

# *Presentation metadata*

Open Data Support is funded by the European Commission under SMART 2012/0107 'Lot 2: Provision of services for the Publication, Access and Reuse of Open Public Data across the European Union, through existing open data portals' (Contract No. 30-CE-0530965/00-17).

© 2013 European Commission



## OPEN DATA SUPPORT

Module de formation 1.4

# Introduction à la gestion des métadonnées

## *Les objectifs de cette formation*

À la fin de ce module de formation, vous devriez avoir une compréhension de:

- Ce que sont les métadonnées;
- La terminologie et les objectifs de la gestion des métadonnées;
- Les différentes dimensions de la qualité des métadonnées;
- L'utilisation de vocabulaires contrôlés pour les métadonnées;
- L'échange et l'agrégation de métadonnées;
- La gestion des métadonnées au sein de Open Data Support.

# Contenu

Ce module contient ...

- Une explication de ce que sont les métadonnées;
- Un aperçu du cycle de vie des métadonnées;
- Une introduction à la qualité des métadonnées;
- An overview of the metadata management and exchange approach implemented by Open Data Support through the Open Data Interoperability Platform.
- Un vue d'ensemble de la gestion des métadonnées et de l'approche d'échange mis en place par « Open Data Support » à travers la plate-forme « Open Data Interoperability ».

# Que sont les métadonnées?

*Définition, exemples et normes réutilisables.*

## Que sont les métadonnées?

*“Les métadonnées sont de l’information structurée qui décrit, explique, localise ou facilite autrement l’obtention, l’utilisation ou la gestion d’une ressource d’information. Les métadonnées sont souvent appelées données des données ou information sur l’information.”*

-- National Information Standards Organization

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

Les métadonnées fournissent des informations permettant de comprendre des **données** (documents, images, bases de données), des **concepts** (par exemple, les systèmes de classification) et des **entités du monde réel** (par exemple, les personnes, les organisations, les lieux, peintures, produits).

## *Types de métadonnées*

- **Métadonnées descriptives**, décrivent une ressource à des fins de découverte et d'identification.
- **Métadonnées structurelles**, par ex. des modèles de données et des données de référence.
- **Métadonnées administratives**, renseignent afin d'aider à la gestion d'une ressource.

Dans ce tutoriel nous nous concentrons principalement sur les métadonnées descriptives pour les ensembles de données.

Les métadonnées administratives sont aussi abordées.

# Des exemples de métadonnées

## Étiquette



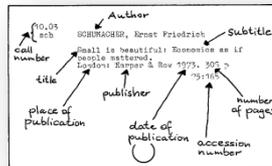
Fournit des métadonnées sur



## Canette



## Carte de catalogue



## Livre



## Description d'ensemble de données (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>
.
```

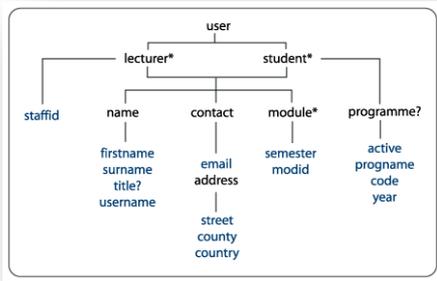


## Dataset

	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

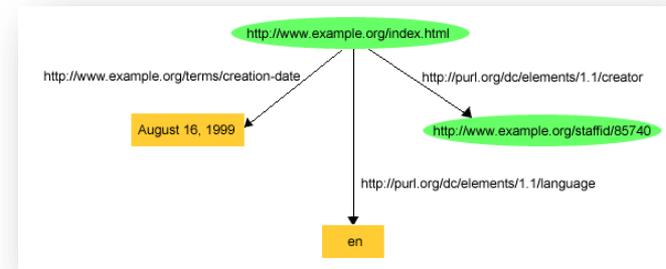
# Deux approches pour fournir des métadonnées sur le Web

## XML (Tree/container approach)



