



**OPEN
DATA
SUPPORT**

Trainingsmodul 1.4

Einführung in das Metadaten- Management



*Die Mitglieder des PwC Netzwerks unterstützen Organisationen und Individuen dabei, die Werte zu schaffen, nach denen sie suchen. Wir sind ein Netzwerk von Unternehmen mit nahezu 180.000 Mitarbeitern in 158 Ländern, die sich dazu verpflichtet fühlen Qualität in den Bereichen Assurance, Tax & Legal sowie Advisory zu liefern. Sagen Sie uns, was Ihnen wichtig ist und besuchen Sie für weitere Informationen unsere Webseite www.pwc.com
PwC bezieht sich auf das PwC Netzwerk und/oder eine oder mehrere Mitgliedsfirmen, von denen jede ein rechtlich selbstständiges Unternehmen ist. Besuchen Sie unsere Webseite www.pwc.com/structure für weitere Details.*

Präsentation Metadaten

Open Data Support wird von der Europäischen Kommission finanziert, gemäß SMART 2012/O107 'Lot 2: Provision of services for the Publication, Access and Reuse of Open Public Data across the European Union, through existing open data portals'(Vertrag No. 30-CE-0530965/00-17).

© 2013 European Commission

Diese Präsentation wurde von PwC erstellt

Autoren:

**Makx Dekkers, Michiel De Keyzer, Nikolaos Loutas
und Stijn Goedertier**

Disclaimers

1. Die Ansichten, die in dieser Präsentation vertreten werden, spiegeln einzig die Meinung des Autors wider und dürfen unter keinen Umständen als offizielle Position der Europäischen Kommission interpretiert werden.

Die Europäische Kommission übernimmt weder eine Garantie für die Genauigkeit der Informationen, die Inhalt dieser Präsentation sind, noch akzeptiert sie jegliche Verantwortung für die Nutzung der selbigen.

Referenzen innerhalb dieser Präsentation zu spezifischen Produkten, Spezifikationen, Prozessen oder Services durch Handelsnamen, Markenzeichen, Hersteller oder ähnliches, implizieren nicht unbedingt deren Unterstützung oder Bevorzugung durch die Europäische Kommission.

Der Autor hat sämtliche Anstrengungen unternommen, um sicherzustellen, dass er/sie, wo nötig, die Erlaubnis erhalten hat, die einzelnen Teile der in dieser Präsentation genutzten Manuskripte zu benutzen. Dies beinhaltet das Einholen einer Nutzungserlaubnis von den Lizenzinhabern oder deren gesetzlichen Vertretern zur Nutzung von Illustrationen, Karten, Schaubildern, die durch Rechte des geistigen Eigentums geschützt sind.

2. Diese Präsentation wurde sorgfältig von PwC zusammengestellt, jedoch gibt PwC keine Erklärung darüber ab und übernimmt keine Garantie dafür (weder ausdrücklich noch implizit), dass die Informationen in dieser Präsentation vollständig oder akkurat sind. PwC ist nicht für die Informationen in dieser Präsentation oder jedwede Entscheidung oder Konsequenz, die aus ihr resultieren, haftbar zu machen. PwC haftet nicht für etwaige Schäden, die durch die Nutzung der Informationen in dieser Präsentation entstehen sollten. Die Informationen in dieser Präsentation sind genereller Natur und einzig dafür bestimmt, eine Orientierungshilfe für Themen allgemeinen Interesses zu sein. Diese Präsentation ist kein Ersatz für professionelle Beratung zu jedwedem speziellen Thema. Kein Leser sollte auf der Basis der Themendarstellungen in dieser Präsentation handeln, ohne vorher angemessene professionelle Beratung einzuholen.

Lernziele

Am Ende dieses Trainingsmoduls sollten Sie verstehen:

- was Metadaten sind
- die Terminologie und Grundsätze der Metadatenverwaltung
- die verschiedenen Dimensionen der Qualität von Metadaten
- die Nutzung von kontrollierten Vokabularien für Metadaten
- Austausch und Erstellung von Metadaten
- Metadatenverwaltung im Open Data Support

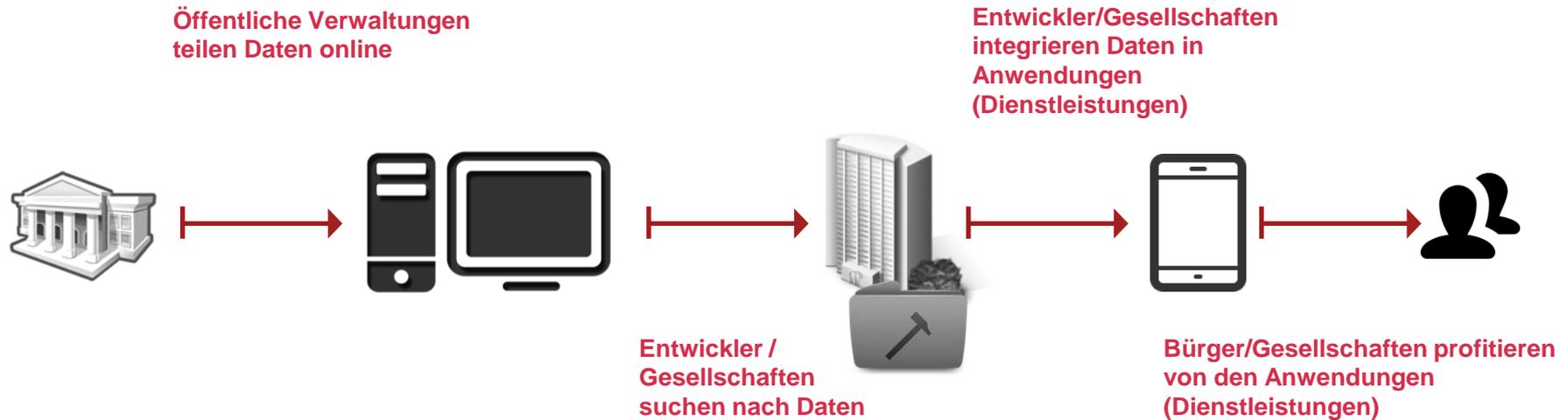
Inhalt

Dieses Modul enthält:

- eine Erklärung, was Metadaten sind
- einen Überblick über den Metadaten-Lebenszyklus
- eine Einführung in die Qualität von Metadaten
- einen Überblick über die Metadatenverwaltung und die Herangehensweise beim Austausch durch die Anwendung von Open Data Support auf der Open Data Interoperability Platform

