

Metadane prezentacji

Projekt Open Data Support jest finansowany przez Komisję Europejską w ramach umowy SMART 2012/0107 „Część 2: Świadczenie usług na rzecz publikacji, udostępniania i wykorzystywania otwartych danych publicznych w Unii Europejskiej poprzez istniejące portale otwartych danych” (nr umowy 30-CE-0530965/00-17).

© Komisja Europejska, 2014



**OPEN
DATA
SUPPORT**

Moduł szkoleniowy 1.4

Zarządzanie metadanymi – wprowadzenie

Cele szkolenia

Po zakończeniu modułu szkoleniowego powinieneś znać:

- definicję metadanych;
- terminologię i cele związane z zarządzaniem metadanymi;
- różne aspekty jakości metadanych;
- zastosowanie słowników kontrolowanych do metadanych;
- proces wymiany i agregacji metadanych;
- proces zarządzania metadanymi w Open Data Support.

Treść

Niniejszy moduł zawiera...

- wyjaśnienie, czym są metadane;
- zarys cyklu życia metadanych;
- wprowadzenie do jakości metadanych;
- ogólny opis systemu zarządzania i wymiany metadanych wdrożony przez projekt Open Data Support poprzez platformę Open Data Interoperability Platform.

Czym są metadane?

Definicja, przykłady i dostępne standardy.

Czym są metadane?

“Metadane to ustrukturyzowane informacje opisujące, tłumaczące, lokalizujące i ułatwiające we wszelki inny sposób odnalezienie, wykorzystanie lub zarządzanie zasobem informacji. Metadane często określa się mianem ,danych o danych’ albo ,informacji o informacjach’.”

-- National Information Standards Organization

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

Metadane dostarczają informacji umożliwiających uporządkowanie **danych** (np. dokumentów, plików graficznych, zbiorów danych), **pojęć** (np. schematów klasyfikacyjnych) i **elementów świata rzeczywistego** (np. ludzi, organizacji, miejsc, obrazów, produktów).

Rodzaje metadanych

- **Metadane opisowe** opisują zasoby w celu ich odnalezienia i identyfikacji.
- **Metadane strukturalne**, np. modele danych i dane referencyjne.
- **Metadane administracyjne** dostarczają informacji ułatwiających zarządzanie zasobami.

Niniejsze szkolenie skupia się na metadanych opisowych dla zbiorów danych.

Częściowo poruszono także kwestię metadanych administracyjnych.

Przykłady metadanych

Etykieta



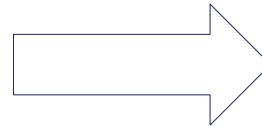
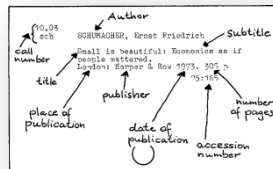
dostarcza metadanych
na temat



puszki



Karta katalogowa



książki



Opis zbioru danych (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>
.
```

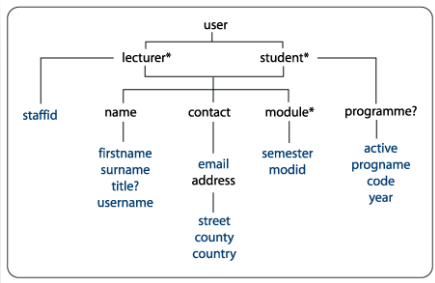


zbioru danych

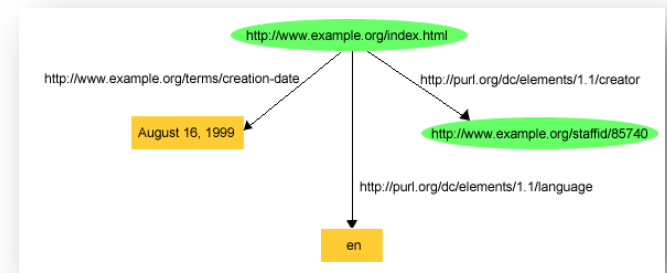
	Temp. °C	Humidity %	Wind direction	Wind speed km/h
Station 1	18.1	60	WSW	18
Station 2	17.5	59	WSW	20
Station 3	18.2	55	SW	22
Station 4	19.0	62	SW	18
Station 5	18.0	65	WSW	19
Station 6	18.2	63	SSW	21
Station 7	17.9	61	SW	22

Dwa sposoby udostępniania metadanych w sieci Web

XML (oparty na strukturze drzewa)



RDF (oparty na trzech elementach)