```
<?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>
```

## RDF (approche basée sur le triplets)



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

# Gérez les métadonnées de vos ensembles de données

## *La gestion des métadonnées est importante*

*Les métadonnées doivent être gérées pour assurer leur...*

- **Disponibilité:** les métadonnées doivent être stockées où elles peuvent être consultées et indexées afin de pouvoir être trouvées.
- **Qualité:** les métadonnées doivent être de qualité constante afin que les utilisateurs sachent qu'ils peuvent y faire confiance.
- **Persistance:** les métadonnées doivent être entretenues au fil du temps.
- **Licence ouverte:** les métadonnées devraient être disponibles sous une licence du domaine public pour permettre leur réutilisation.

*Le cycle de vie de métadonnées **est plus grand** que le cycle de vie des données:*

- Les métadonnées **peuvent être créées avant que les données ne soient créées** ou capturées, par exemple, pour informer sur les données qui seront disponibles dans le future.
- Les métadonnées doivent être **conservées après que les données ont été supprimées**, par exemple, pour informer sur les données qui ont été déclassées ou retirées.

# Schéma de métadonnées

« Un système de labellisation, marquage et codage utilisé pour l'enregistrement d'information de catalogue ou pour structurer des notices descriptives. Un schéma de métadonnées établit et définit les éléments de données et les règles d'application pour l'usage de ces éléments de données pour décrire une ressource. »

## Schéma 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>
      <dcterms: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"/>
    </dcterms:modified>
  - <rdf:Description>
    <rdf:Description rdf:about="http://purl.org/dc/terms/subject">
      <rdfs:label xml:lang="en">Subject</rdfs:label>
```

## Schéma RDF

## *Réutiliser des vocabulaires existants pour fournir des métadonnées à vos ressources*

*Normes et spécifications à usage général :*

- **Dublin Core** pour les documents publiés (textes, images), <http://dublincore.org/documents/dcmi-terms/>
- **FOAF** pour les personnes et les organisations, <http://xmlns.com/foaf/spec/>
- **SKOS** pour les collections conceptuelles, <http://www.w3.org/TR/skos-reference>
- **ADMS** pour les ressources d'interopérabilité, <http://www.w3.org/TR/vocab-adms/>

*Norme spécifique pour les ensembles de données:*

- **Vocabulaire pour les Catalogue de Données DCAT**, <http://www.w3.org/TR/vocab-dcat/>

*Usage spécifique de DCAT et d'autres vocabulaires pour soutenir l'interopérabilité des portails de données à travers l'Europe:*

**Profil applicatif DCAT pour les portails de données en Europe**, [http://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](http://joinup.ec.europa.eu/asset/dcat_application_profile/description)

# Concevoir votre schéma de métadonnées avec RDF Schema (RDFS) - à réutiliser autant que possible

Schéma RDF est particulièrement bon en termes de combinaison de différentes normes et spécifications.

**Ne réinventez pas** des termes qui sont déjà définis ailleurs, lors de la conception de schémas RDF – réutilisez des termes lorsque cela est possible.

- Par exemple, le profil applicatif DCAT pour les portails de données en Europe (DCAT-AP) réutilise les termes de DCAT, Dublin Core, FOAF, SKOS, ADMS et autres.

**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

# Exemple: description d'un ensemble de données ouvert avec le EDSC-AP

## Description du Catalogue

```
: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 ;
  .
```

## Description de l'ensemble des données

```
: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 ;
  .
```

## Description de la distribution

```
: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 ;
  .
```

# Les vocabulaires contrôlés

*Utiliser les thésauri, taxonomies et les listes normalisées de termes pour attribuer des valeurs aux propriétés de métadonnées.*

## Que sont les vocabulaires contrôlés?

*Un vocabulaire contrôlé est une liste prédéfinie de valeurs à utiliser comme valeur pour une propriété spécifique dans votre schéma de métadonnées.*

- En plus de la conception minutieuse des schémas, des écarts de valeurs pour les propriétés des métadonnées sont importants pour l'échange d'information, et donc pour l'interopérabilité.
- Des vocabulaires contrôlés communs pour les écarts de valeurs rendent les métadonnées compréhensibles à travers les systèmes.

# Quel vocabulaire contrôlé utilisez pour quel type de propriété

- Utilisez des **listes de codes** comme vocabulaire contrôlé pour un texte libre ou des propriétés utilisant des suites de caractères.

- Exemple de propriété DCAT-AP:

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

- Exemple de liste de codes-ObjectInCrimeClass (Listpoint)

Code	Meaning
1	<a href="#">Used</a>
2	<a href="#">Taken Without Consent</a>
3	<a href="#">Eliminated</a>
4	<a href="#">Not Involved</a>
5	<a href="#">Involved</a>
6	<a href="#">Crime Weapon</a>
7	<a href="#">Handled / Received</a>
11	<a href="#">Exhibit</a>

Utilisez des **concepts identifiés par un URI** comme référence aux «choses».

