlinie
 - Systematisiert die Begutachtung und macht transparent, welche Anforderungen KIM an ein Metadatenprofil stellt
 - Basiert auf dem Singapore Framework = Standard für die Dokumentation von Metadatenprofilen



Singapore Framework

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

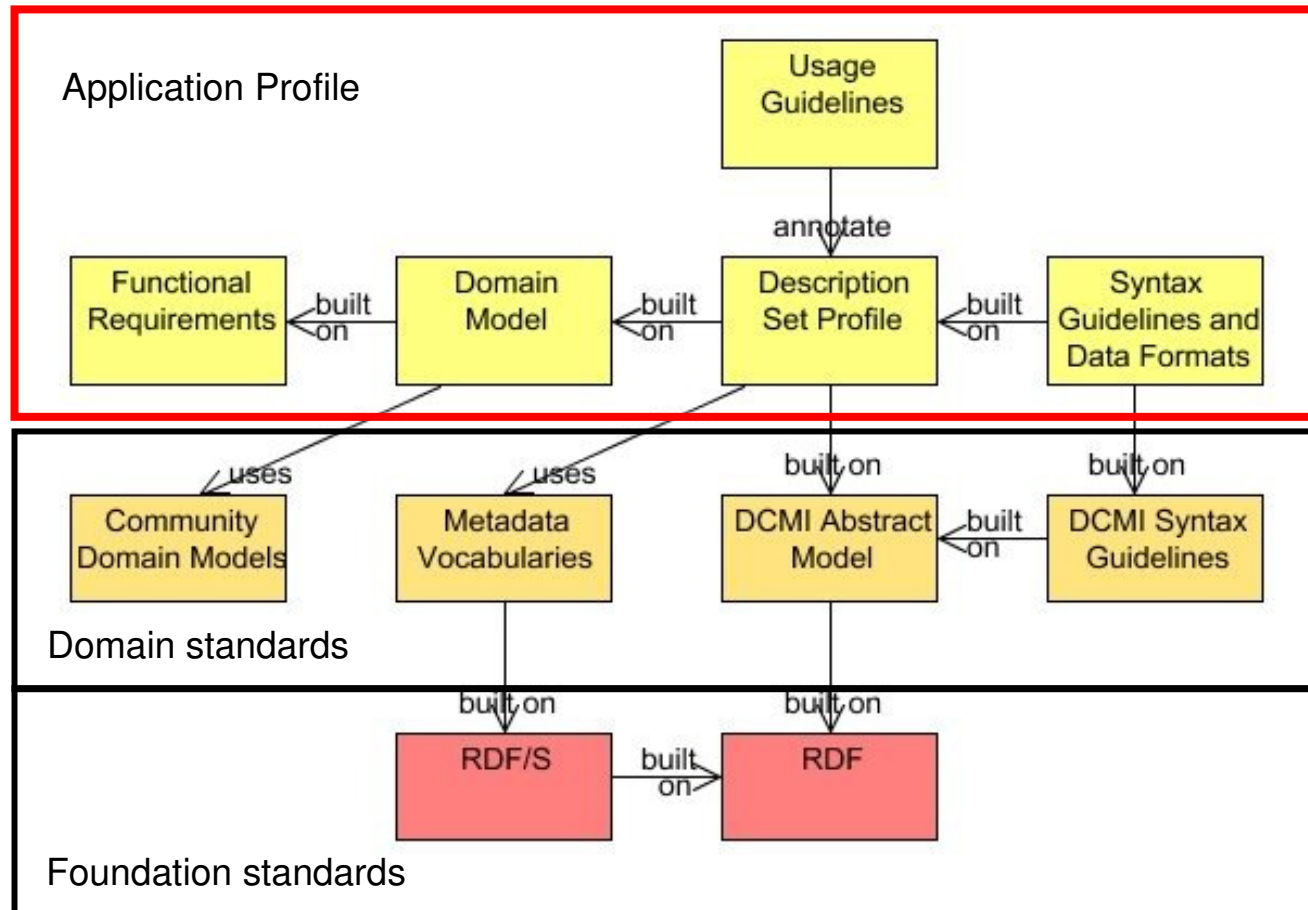
„...a framework for designing metadata applications for maximum interoperability and for documenting such applications for maximum reusability.“