```
<?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">
      <programe>M.Eng in Electronic Systems</programe>
      <code>MEN</code>
      <year>1</year>
    </programme>
    <module semester="2">
      <modid>EE557</modid>
    </module>
    <module semester="1">
      <modid>EE553</modid>
    </module>
  </student>
```

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


Zarządzanie metadanymi w bazie danych

Zarządzanie metadanymi jest istotne

Zarządzanie metadanymi jest potrzebne, aby zapewnić...

- **Dostępność:** metadane muszą być przechowywane w dostępnym miejscu oraz indeksowane, aby można było je znaleźć.
- **Jakość:** metadane zawsze muszą być dobrej jakości, aby były wiarygodne dla użytkowników.
- **Trwałość:** metadane muszą być przechowywane przez określony czas.
- **Otwartą licencję:** metadane powinny być dostępne w domenie publicznej, aby mogły być wielokrotnie wykorzystywane.

*Cykl życia metadanych jest **dłuższy** od cyklu życia danych:*

- Metadane mogą zostać **utworzone zanim zostaną utworzone lub zebrane dane**, np. w formie informacji o danych dostępnych w przyszłości.
- Metadane muszą być **przechowywane po usunięciu danych**, np. w formie informacji o zlikwidowanych bądź wycofanych danych.

Schemat metadanych

“System etykietowania, oznaczania lub kodowania wykorzystywany do rejestrowania informacji o zasobach lub strukturyzowania metadanych opisowych. Schemat metadanych ustanawia i definiuje elementy danych oraz zasady wykorzystywania elementów danych do opisu zasobu.”

Schemat XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE RDF>
<rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:rd="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:dcam="http://purl.org/dc/dcam/" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:owl="http://www.w3.org/2002/07/owl#">
  <rdf:Description rdf:about="http://purl.org/dc/terms/">
    <dcterms:title xml:lang="en">DCMI Metadata Terms - other</dcterms:title>
    <dcterms:publisher rdf:resource="http://purl.org/dc/aboutdcmi#DCMI"/>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-06-14</dcterms:modified>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.org/dc/terms/title">
    <rdfs:label xml:lang="en">Title</rdfs:label>
    <rdfs:comment xml:lang="en">A name given to the resource.</rdfs:comment>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#title-002"/>
    <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/title"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.org/dc/terms/creator">
    <rdfs:label xml:lang="en">Creator</rdfs:label>
    <rdfs:comment xml:lang="en">An entity primarily responsible for making the resource.</rdfs:comment>
    <dcterms:description xml:lang="en">Examples of a Creator include a person, an organization, or a service.</dcterms:description>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#creator-002"/>
    <rdfs:range rdf:resource="http://purl.org/dc/terms/Agent"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/contributor"/>
    <owl:equivalentProperty rdf:resource="http://xmlns.com/foaf/0.1/maker"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.org/dc/terms/subject">
    <rdfs:label xml:lang="en">Subject</rdfs:label>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE RDF>
<rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:rd="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:dcam="http://purl.org/dc/dcam/" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:owl="http://www.w3.org/2002/07/owl#">
  <rdf:Description rdf:about="http://purl.org/dc/terms/">
    <dcterms:title xml:lang="en">DCMI Metadata Terms - other</dcterms:title>
    <dcterms:publisher rdf:resource="http://purl.org/dc/aboutdcmi#DCMI"/>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-06-14</dcterms:modified>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.org/dc/terms/title">
    <rdfs:label xml:lang="en">Title</rdfs:label>
    <rdfs:comment xml:lang="en">A name given to the resource.</rdfs:comment>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#title-002"/>
    <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/title"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.org/dc/terms/creator">
    <rdfs:label xml:lang="en">Creator</rdfs:label>
    <rdfs:comment xml:lang="en">An entity primarily responsible for making the resource.</rdfs:comment>
    <dcterms:description xml:lang="en">Examples of a Creator include a person, an organization, or a service.</dcterms:description>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#creator-002"/>
    <rdfs:range rdf:resource="http://purl.org/dc/terms/Agent"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/contributor"/>
    <owl:equivalentProperty rdf:resource="http://xmlns.com/foaf/0.1/maker"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://purl.org/dc/terms/subject">
    <rdfs:label xml:lang="en">Subject</rdfs:label>
```