- Exemple de propriété DCAT-AP:

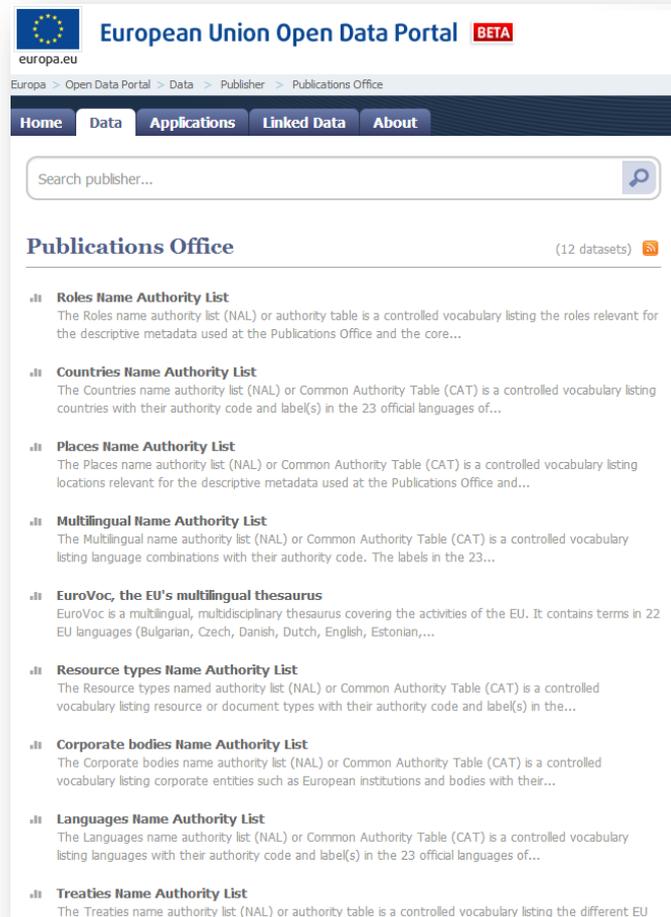
theme/ category	dcat:theme, subproperty of dct:subject	skos:Concept	This property refers to a category of the dataset.
-----------------	--	--------------	--