- **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 – Zertifizierungskriterien

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

According to the Singapore Framework, a Dublin Core Application Profile is packet of documentation which consists of the following components:

- “Functional requirements“
 - „The functional requirements of a Dublin Core Application Profile describe the functions that the application profile is designed to support, as well as functions that are out of scope.“

http://www.kim-forum.org/material/pdf/zertifizierungsrichtlinien_20101503.pdf

Ein Metadatenprofil setzt sich aus den folgenden Komponenten zusammen:

- (Für wen?) Ein Metadatenprofil muss Zielsetzung und Geltungsbereich des Profils beschreiben,
 - Indem es den Kontext nennt, in dem das Profil verwendet wird.
- (Wozu?) Ein Metadatenprofil muss die funktionellen Anforderungen beschreiben,
 - die die Anwendung in Hinblick auf Benutzerbedürfnisse erfüllen soll.





Singapore Framework – Zertifizierungskriterien

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

According to the Singapore Framework, a Dublin Core Application Profile is packet of documentation which consists of the following components:

- “Domain model“
 - „The domain model defines the basic entities described by the application profile and their fundamental relationships.“

http://www.kim-forum.org/material/pdf/zertifizierungsrichtlinien_20101503.pdf

Ein Metadatenprofil setzt sich aus den folgenden Komponenten zusammen:

- (Was?) Ein Metadatenprofil muss ein Datenmodell (Modell der beschriebenen Entitäten) bereitstellen,
 - Das die zu beschreibenden Objekte und die Beziehungen zwischen diesen Objekten beschreibt





Singapore Framework – Zertifizierungskriterien

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

According to the Singapore Framework, a Dublin Core Application Profile is packet of documentation which consists of the following components:

- “Description Set Profile“
 - „A Description Set Profile (DSP) defines a set of metadata records that are valid instances of an application profile ... The Dublin Core Description Set Profile model is designed to offer a simple constraint language for Dublin Core metadata, based on the DCMI Abstract Model. A DSP constrains the resources that may be described by descriptions in a description set conforming to the application profile, the properties that may be used, and the ways a value may be referenced.“

http://www.kim-forum.org/material/pdf/zertifizierungsrichtlinien_20101503.pdf

Ein Metadatenprofil setzt sich aus den folgenden Komponenten zusammen:

- (Womit?) Ein Metadatenprofil muss ein Beschreibungssatzprofil bereitstellen,
 - Das ...





Was ist ein Beschreibungssatzprofil?