Schemat RDF

Wykorzystanie dostępnych słowników do wprowadzenia metadanych do zasobów

Ogólne standardy i specyfikacje:

- **Dublin Core** do publikowanych materiałów (tekstu, obrazów), <http://dublincore.org/documents/dcmi-terms/>
- **FOAF** do osób i organizacji, <http://xmlns.com/foaf/spec/>
- **SKOS** do zbiorów pojęć, <http://www.w3.org/TR/skos-reference>
- **ADMS** do zasobów interoperacyjności, <http://www.w3.org/TR/vocab-adms/>

Szczególne standardy zasobów danych:

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

Szczególne zastosowanie standardu DCAT i innych słowników dla wsparcia interoperacyjności portali danych w Europie:

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

Tworzenie schematu metadanych przy pomocy schematu RDF (RDFS) – wykorzystuj dostępne terminy

Schemat RDF jest szczególnie dobrym narzędziem do łączenia terminów z różnych standardów i specyfikacji.

Nie twórz od nowa terminów, które zostały zdefiniowane w innym miejscu, tworząc schematy RDF – w miarę możliwości **wykorzystuj** istniejące terminy.

- Np. DCAT Application Profile for data portals in Europe (DCAT-AP) wykorzystuje terminy z DCAT, Dublin Core, FOAF, SKOS, ADMS i in.

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/	dcat:keyword	rdfs:Literal	This property contains a keyword or tag describing the dataset.	0..n

Przykład: opis otwartego zbioru danych przy użyciu DCAT-AP

Opis katalogu

```
: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 ;
  .
```

Opis zbioru danych

```
: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 ;
  .
```

Opis dystrybucji

```
: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 ;
  .
```

Słowniki kontrolowane

Wykorzystanie tezaursów, taksonomii i standaryzowanych list pojęć do przyporządkowywania wartości właściwościom metadanych.

Czym są słowniki kontrolowane?

Słownik kontrolowany to ustalona lista wartości przypisanych do określonej właściwości w schemacie metadanych.

- Poza starannym tworzeniem schematów, ważne są pola wartości poszczególnych właściwości metadanych ze względu na wymianę informacji, a co za tym idzie – interoperacyjność.
- Wspólne słowniki kontrolowane dla pól metadanych pozwalają na odczytywanie metadanych w różnych systemach.

Jakiego słownika kontrolowanego użyć do jakiego rodzaju właściwości

- Użyj **list kodów** jako słownika kontrolowanego w przypadku wolnego tekstu lub ciągu znaków.
- Przykładowa właściwość DCAT-AP
- Użyj **pojęć zidentyfikowanych przez URI** w odniesieniu do „obiektów”:
- Przykładowa właściwość DCAT-AP :

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

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

- Przykładowa lista kodów - ObjectInCrimeClass (ListPoint)

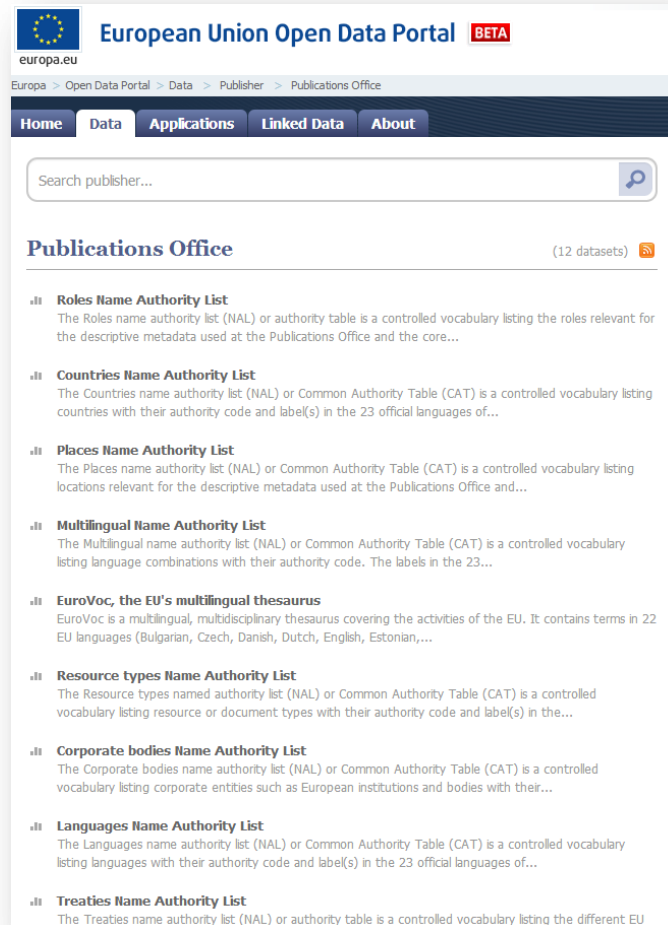
Code	Meaning
1	Used
2	Taken Without Consent
3	Eliminated
4	Not Involved
5	Involved
6	Crime Weapon
7	Handled / Received
11	Exhibit