Open Data hat großes Potenzial, um soziale und wirtschaftliche Werte zu schaffen

Veröffentlichung von Daten

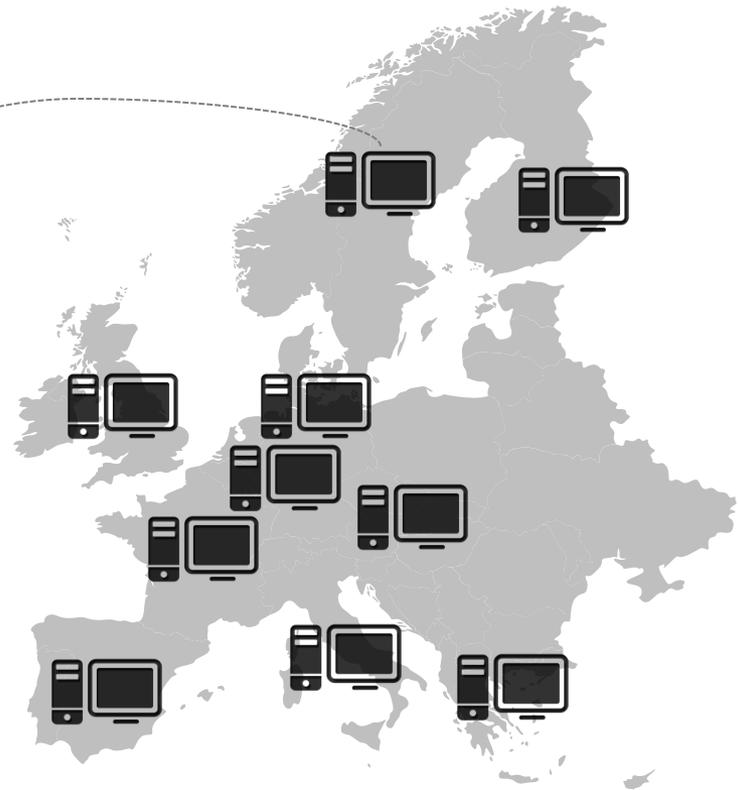


Weiterverwendung von Daten

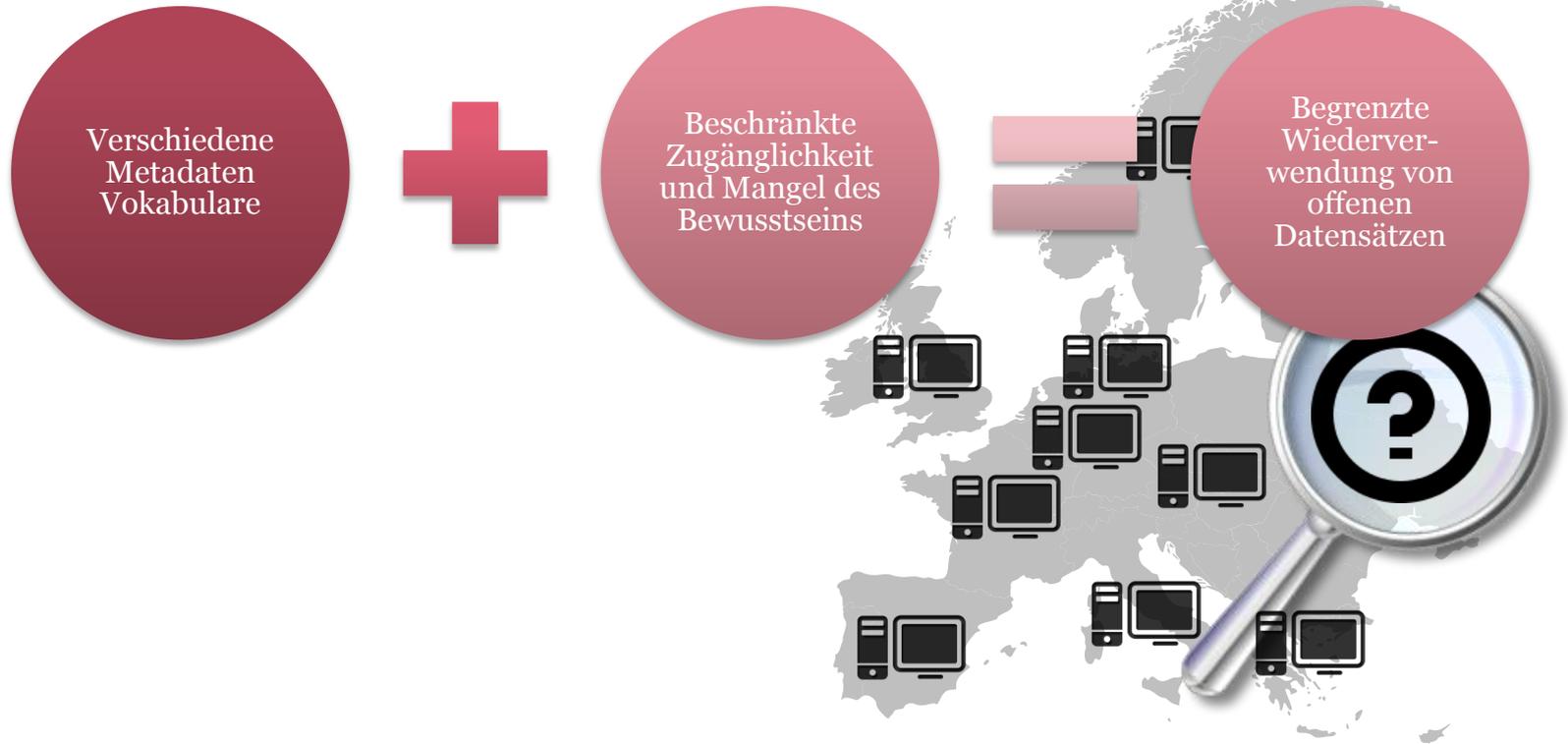
Es gibt mehr als 150 Portale in Europa, die Open Government Data sammeln

Existierende OGD Portale

150+



Beschränkte Zugänglichkeit und Fehlen von einem (grenzüberschreitendem/branchenspezifischem) Bewusstsein für offene Datensätze



Keine Weiterverwendung = kein sozialer und wirtschaftlicher Wert



Was sind Metadaten?

Definition, Beispiele und Standards der Wiederverwendung

Was sind Metadaten?