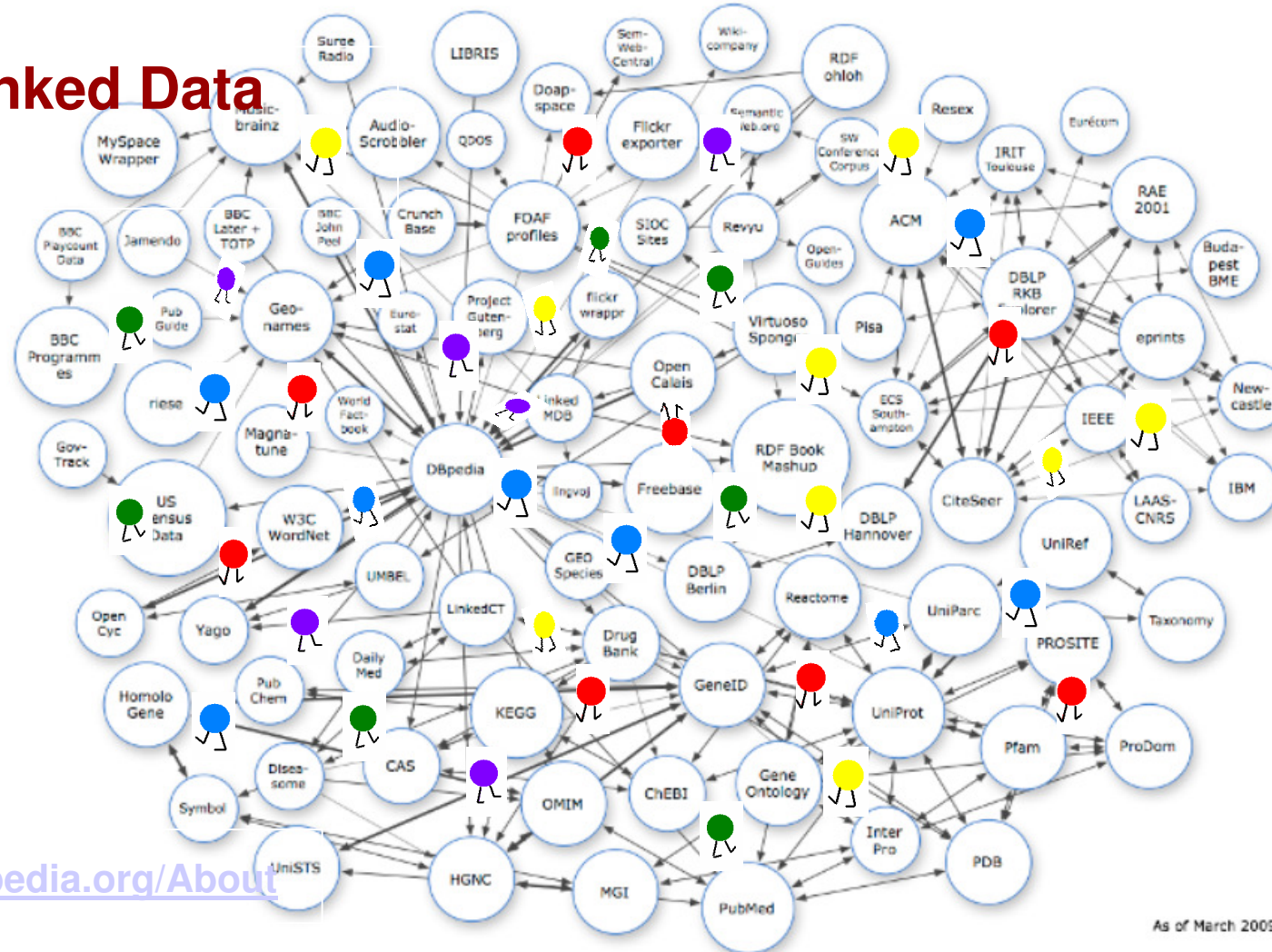
Und wozu brauche ich das überhaupt?





Unser Ziel ... das semantische Netz

Linked Data



<http://dbpedia.org/About>

As of March 2009





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**





“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.





... 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://dublincor
e.org/pages/home">

<dcds:statement
  dcds:propertyURI="http://purl.org/d
c/terms/title">

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

<dcds:statement
  dcds:propertyURI="http://purl.org/d
c/terms/publisher"
  dcds:valueURI="http://example.org/
agents/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
a#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>
```



Zertifizierungsrichtlinie

- **Das Beschreibungssatzprofil (DSP) muss für jede Klasse von Objekten, die im Datenmodell genannt werden, ein eigenes Beschreibungsschema enthalten**
 - z. B. ein Beschreibungsschema für Bücher, ein anderes für Personen, ein anderes wieder für Homepages usw
- **Ein Beschreibungsschema muss aus mindestens einem Aussagenschema (statement template) bestehen**
- **Das Aussagenschema muss die Terme nennen und beschreiben, die für die Beschreibung einer Klasse von Objekten verwendet werden**
- **Den Termen müssen korrekte URIs zugewiesen sein.**
 - Will ich bei der Beschreibung einer Klasse von Objekten die Eigenschaft „Datum“ berücksichtigen, so muss ich ein Aussagenschema entwickeln, in dem ich diese Eigenschaft definiere und mit einer URI identifiziere.
 - Wohldefinierte Terme, damit Mensch und Maschine wissen, was gemeint ist
 - Terme aus bereits vorhandenen Standards, verbessert die Interoperabilität auf der semantischen Ebene und fördert so die Nachhaltigkeit und verringert die Kosten bei Datenaustausch.





Vielen Dank

Beispiel folgt!