- Przykładowa taksonomia pojęć z identyfikatorami URI - EuroVoc

```
<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">internationaal krediet</s04:prefLabel>
</rdf:Description>
```

Przykład – Listy organów krajowych Urzędu Publikacji

- Listy organów krajowych oferują słowniki kontrolowane:
 - państw
 - osób prawnych
 - rodzajów dokumentów
 - procedur międzyinstytucjonalnych
 - języków
 - w wersji wielojęzycznej
 - rodzajów zasobów
 - ról
 - traktatów



The screenshot displays the 'European Union Open Data Portal' interface. At the top, there is a search bar with the text 'Search publisher...'. Below the search bar, the 'Publications Office' dataset is highlighted, showing '(12 datasets)'. The list of datasets includes:

- 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

Cykl życia metadanych

*Tworzenie, zarządzanie, aktualizacja, przechowywanie
i publikacja metadanych oraz usuwanie danych.*

Tworzenie metadanych

Tworzenie metadanych mogą wspomagać procesy (pół)automatyczne.

- Właściwości dokumentu wygenerowane przez narzędzia (komputerowe), np. data utworzenia
- Informacje dot. czasu i miejsca zarejestrowane przez aparaty, czujniki...
- Informacje z danych publikacji, np. lokalizacja pliku lub adres URL

Mimo to inne właściwości wymagają ręcznego uzupełnienia:

- Czego dotyczy zasób (np. skojarzenie ze słownikiem tematycznym)?
- Jak można wykorzystać zasób (np. skojarzenie z licencją)?
- Gdzie mogę znaleźć więcej informacji nt. zasobu (np. skojarzenie ze stroną internetową albo dokumentacją opisującą zasób)?
- W jaki sposób dołączyć informację o jakości?

Zarządzanie metadanymi

Sposób zarządzania metadanymi musi być odpowiedni do rodzaju publikowanych danych.

- Jeśli **dane nie ulegają zmianie**, **metadane** mogą być względnie **stabilne**. W razie potrzeby zmiany (masowe przekształcenia) mogą być wprowadzane w trybie offline.
- Jeśli **dane zmieniają się regularnie** (np. dane w czasie rzeczywistym), **metadane** muszą być ściśle połączone z obiegiem danych, a **zmiany** muszą być praktycznie **natychmiastowe**.

Aktualizacja metadanych – planowanie zmiany

Metadane funkcjonują w kontekście globalnym, w którym zachodzą zmiany!

- **Organizacja** – dochodzi do powstawania i łączenia się departamentów, przekazywania kompetencji.
- **Użycie danych** – dane znajdują nowe zastosowania.
- **Dane referencyjne** – słowniki kontrolowane zmieniają i łączą się.
- **Standardy i technologie danych** – cykl życia technologii skraca się coraz bardziej; jak w przyszłości będzie wyglądała sieć Web?
- **Narzędzia i systemy** – ewolucja systemu przechowywania, przepustowości, telefonii komórkowej...

Metadane muszą być aktualizowane w możliwie największym stopniu, przy uwzględnieniu dostępnego czasu i budżetu.

Przechowywanie metadanych – jaki jest wybór?

W zależności od wymogów operacyjnych, metadane mogą zostać osadzone w danych lub przechowywane oddzielnie.

- Osadzanie metadanych w danych (np. dokumenty, pliki mp3, jpg, dane RDF) ułatwia wymianę danych.
- Oddzielne przechowywanie metadanych (np. w bazie danych) z odnośnikami do odpowiednich danych ułatwia zarządzanie nimi.