“Metadaten sind strukturierte Informationen, die beschreiben, erklären, lokalisieren, oder es anderweitig einfacher machen, eine Informationsquelle abzurufen, zu verwenden oder zu verwalten. Metadaten werden oft Daten zu bestimmten Daten oder Informationen zu bestimmten Informationen genannt.”

-- National Information Standards Organization

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Metadaten liefern die nötigen Informationen dafür, dass **Daten** (z.B. Dokumente, Bilder, Datensätze), **Konzepte** (z.B. Klassifikationen) und **reale Begebenheiten** (z.B. Personen, Organisationen, Standorte, Bilder, Produkte) zusammenpassen.

Arten von Metadaten

- **Beschreibende Metadaten** beschreiben eine Ressource zum Zwecke der Entdeckung und der Identifizierung
- **Strukturelle Metadaten**, z.B. Datenmodelle und Referenzdaten
- **Administrative Metadaten** bieten Informationen zur Verwaltung einer Ressource

In diesem Tutorial werden wir uns vor allem auf beschreibende Metadaten für Datensätze konzentrieren.

Auch administrative Metadaten werden teilweise beschrieben.

Beispiele für Metadaten

Etikett

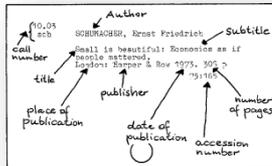


liefert Metadaten für



Dose

Katalogzettel



Buch

Datensetbeschreibung (DCAT)

```
:weather1-7 a dcat:Dataset ;
dct:title "Measurements from weather stations 1-7" ;
dct:description "Data from seven weather stations
showing temperature, humidity,
wind direction and wind speed" ;
dct:modified "2013-07-01" ;
dct:publisher <http://myweather.com/id/myweather> ;
dcat:keyword "weather" ;
dcat:landingpage <http://myweather.com/stations1-7.html> ;
dcat:distribution :weatherdata-xlsx
.

:weatherdata1-7-xlsx a dcat:Distribution ;
dct:format <http://publications.europa.eu/resource/authority/file-type/XLSX> ;
dct:licence <http://creativecommons.org/licenses/by/4.0/> ;
dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```

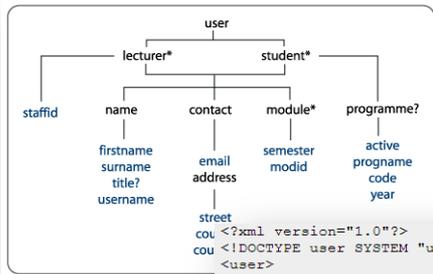


| | Temp. °C | Humidity % | Wind direction | Wind speed km/h |
|-----------|----------|------------|----------------|-----------------|
| Station 1 | 18.1 | 60 | WSW | 18 |
| Station 2 | 17.5 | 59 | WSW | 20 |
| Station 3 | 18.2 | 55 | SW | 22 |
| Station 4 | 19.0 | 62 | SW | 18 |
| Station 5 | 18.0 | 65 | WSW | 19 |
| Station 6 | 18.2 | 63 | SSW | 21 |
| Station 7 | 17.9 | 61 | SW | 22 |

Datenset

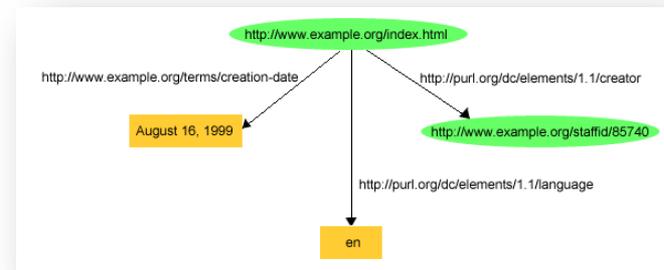
Zwei Ansätze, um Metadaten im Web zur Verfügung zu stellen

XML (Baumstruktur- / Containerstruktur-Ansatz)



```
<?xml version="1.0"?>
<!DOCTYPE user SYSTEM "users.dtd">
<user>
  <student>
    <name>
      <firstname>Joe</firstname>
      <surname>Smith</surname>
      <title>Mr.</title>
      <username>smithj</username>
    </name>
    <contact>
      <address>
        <street>54 Maple Rise, Santry</street>
        <country>Dublin</country>
        <country>Ireland</country>
      </address>
      <email>smithj@dcu.ie</email>
    </contact>
    <programme active="true">
      <progrname>M.Eng in Electronic Systems</progrname>
      <code>MEN</code>
      <year>1</year>
    </programme>
    <module semester="2">
      <modid>EE557</modid>
    </module>
    <module semester="1">
      <modid>EE553</modid>
    </module>
  </student>
```

RDF (Triple-basierter Ansatz)



```
ex:index.html dc:creator exstaff:85740 .
ex:index.html exterms:creation-date "August 16, 1999" .
ex:index.html dc:language "en" .
```

Verwalten Sie die
Metadaten Ihrer
Datensätze

Metadaten-Management ist wichtig

Metadaten müssen verwaltet werden, um Folgendes sicherzustellen ...

- **Verfügbarkeit:** Metadaten müssen dort gespeichert werden, wo Zugriff und Registrierung möglich sind, damit man sie finden kann
- **Qualität:** Metadaten müssen von gleichbleibender Qualität sein, so dass die Benutzer wissen, dass sie diesen vertrauen können
- **Persistenz:** Metadaten müssen über lange Zeiträume gespeichert werden
- **Offene Lizenz:** Metadaten sollten unter einer öffentlichen Domain-Lizenz zur Verfügung stehen, um ihre Weiterverwendungen zu ermöglichen

*Der Metadaten-Lebenszyklus ist **größer** als der Lebenszyklus der Daten:*

- **Metadaten können erstellt werden, bevor Daten erstellt oder erfasst werden**, z.B. um über Daten zu informieren, die in der Zukunft verfügbar sein werden
- **Metadaten müssen aufbewahrt werden, nachdem Daten entfernt worden sind**, z.B. um über Daten zu informieren, die außer Betrieb oder zurückgenommen wurden

Weiterverwendung von vorhandenen Vokabeln, um Ihre Ressourcen mit Metadaten zu versorgen

Universelle Standards und Spezifikationen:

- **Dublin Core** für veröffentlichtes Material (Text, Bild), <http://dublincore.org/documents/dcmi-terms/>
- **FOAF** für Personen und Organisationen, <http://xmlns.com/foaf/spec/>
- **SKOS** für Konzeptsammlungen, <http://www.w3.org/TR/skos-reference>
- **ADMS** für “Interoperability Assets”, <http://www.w3.org/TR/vocab-adms/>

Spezifischer Standard für Datensätze:

- **Data Catalog Vocabulary DCAT**, <http://www.w3.org/TR/vocab-dcat/>

Spezifische Verwendung von DCAT und anderen Vokabularien, um die Interoperabilität von Datenportalen in ganz Europa zu unterstützen:

- **DCAT application profile for data portals in Europe**, http://joinup.ec.europa.eu/asset/dcat_application_profile/description

Entwerfen Sie Ihr Metadaten-Schema mit RDF Schema (RDFS) – wenn möglich wiederverwendbar

RDF Schema ist besonders gut geeignet für die Kombination von Begriffen aus verschiedenen Standards und Spezifikationen

Erfinden Sie keine Begriffe neu, die schon woanders definiert werden. Wenn sie RDF-Schemata gestalten, **verwenden Sie** Begriffe möglichst wieder

- Zum Beispiel verwendet das Anwendungsprofil DCAT für Datenportale in Europa (DCAT-AP) Begriffe aus DCAT, Dublin Core, FOAF, SKOS, ADMS und anderen Quellen wieder

7.3. Dataset

7.3.1. Mandatory properties for Dataset

| Property | URI | Range | Usage note | Card |
|-------------|---------------------------------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|------|
| description | dct:description | rdfs:Literal | This property contains a free-text account of the dataset. This property can be repeated for parallel language versions of the description. | 1..n |
| publisher | dct:publisher | foaf:Organization | This property refers to an organisation responsible for making the dataset available. | 1..1 |
| title | dct:title | rdfs:Literal | This property contains a name given to the dataset. This property can be repeated for parallel language versions of the name. | 1..n |

7.3.2. Recommended properties for Dataset

| Property | URI | Range | Usage note | Card |
|----------------------|----------------------------------------------|------------------------------|--------------------------------------------------------------------------------------------------------------------|------|
| contact point | adms:contactPoint | v:VCard | This property contains contact information that can be used for flagging errors in the dataset or sending comments | 0..n |
| dataset distribution | dcat:distribution | dcat:Distribution | This property links the dataset to an available distribution. | 0..n |
| theme/ category | dcat:theme, subproperty of dct:subject | skos:Concept | This property refers to a category of the dataset. A dataset can have multiple themes. | 0..n |

7.3.3. Optional properties for Dataset

| Property | URI | Range | Usage note | Card |
|-----------------|------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------|------|
| frequency | dct:accrualPeriodicity | dct:Frequency | This property refers to the frequency at which dataset is published. | 0..1 |
| identifier | dct:identifier | rdfs:Literal | This property contains the main identifier for the dataset, e.g. the URI or other unique identifier in the context of the Catalog | 0..n |
| keyword/ tag | dcat:keyword | rdfs:Literal | This property contains a keyword or tag describing the dataset. | 0..n |

Beispiel: Beschreibung eines offenen Datensatzes mit DCAT-AP

Beschreibung des Katalogs

```
:catalog
  a dcat:Catalog ;
  dct:title "Imaginary Catalog" ;
  rdfs:label "Imaginary Catalog" ;
  foaf:homepage <http://example.org/catalog> ;
  dct:publisher :transparency-office ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dcat:dataset :dataset-001 , :dataset-002 , :dataset-003 ;
  .
```

Beschreibung des Datensatzes

```
:dataset-001
  a dcat:Dataset ;
  dct:title "Imaginary dataset" ;
  dcat:keyword "accountability","transparency" ,"payments" ;
  dct:issued "2011-12-05"^^xsd:date ;
  dct:modified "2011-12-05"^^xsd:date ;
  dct:publisher :finance-ministry ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dcat:distribution :dataset-001-csv ;
  .
```

Beschreibung der Verteilung

```
:dataset-001-csv
  a dcat:Distribution ;
  dcat:downloadURL <http://www.example.org/files/001.csv> ;
  dct:title "CSV distribution of imaginary dataset 001" ;
  dcat:mediaType "text/csv" ;
  dcat:byteSize "5120"^^xsd:decimal ;
  .
```

Kontrollierte Vokabularien

Thesauren, Taxonomien und standardisierte Listen von Begriffen können für die Zuweisung von Werten zu Metadateneigenschaften verwendet werden

Was sind kontrollierte Vokabularien?

Ein kontrolliertes Vokabular ist eine vordefinierte Liste von Werten, die als Werte für eine bestimmte Eigenschaft in Ihrem Metadaten-Schema verwendet werden

- Zusätzlich zur sorgfältigen Gestaltung von Schemata sind die Wertebereiche von Metadateneigenschaften wichtig für den Informationsaustausch und somit für die Interoperabilität
- Gemeinsame kontrollierte Vokabularien für Wertebereiche machen Metadaten systemübergreifend verständlich

Welches kontrollierte Vokabular eignet sich für welche Objektart?

- Verwendung der **Codelisten** als kontrolliertes Vokabular für freien Text oder “String” Eigenschaften
- Verwenden Sie **Konzepte**, die von einem **URI** in Bezug auf “Dinge” **identifiziert** werden

Beispiel DCAT-AP Eigenschaft:

| | | | |
|--------------|--------------|--------------|-----------------------------------------------------------------|
| keyword/ tag | dcat:keyword | rdfs:Literal | This property contains a keyword or tag describing the dataset. |
|--------------|--------------|--------------|-----------------------------------------------------------------|

Beispiel DCAT-AP Eigenschaft:

| | | | |
|--------------------|----------------------------------------------|--------------|----------------------------------------------------|
| theme/ category | dcat:theme, subproperty of dct:subject | skos:Concept | This property refers to a category of the dataset. |
|--------------------|----------------------------------------------|--------------|----------------------------------------------------|

Beispiel Codelist - ObjectInCrimeClass (ListPoint):

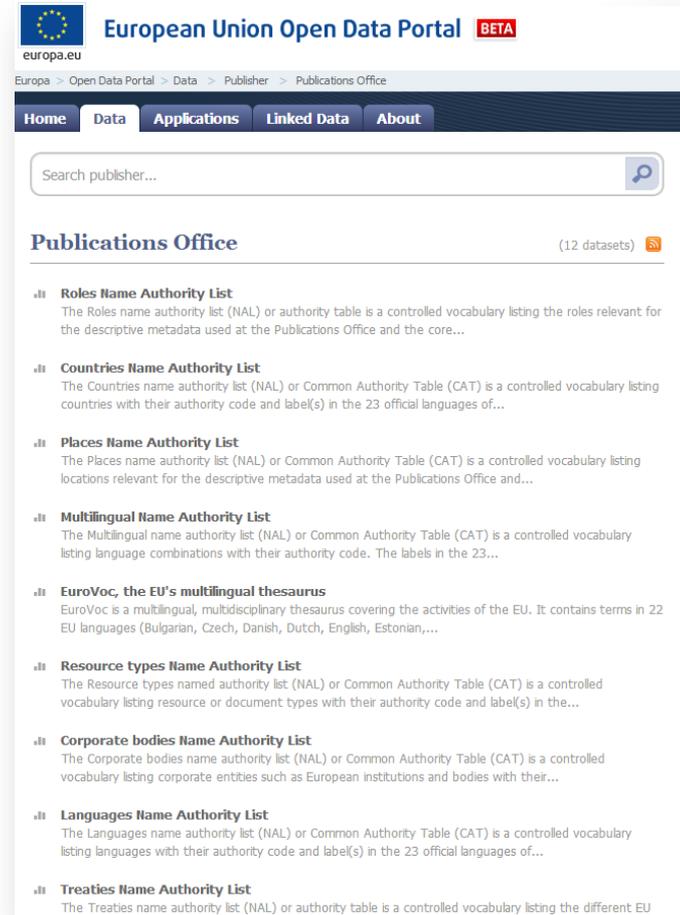
| Code | Meaning |
|------|---------------------------------------|
| 1 | Used |
| 2 | Taken Without Consent |
| 3 | Eliminated |
| 4 | Not involved |
| 5 | Involved |
| 6 | Crime Weapon |
| 7 | Handled / Received |
| 11 | Exhibit |

Beispiel Taxonomie mit Begriffen, die ein URI – EuroVoc – haben:

```
<rdf:Description rdf:about="http://eurovoc.europa.eu/300">
  <xl:altLabel rdf:resource="http://eurovoc.europa.eu/415040"/>
  <s04:prefLabel xml:lang="da">international kredit</s04:prefLabel>
  <s04:prefLabel xml:lang="sv">internationell kredit</s04:prefLabel>
  <s04:prefLabel xml:lang="en">international credit</s04:prefLabel>
  <s04:prefLabel xml:lang="de">internationaler kredit</s04:prefLabel>
  <s04:prefLabel xml:lang="nl">international krediet</s04:prefLabel>
</rdf:Description>
```

Beispiel – Auswahllisten (Named Authority Lists) des Amts für Veröffentlichungen

- Die Named Authority Lists bieten wiederverwendbare kontrollierte Vokabularien für:
 - Länder
 - Unternehmen
 - Dateitypen
 - interinstitutionelle Verfahren
 - Sprachen
 - Mehrsprachigkeit
 - Ressourcentypen
 - Rollen
 - Verträge



The screenshot shows the European Union Open Data Portal (BETA) interface. The page title is "Publications Office" with a sub-header "(12 datasets)". Below the title, there is a list of Named Authority Lists (NALs) with brief descriptions:

- Roles Name Authority List**: The Roles name authority list (NAL) or authority table is a controlled vocabulary listing the roles relevant for the descriptive metadata used at the Publications Office and the core...
- Countries Name Authority List**: The Countries name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing countries with their authority code and label(s) in the 23 official languages of...
- Places Name Authority List**: The Places name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing locations relevant for the descriptive metadata used at the Publications Office and...
- Multilingual Name Authority List**: The Multilingual name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing language combinations with their authority code. The labels in the 23...
- EuroVoc, the EU's multilingual thesaurus**: EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU. It contains terms in 22 EU languages (Bulgarian, Czech, Danish, Dutch, English, Estonian,...
- Resource types Name Authority List**: The Resource types named authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing resource or document types with their authority code and label(s) in the...
- Corporate bodies Name Authority List**: The Corporate bodies name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing corporate entities such as European institutions and bodies with their...
- Languages Name Authority List**: The Languages name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing languages with their authority code and label(s) in the 23 official languages of...
- Treaties Name Authority List**: The Treaties name authority list (NAL) or authority table is a controlled vocabulary listing the different EU

Der Metadaten- Lebenszyklus

*Erstellung, Pflege, Aktualisierung, Speicherung,
Veröffentlichung von Metadaten und der Umgang mit
Datenlöschung*

Das Erstellen Ihrer Metadaten

Metadatenerstellung kann durch (semi-)automatische Prozesse unterstützt werden

- *Dokumenteigenschaften in (Office)-Tools generiert, z.B. Erstellungsdatum*
- *Räumliche und zeitliche Information erfasst von Kameras, Sensoren ...*
- *Information aus Publikationsworkflows z.B. der Speicherort der Datei oder URL*

Allerdings erfordern andere Eigenschaften menschliche Eingriffe

- Was ist die Ressource (z.B. Verknüpfung zu einem Fachvokabular)?
- Wie kann die Ressource verwendet werden (z.B. Verknüpfung zu einer Lizenz)?
- Wo kann ich weitere Informationen über diese Ressource finden (z.B. Verknüpfung zu einer Website oder Dokumentation, die die Ressource beschreibt)?
- Wie können hochwertige Informationen eingeschlossen werden?

Die Pflege Ihrer Metadaten

Ansätze für die Pflege von Metadaten müssen für den Datentyp, der veröffentlicht wird, geeignet sein

- Wenn die **Daten** sich nicht **ändern**, können **Metadaten** relativ **stabil** sein. Änderungen (Massenumwandlung) können offline stattfinden, wenn nötig
- Wenn **Daten** sich **häufig ändern** (z.B. Echtzeitsensordaten), müssen **Metadaten** mit dem Datenworkflow eng gekoppelt und **Änderungen** praktisch **unmittelbar** sein

Das Aktualisieren Ihrer Metadaten – Änderungen planen