W zależności od dostępności narzędzi i wymogów dotyczących wydajności i pojemności, metadane mogą być przechowywane w „klasycznej” relacyjnej bazie danych lub w **potrójnym modelu RDF.**

Usuwanie danych

W wielu przypadkach metadane muszą zostać zachowane nawet po usunięciu opisywanych przez nich danych.

Wycofanie lub usunięcie danych ma miejsce m.in. :

- Gdy dane są już niepotrzebne
- Gdy dane są już nieaktualne
- Gdy dane są nieprawidłowe
- Gdy dane zostają wycofane przez właściciela/publikującego.

W takim przypadku metadane powinny **zawierać informację** o usunięciu danych lub, w przypadku **archiwizacji** danych, jak i gdzie można **uzyskać dostęp do kopii archiwum**.

Publikacja metadanych – jaki jest wybór?

- Publikacja „otwarta”: bezpośredni dostęp przez URI
 - Opcja w największym stopniu zgodna z ideą danych połączonych, pozwalająca na swobodne poruszanie się w obrębie danych.
- Udostępnienie metadanych przez **punkt końcowy SPARQL**
 - Umożliwia to systemom zewnętrznym realizować zapytania dla plików RDF.
 - Wymaga wiedzy na temat schematu używanego w trójkach RDF.
- Publikacja odroczonej: dostęp do wyeksportowanego pliku w RDF
 - Konwersja danych do formatu RDF.
 - Umożliwia masowe przetwarzanie i buforowanie zbiorów danych w trybie offline.
 - Umożliwia wdrożenie kontroli dostępu.

Zobacz również:

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

Jakość metadanych

Jakość i kompletność metadanych opisowych zbiorów danych bezpośrednio wpływa na możliwość ich wyszukiwania i wykorzystania.

Jakość metadanych polega na... (1/3)

- **precyzji** metadanych – czy właściwości zasobu zostały właściwie odzwierciedlone?
 - *np. podanie właściwego tytułu, licencji i publikującego umożliwia użytkownikom znalezienie poszukiwanych zasobów.*
- **dostępności** metadanych – czy do metadanych można uzyskać dostęp teraz i w przyszłości?
 - *np. umożliwienie indeksowania i pobierania metadanych oraz regularne tworzenie kopii zapasowych.*
- **kompletności** metadanych – czy zawarte zostały wszystkie istotne właściwości zasobu (w granicach praktycznej i finansowej wykonalności oraz w miarę potrzeby)?
 - *np. oznaczenie licencji określającej warunki użycia lub formatu dystrybucji umożliwia filtrowanie metadanych pod tym kątem.*

Jakość metadanych polega na... (2/3)

- **zgodności** metadanych z przyjętymi standardami – czy metadane są zgodne z jednym ze szczególnych standardów lub standardem AP?
 - *np. opis zbioru danych zgodny ze standardem DCAT-AP.*
- **spójności** metadanych – czy dane nie są sprzeczne?
 - *np. brak powtarzających się i sprzecznych informacji dot. licencji odnoszących się do tych samych danych.*
- **wiarygodności i pochodzeniu** metadanych – czy metadane pochodzą z wiarygodnych źródeł?
 - *np. odnośnik do danych referencyjnych opublikowanych i zarządzanych przez wiarygodną organizację (np. Urząd Publikacji UE).*

Jakość metadanych polega na... (3/3)

- **czytelności** metadanych – czy metadane nadają się do odczytu komputerowego?
 - *np. udostępnienie metadanych zbioru w formacie RDF i/lub XML zamiast jako wolnego tekstu.*
- **istotności** metadanych – czy metadane zawierają wystarczające informacje do realizacji bieżącego zadania?
 - *np. optymalne dostosowanie ilości informacji do potrzeb użytkowników.*
- **aktualności** metadanych – czy metadane odpowiadają aktualnym (obecnym) właściwościom zasobu i czy zostały opublikowane w porę?
 - *np. oznaczenie ostatniej daty modyfikacji zasobu, a tym samym upewnienie się, że metadane są aktualne, a użytkownicy widzą najnowsze informacje.*

Wymiana metadanych zbiorów

Mapping your metadata to a common metadata vocabulary, such as the DCAT-AP, and exchanging the metadata across platforms.