Metadaten stehen in einem globalen Kontext, der ständig Änderungen ausgesetzt ist!

- **Organisation** – Abteilungen werden gebildet, mit anderen zusammengeführt, Verantwortlichkeiten werden weitergegeben
- **Die Verwendung der Daten** – rund um die Daten entstehen neue Anwendungen
- **Referenzdaten** – kontrollierte Vokabularien werden entwickelt und miteinander verknüpft
- **Datenstandards und Technologien** – Der Lebenszyklus von Technologien wird immer kürzer; wie wird das Web von morgen sein?
- **Werkzeuge und Systeme** – die Entwicklung von Speicher, Bandbreite, Mobile Systeme...

Metadaten sollten so aktuell wie möglich gehalten werden, wobei die verfügbare Zeit und das Budget berücksichtigt werden sollten

Speichern Sie Ihre Metadaten – was sind die Optionen?

Je nach Betriebserfordernissen können Metadaten in die Daten eingebettet oder gesondert von den Daten gespeichert werden

- Die Einbettung der Metadaten in die Daten (z.B. Office Dokumente, MP3, JPG, RDF Daten) macht den Datenaustausch leichter
- Die Trennung der Metadaten von den Daten (z.B. in einer Datenbank) mit Links zu den entsprechenden Dateien macht die Verwaltung einfacher

Metadaten können in einer **“klassischen” relationalen Datenbank** oder einem **RDF Triple Speicher** gespeichert werden, dies hängt von der Verfügbarkeit von Werkzeugen und den Anforderungen an Leistung und Kapazität ab

Das Löschen von Daten

In vielen Fällen müssen Metadaten sogar nach der Löschung der Daten, die sie beschreiben, erhalten bleiben

Stilllegung oder Löschung von Daten geschieht zum Beispiel:

- Wenn Daten nicht mehr erforderlich sind
- Wenn Daten nicht mehr gültig sind
- Wenn Daten falsch sind
- Wenn Daten vom Eigentümer/Herausgeber abgezogen werden

In diesem Fall sollten die Metadaten **Informationen enthalten**, dass die Daten **gelöscht** wurden, und sofern diese **archiviert** wurden, wie und wo eine **Archivierungskopie angefordert** werden kann

Veröffentlichung von Metadaten – Welche Optionen haben Sie?

- ‘Offene’ Veröffentlichung: direkter Zugang auf URIs:
 - Dies ist die Option, die am ehesten im Einklang mit der Vision von Linked Open Data steht. Sie ermöglicht das “der-Nase-nach”-Prinzip
- Machen Sie Ihre Metadaten durch einen **SPARQL-Endpunkt** verfügbar:
 - Dies ermöglicht, dass externe Systeme Anfragen zu einem RDF Tripel Speicher senden
 - Dies erfordert Kenntnisse über das in dem Tripel Speicher verwendete Schema
- Verzögerte Veröffentlichung: Zugang zu exportierten Dateien in RDF
 - Erzeugt durch die Umwandlung von Nicht-RDF-Daten in RDF-Daten
 - Ermöglicht das offline Herunterladen von Großvolumen und das Zwischenspeichern von Datensammlungen
 - Ermöglicht die Einführung einer Zugriffskontrolle

Siehe auch: <http://www.slideshare.net/OpenDataSupport/licence-your-data-metadata>

Metadaten Qualität

Die Qualität und Vollständigkeit der Metadatenbeschreibung Ihrer Datensätze haben einen direkten Einfluss auf ihre Auffindbarkeit und Weiterverwendung

Bei der Metadaten Qualität geht es um... (1/3)

- Die **Genauigkeit** Ihrer Metadaten: Werden die Eigenschaften der Ressourcen richtig wiedergegeben?
 - *Wenn man z.B. den richtigen Titel, die richtige Lizenz, den richtigen Herausgeber angibt, ermöglicht dies den Benutzern, diejenigen Ressourcen zu entdecken, die sie brauchen.*
- Die **Verfügbarkeit** Ihrer Metadaten: Kann man jetzt und in Zukunft auf die Metadaten zugreifen?
 - *z.B. sie für das Katalogisieren und Herunterladen verfügbar machen und sie einem regelmäßigen Backup-Prozess unterziehen*
- Die **Vollständigkeit** Ihrer Metadaten: Sind alle relevanten Eigenschaften der Ressource genannt (soweit praktisch und wirtschaftlich möglich und für die Anwendung notwendig)?
 - *z.B. Angabe der Lizenz, die die Weiterverwendung regelt oder das Format der Verteilung, das Filter auf solche Aspekte ermöglicht*

Siehe auch: <http://www.slideshare.net/OpenDataSupport/open-data-quality>

Bei der Metadaten Qualität geht es um... (2/3)

- Die **Konformität** Ihrer Metadaten mit anerkannten Standards: Stimmen die Metadaten mit spezifischen Metadaten-Standards oder einem Anwendungsprofil überein?
 - *z.B. Die Konformität der Beschreibung eines Datensatz mit dem DCAT-AP*
- Die **Konsistenz** Ihrer Metadaten: Enthalten die Daten keine Widersprüche?
 - *Sie sollten z.B. keine mehrfachen und widersprüchlichen Lizenzklärungen für dieselben Daten haben*
- Die **Glaubwürdigkeit** und **Herkunft** Ihrer Metadaten: Beruhen die Metadaten auf vertrauenswürdigen Quellen?
 - *z.B. Verknüpfung zu veröffentlichten Referenzdaten und der Verwaltung von einer soliden Organisation (z.B. dem EU-Amt für Veröffentlichung)*

Bei der Metadaten Qualität geht es um... (3/3)

- Die **Verarbeitbarkeit** der Metadaten: Sind die Metadaten richtig maschinenlesbar?
 - *z.B. Werden Sie die Metadaten eines Datensatzes in RDF und/oder XML, und nicht als Freitext machen?*
- Die **Relevanz** der Metadaten: Enthalten die Metadaten die richtige Menge an Information für die Aufgabe?
 - *Begrenzen sie z.B. die Menge der Informationen, um die Bedürfnisse der Benutzer optimal zu bedienen*
- Die **Aktualität** Ihrer Metadaten: Entsprechen die Metadaten, den tatsächlichen (aktuellen) Eigenschaften der Ressourcen und werden sie früh genug veröffentlicht?
 - *z.B. Zeigen Sie das letzte Änderungsdatum der Ressource an, um so sicherzustellen, dass die Metadaten frisch sind und die Benutzer die neuesten Informationen sehen*

Austausch von Metadaten für Datensätze

*Ordnen Sie Ihre Metadaten einem gemeinsamen
Metadatenvokabular zu und tauschen Sie die Metadaten
über Plattformen aus*

Vereinheitlichung von Metadaten

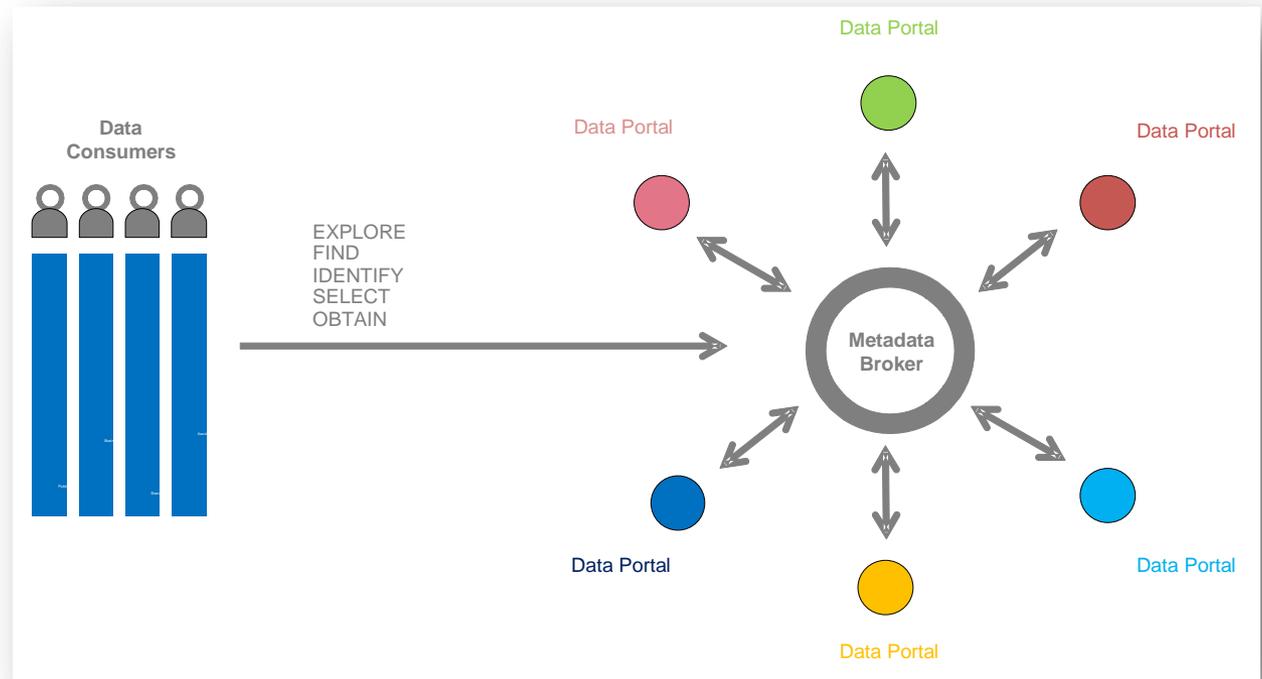
Metadaten sollten nach einem gemeinsamen Schema zugeordnet werden, wenn sie innerhalb von Systemen ausgetauscht werden. Dadurch haben Absender und Empfänger ein gemeinsames Verständnis von der Bedeutung der Metadaten

- Auf der Schema-Ebene können Metadaten, die aus verschiedenen Quellen kommen, **auf verschiedenen Metadaten-Schemata basieren**, z.B. DCAT, schema.org, CERIF, eigenes internes Modell...
- Auf der Daten-(Wert)**Ebene** sollten die Metadateneigenschaften den Werten von **verschiedenen kontrollierten Vokabularien** oder **Syntaxen** zugewiesen werden, z.B.:
 - Sprache: Englisch kann ausgedrückt werden als:
<http://publications.europa.eu/resource/authority/language/ENG> oder als
<http://id.loc.gov/vocabulary/iso639-1/en>
 - Datum: ISO8601 (“20130101”) versus W3C DTF (“2013-01-01”)

Beispiel: Das Homogenisieren von Metadaten zu Datensätzen

Das DCAT Anwendungsprofil für Datenportale in Europa

Das DCAT-AP kann als gemeinsames Modell für den Austausch von Metadaten mit offenen Datenplattformen in Europa verwendet werden und/oder mit einer Datenvermittlung (z.B. die Open Data Interoperability Platform - ODIP)



Siehe auch: http://joinup.ec.europa.eu/asset/dcat_application_profile/home

Abbildungsbeispiel – GovData

dct:publisher

adms:contactPoint

Veröffentlichende Stelle: Statistisches Bundesamt (poststelle@destatis.de)

dct:title (Dataset)



Produktionsindex für das Verarbeitende Gewerbe: Deutschland, Monate, Original- und bereinigte Daten, Wirtschaftszweige (Hauptgruppen und Aggregate)

Produktionsindex für das Verarbeitende Gewerbe: Deutschland, Monate, Original- und bereinigte Daten, Wirtschaftszweige (Hauptgruppen und Aggregate)
Klicken Sie hier für weitere Informationen

dct:description

Links zu den Datendateien:

CSV-Datei der Tabelle '42153-0001'

dct:title (Distribution)

CSV

XLSX-Datei der Tabelle '42153-0001'

XLSX

Link zu CKAN:

<http://www.govdata.de/ckan/api/rest/dataset/destatis-service-42153-0001>

dcat:accessURL

Kategorien: Bevölkerung *dct:theme*

Schlüsselwörter: bv41-kalender--und-saisonbereinigt, bv41-trend, originalwerte, x12-arima-arbeitstaglich--und-saisonbereinigt, x12-arima-arbeitstaglich-bereinigt *dcat:keyword*

dct:license

Freie Nutzung

Nutzungsbedingungen:
Datenlizenz Deutschland
Namensnennung

Stand Metadaten:
11.9.2014

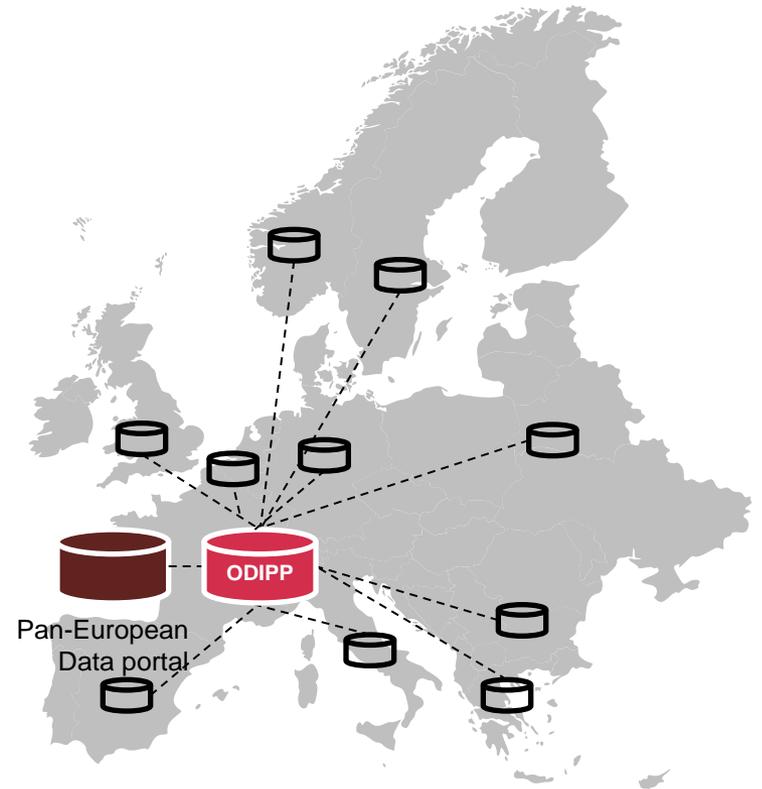
Veröffentlicht: 12.9.2014 *dct:issued*

Zeitraum: 12.9.2014 - 12.9.2014 *dct:periodOfTime*

☆☆☆☆☆ (0)

Was kann die Open Data Interoperability Platform?

- Sie **sammelt** Metadaten von einem offenen Datenportal
- Sie **verwandelt** die Metadaten in RDF
- Sie **vereinheitlicht** die RDF-Metadaten, die in den vorherigen Schritten mit DCAT-AP produziert worden sind
- Sie **bestätigt** die harmonisierten Metadaten mit Hilfe der DCAT-AP