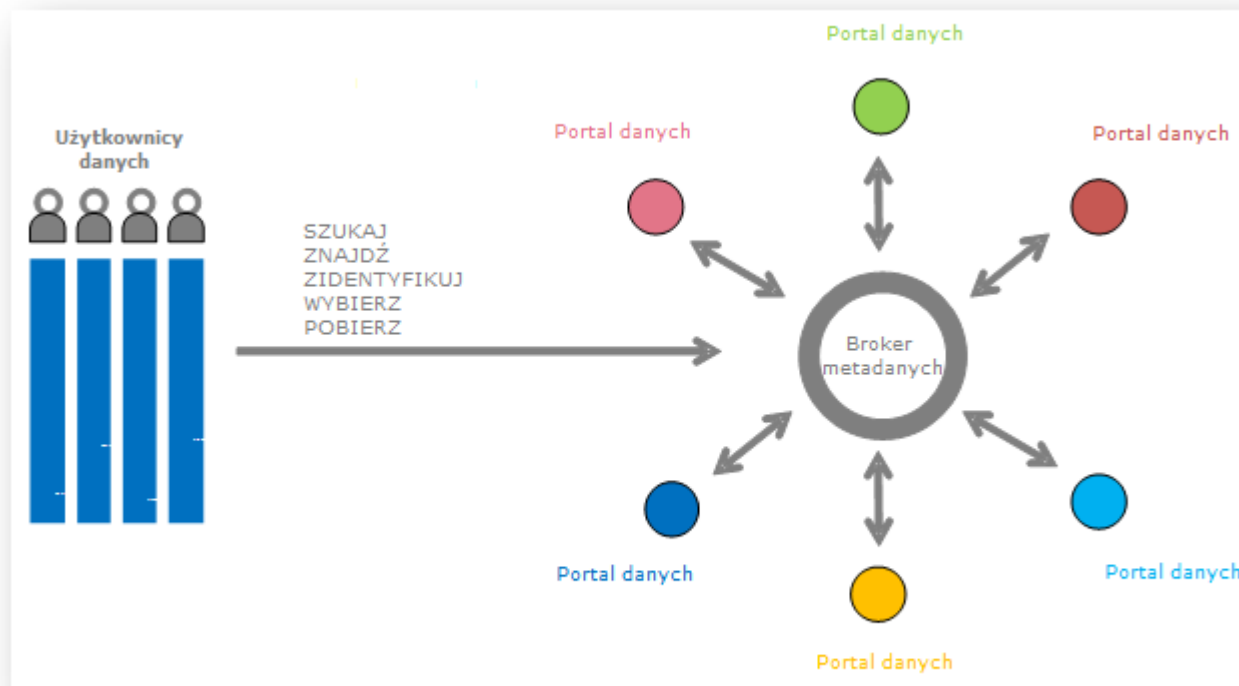
Ujednolicanie metadanych

W procesie wymiany między systemami metadane powinny być mapowane do wspólnego modelu, tak aby nadawca i odbiorca w ten sam sposób interpretowali metadane.

- Na **poziomie schematu**, metadane z różnych źródeł mogą zostać oparte na różnych schematach metadanych, np. DCAT, schema.org, CERIF, własnym modelu wewnętrznym...
- Na **poziomie danych** (wartości), właściwości metadanych powinny mieć przypisane wartości z **różnych słowników kontrolowanych** lub **składni**, np.:
 - Język: angielski może zostać wyrażony jako <http://publications.europa.eu/resource/authority/language/ENG> lub <http://id.loc.gov/vocabulary/iso639-1/en>
 - Daty: ISO8601 (“20130101”) lub W3C DTF (“2013-01-01”)

Przykład: Ujednolicanie metadanych dot. zbiorów Standard DCAT-AP dla portali danych w Europie

Standard DCAT-AP może być zastosowany jako wspólny model wymiany metadanych z otwartymi platformami danych w Europie i/lub brokerem danych (np. Platforma Interoperacyjności Otwartych Danych - ODIP).



Zobacz również:

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

Przykład mapowania – data.gov.uk

Scottish Road Accident Statistics

dct:title (Dataset)

Data about injury road accidents, accident costs, vehicles involved, drivers and riders, drink-drive accidents, drivers breath tested, casualties and international comparisons.