- Sie **veröffentlicht** die Beschreibung der Metadaten als Linked Open Data



Siehe auch: <http://www.slideshare.net/OpenDataSupport/promoting-the-re-use-of-open-data-through-odip>

Schlussfolgerungen

- Metadaten liefern Informationen über Ihre Daten und Ressourcen. Die Qualität der Metadaten beeinflusst direkt die Auffindbarkeit und Weiterverwendung Ihrer Ressourcen
- Für die Metadatenverwaltung sollte ein strukturierter Ansatz verfolgt werden
- Der Metadaten-Lebenszyklus verlängert die Lebensdauer von Datensätzen (Metadaten vor der Veröffentlichung und nach dem Löschen)
- Vereinheitlichte Metadaten ermöglichen den Einsatz von Metadaten-Brokern, die die Zugangsbarrieren zu Ihren Ressourcen wiederum senken, was zu einer verbesserten Sichtbarkeit und Auffindbarkeit führen sowie das Wiederverwendungspotenzial steigern kann

Gruppenübung und Fragen



<http://www.visualpharm.com>

Wählen Sie in Zweiergruppen einen Datensatz aus Ihrem Land aus und beschreiben Sie ihn mit dem DCAT Anwendungsprofil.



<http://www.visualpharm.com>

Hat Ihre Organisation einen minimalen Satz von Metadaten, die zusammen mit Open Data zur Verfügung gestellt werden können?



<http://www.visualpharm.com>

Was sind Ihrer Meinung nach die größten Hindernisse in Ihren Metadaten für die (Wieder-)Verwendung von kontrollierten Standardvokabularien?



<http://www.visualpharm.com>

Haben Sie auf der betrieblichen Ebene eine Methodik für die Verwaltung von Daten und/oder Metadaten?

Machen Sie auch den Online-Test!

Vielen Dank!

... Und jetzt IHRE Fragen?

Referenzen

Folie 6,7:

- NISO. Understanding Metadata. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Folie 9:

- Dublin City University. Chapter 3: Introduction to XML. <http://wiki.eeng.dcu.ie/ee557/g2/326-EE.html>
- W3C. RDF Primer. <http://www.w3.org/TR/rdf-primer/>

Folie 12:

- http://gondolin.rutgers.edu/MIC/text/how/catalog_glossary.htm
- Dublin Core. Example XML Schema. <http://dublincore.org/schemas/xmls/qdc/dc.xsd>
- Dublin Core, Example RDF Schema. <http://dublincore.org/2012/06/14/dcterms.rdf>

Folien 14, 33:

- The ISA Programme. DCAT Application Profile for Data Portals in Europe - Final Draft. https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final-draf

Folie 18:

- ListPoint. ObjectInCrimeClass. <http://www.listpoint.co.uk/CodeList/details/ObjectInCrimeClass/1.2/1>

Folien 19:

- Publications Office. Countries Name Authority List. <http://open-data.europa.eu/en/data/dataset/2nM4aG8LdHG6RBMumfkNzQ>

Weiterführende Lektüre



Understanding Metadata, NISO.

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>



Ben Jareo and Malcolm Saldanha. The value proposition of a metadata driven data governance program. Best Practices Metadata. May 2012.

https://community.informatica.com/mpresources/Communities/IW2012/Docs/bos_30.pdf



John R. Friedrich, II. Metadata Management Best Practices and Lessons Learned. The 10th Annual Wilshire Meta-Data Conference and the 18th Annual DAMA International Symposium. April 2006.

<http://www.metaintegration.net/Publications/2006-Wilshire-DAMA-MetaIntegrationBestPractices.pdf>

Verwandte Initiativen



Metadata Management. Trainer screencasts, <http://managemetadata.com/screencasts/msa/>



MIT Libraries. Data Management and Publishing. Reasons to Manage and Publish Your Data, <http://libraries.mit.edu/guides/subjects/data-management/why.html>



ISA Programme. DCAT Application Profile for European Data Portals, https://joinup.ec.europa.eu/asset/dcat_application_profile/description



Generating ADMS-based descriptions of assets using Open Refine RDF, <https://joinup.ec.europa.eu/asset/adms/document/generate-adms-asset-descriptions-spreadsheet-refine-rdf>



The Dublin Core Metadata Initiative, <http://dublincore.org/>

Werden Sie Teil unseres Teams...

*Finden Sie
uns auf*



[Open Data Support](http://www.slideshare.net/OpenDataSupport)

<http://www.slideshare.net/OpenDataSupport>



[Open Data Support](http://goo.gl/y9ZZI)

<http://goo.gl/y9ZZI>

*Begleiten Sie
uns auf*



joinup

<http://www.opendatasupport.eu>

*Folgen Sie
uns auf*



[@OpenDataSupport](https://twitter.com/OpenDataSupport)

*Kontaktieren
Sie uns unter*

contact@opendatasupport.eu