Source agency: Scottish Government

Designation: National Statistics

Language: English **dct:language**

Alternative title: Scottish Road Accident Statistics

dct:description

Licence **dct:license**

UK Open Government Licence (OGL) [OPEN DATA](#)

Data Resources **2**

Key statistics for 2007

2007 Volume

dct:title (Distribution)

Details Download ▾
Details Download ▾

Dcat:accessURL

**dcat:downloadURL, dct:issued,
dct:format, dct: description**

Additional Information

Openness score	☆☆☆☆☆
Geographic coverage	Scotland dct:spatial
National statistic	yes
ONS Category	Travel and Transport dct:theme
Temporal coverage	No value dct:temporal
Date added computed	No value
Date updated computed	No value

dct:publisher

Publisher

Scottish Government

Enquiries:

No details supplied

FOI Contact:

adms:contactPoint

• Web:

<http://www.whatdotheyknow.com...>

Tags

accident health-well-being-and-care
road road-accidents road-safety
roads safety transport
transport-accidents-and-casualties
travel-and-transport

dcat:keyword

About this dataset

- Added to data.gov.uk: 10/12/2011 **dct:issued**
- Modified on data.gov.uk: 10/06/2013 **dct:modified**
- History of changes
- JSON, API and URI for developers

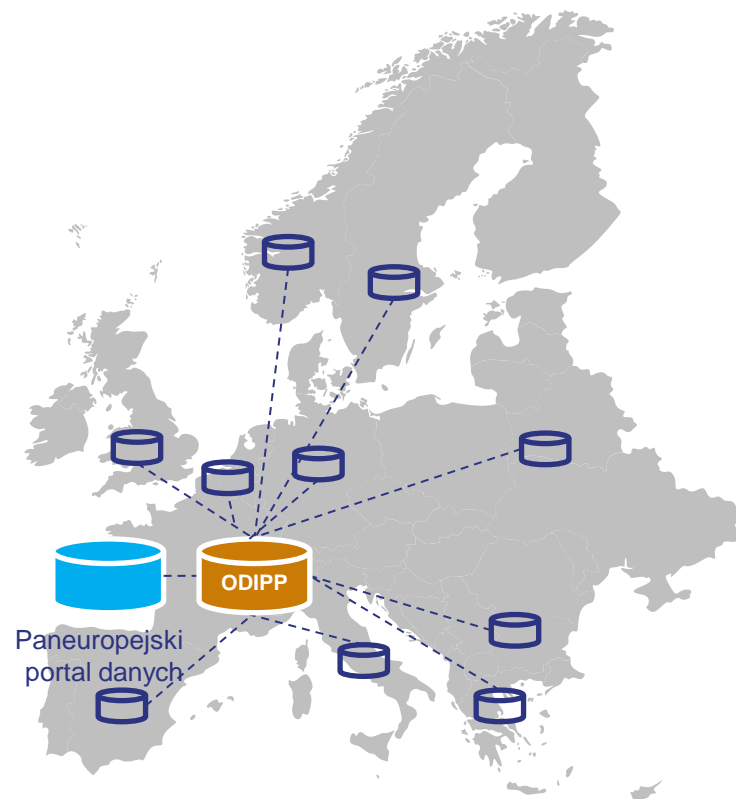
Do more with this data

- Share your app
- Share an idea
- Request new data

[Tweet](#) [Share](#) [Google +1](#)

Jakie są możliwości Platformy Interoperacyjności Otwartych Danych?

- **Pozyskanie** metadanych z portalu otwartych danych.
- **Konwersja** metadanych do formatu RDF.
- **Harmonizacja** metadanych RDF utworzonych w poprzednich etapach w standardzie DCAT-AP.
- **Walidacja** zharmonizowanych metadanych względem standardu DCAT-AP.
- **Publikacja** metadanych opisowych w formie otwartych danych połączonych.



Zobacz również:

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

Podsumowanie

- Metadane dostarczają informacji na temat danych i zasobów. Jakość metadanych bezpośrednio wpływa na możliwość wyszukiwania i wykorzystania zasobów.
- W zarządzaniu metadanymi należy przyjąć podejście strukturalne.
- Cykl życia metadanych jest dłuższy od cyklu życia zbiorów danych (metadane przed publikacją i po usunięciu danych).
- Ujednolicone metadane umożliwiają funkcjonowanie brokerów metadanych, co z kolei zwiększa dostępność zasobów, a tym samym zwiększa ich widoczność i ułatwia ich znalezienie, a w konsekwencji zwiększa możliwości ich wykorzystania.

Ćwiczenia i pytania dla grup



<http://www.visualpharm.com>

W parach wybierzcie i opiszcie jeden zbiór danych z Waszego kraju używając specyfikacji DCAT-AP.



<http://www.visualpharm.com>

Czy Wasza organizacja dysponuje minimalnym zbiorem metadanych przeznaczonym do otwartych danych?



<http://www.visualpharm.com>

Jakie są Waszym zdaniem główne przeszkody w wykorzystaniu standaryzowanych słowników kontrolowanych do Waszych metadanych?



<http://www.visualpharm.com>

Czy posiadacie metodologię zarządzania danymi i/lub metadany na poziomie korporacyjnym?

Rozwiąż test online!

Dziękujemy!
...i czekamy na pytania.

Prezentacja została przygotowana przez Open Data Support

Zastrzeżenia

1. Poglądy wyrażone w niniejszej prezentacji odzwierciedlają jedynie opinie autorów i w żadnym wypadku nie stanowią oficjalnego stanowiska Komisji Europejskiej.

Komisja Europejska nie gwarantuje prawidłowości informacji zawartych w niniejszej prezentacji i nie ponosi odpowiedzialności za ich wykorzystywanie.

Komisja Europejska nie promuje, nie poleca ani nie popiera żadnych produktów, specyfikacji, procesów lub usług, o których wspomniano w prezentacji pod postacią nazwy handlowej, znaku towarowego, nazwy producenta bądź inną.

Autor uzyskał zgodę od właścicieli praw autorskich lub ich przedstawiciela na wykorzystanie wszelkich fragmentów dokumentów objętych ochroną własności intelektualnej, w tym ilustracji, map i wykresów.

2. Niniejsza prezentacja została starannie opracowana przez firmę PwC, która nie gwarantuje jednak, że zawarte w niej informacje są pełne i prawidłowe. Firma PwC nie ponosi odpowiedzialności za informacje ani wszelkie decyzje bądź następstwa decyzji podjętych na podstawie informacji zawartych w niniejszej prezentacji. PwC nie ponosi odpowiedzialności za ewentualne szkody powstałe w wyniku wykorzystania informacji zawartych w prezentacji. Przedstawione informacje mają charakter ogólny i powinny służyć jedynie jako wskazówki w kwestiach będących przedmiotem zainteresowania ogólnego. Niniejsza prezentacja nie może zastępować profesjonalnej porady w jakiegokolwiek kwestii. Nikt nie powinien podejmować decyzji na podstawie jakiegokolwiek kwestii poruszonej w niniejszej publikacji bez uprzedniego zasięgnięcia profesjonalnej porady.

Autorzy:

Makx Dekkers, Michiel De Keyzer, Nikolaos Loutas and Stijn Goedertier

Oдноšniki

Slajdy 6, 7:

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

Slajd 9:

- Dublin City University. Chapter 3: Introduction to XML. <http://wiki.eeng.dcu.ie/ee557/q2/326-EE.html>
- W3C. RDF Primer. <http://www.w3.org/TR/rdf-primer/>

Slajd 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>

Slajdy 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

Slajd 18:

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

Slajd 19:

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

Literatura uzupełniająca



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>

Podobne inicjatywy



Zarządzanie metadanymi. Filmy instruktażowe,
<http://managemetadata.com/screencasts/msa/>



Biblioteki MIT. Zarządzanie danymi i publikacja danych. Dlaczego warto zarządzać danymi i publikować je,
<http://libraries.mit.edu/guides/subjects/data-management/why.html>



Program ISA. Standard DCAT-AP dla europejskich portali danych,
https://joinup.ec.europa.eu/asset/dcat_application_profile/description



Tworzenie opisów zasobów w standardzie ADMS przy pomocy Open Refine RDF,
<https://joinup.ec.europa.eu/asset/adms/document/generate-adms-asset-descriptions-spreadsheet-refine-rdf>



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

Dołącz do nas...

Odwiedź nas



Open Data Support

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



Open Data Support

<http://goo.gl/y9ZZI>

Śledź nas



@OpenDataSupport

Dołącz do nas



joinup

<http://www.opendatasupport.eu>

Kontakt

contact@opendatasupport.eu