- Un exemple de taxonomie de termes ayant un 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>
```

# Exemple – “Publications Office’s Named Authority Lists”

- Les «Named Authority Lists» offrent des vocabulaires contrôlés réutilisables pour:
  - Les pays
  - Les organismes
  - Les types de fichiers
  - Les procédures interinstitutionnelles
  - Les langues
  - Multilingue
  - Les types de ressources
  - Les rôles
  - Les Traités



The screenshot displays the European Union Open Data Portal interface. At the top, the logo of the European Union and the text "European Union Open Data Portal" are visible, along with a "BETA" badge. Below the header, there is a navigation menu with options: Home, Data, Applications, Linked Data, and About. A search bar is present with the placeholder text "Search publisher...". The main content area is titled "Publications Office" and indicates "(12 datasets)". A list of datasets is shown, each with a title and a brief description:

- 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

# Le cycle de vie des métadonnées

*Création, entretien, mise à jour, stockage, publication des métadonnées et manipulation de la suppression des données.*

## Créer vos métadonnées

*La création de métadonnées peut être prise en charge par des processus (semi-)automatiques.*

- Les propriétés de documents générées par des outils (de bureau), par exemple la date de création d'un document.
- Informations spatiales et temporelles capturées par des caméras, des capteurs ...
- Informations issues du processus de publication, par exemple l'emplacement de fichier ou l'URL

*Cependant, d'autres caractéristiques requièrent une intervention humaine:*

- Quel est l'objet de la ressource (par exemple un lien vers le vocabulaire d'un sujet)?
- Comment peut-on utiliser les ressources (par exemple un lien vers une licence)?
- Où puis-je trouver plus d'information sur cette ressource (par exemple un lien vers un site Web ou de la documentation qui décrit la ressource)?
- Comment de l'information de qualité peut être incluse?

## *Entretenir vos métadonnées*

*Les approches pour l'entretien des métadonnées doivent être appropriées pour le type de données qui sont publiées.*

- Si **les données ne changent pas**, **les métadonnées** peuvent être relativement **stables**. Des changements (conversions en vrac) peuvent avoir lieu hors ligne lorsque cela est nécessaire.
- Si **les données changent fréquemment** (par exemple, les données de capteurs en temps réel), **les métadonnées** doivent être étroitement couplé au flux de données et **les changements** doivent être pratiquement **instantanés**.

# *Mise à jour de vos métadonnées - planifier pour le changement*

*Les métadonnées opèrent dans un contexte mondial qui fait l'objet de changements!*

- **Organisation** - des départements sont établis, fusionne avec d'autres et des responsabilités sont transmises.
- **L'utilisation des données** - de nouvelles applications émergent autour des données.
- **Les données de référence** - les vocabulaires contrôlés évoluent et se relient.
- **Normes de données et de technologies** – le cycle de vie des technologies se raccourcit tout le temps; à quoi ressemblera le Web de demain?
- **Outils et systèmes** - Évolution du stockage, de la bande passante, de la mobilité ...

*Les métadonnées doivent être gardées à jour dans la mesure du possible, en tenant compte du temps et des budgets à disposition.*

## Stocker vos métadonnées - quelles sont les options?

*En fonction des exigences opérationnelles, les métadonnées peuvent être intégrées avec les données ou stockées séparément des données.*

- Intégrer les métadonnées dans les données (documents bureautiques, MP3, JPG, données RDF) facilite l'échange de données.
- La séparation des métadonnées et des données (par exemple dans une base de données), avec des liens vers des fichiers de données correspondants rend la gestion plus facile.

En fonction de la disponibilité des outils et des exigences de performances et de capacité, les métadonnées peuvent être stockées dans une **base de données relationnelle «classique»** ou un **« store » de triplets en RDF.**

# *La gestion de la suppression des données*

*Dans de nombreux cas, les métadonnées doivent survivre même après la suppression des données qu'elles décrivent.*

**Le déclassé**ment ou **la suppression** des données peut se produire, par ex.:

- Lorsque les données ne sont plus nécessaires
- Lorsque les données ne sont plus valables.
- Lorsque les données sont mauvaises.
- Lorsque les données sont retirées par le propriétaire / éditeur

Dans ce cas, les métadonnées doivent **contenir l'information** que les données ont été **supprimées**, et si elles ont été **archivées**, comment et où une **copie d'archivage** peut être demandée.

## *Publier vos métadonnées - quelles sont les options?*

- Publication « ouverte »: accès direct par les URIs
  - Cette option est la plus conforme à la vision des données liées ouvertes.
- Rendez vos métadonnées disponibles par le biais d'un « **SPARQL endpoint** »
  - Cela permet à des systèmes externes d'envoyer des requêtes à un « store » de triplets en RDF.
  - Il faut connaître le schéma utilisé dans le « store » de triplets.
- Publication différée: l'accès au dossier exporté en RDF
  - Produit par la conversion des données non-RDF en RDF.
  - Permet la récolte en vrac hors-ligne et la mise en cache des collections de données.
  - Permet l'implémentation du contrôle d'accès.

**Voir aussi:**

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

# Qualité des métadonnées

*La qualité et l'exhaustivité des métadonnées de description de vos bases de données influent directement sur leur visibilité et leur réutilisation.*

## La qualité des métadonnées porte sur... (1/3)

- La **précision** de vos métadonnées – est-ce que les caractéristiques de la ressource sont correctement reflétée?
  - *par ex. indiquer le bon titre, le bon droit de licence, le bon droit éditeur permet aux utilisateurs de découvrir les ressources dont ils ont besoin.*
- La **disponibilité** de vos métadonnées – est-ce que les métadonnées sont accessible dès maintenant et au fil du temps dans le futur?
  - *par ex. rendre les données disponibles pour l'indexation et le téléchargement, et les inclure dans un processus de back-up régulier.*
- L'**exhaustivité** de vos métadonnées – est-ce que toutes les caractéristiques pertinentes de la ressource sont capturées (pour autant que c'est faisable pratiquement et économiquement ainsi que nécessaire pour l'application)?
  - *par ex. indiquer la licence qui régit la réutilisation ou le format de la distribution permet de filtrer sur ces aspects.*

## La qualité des métadonnées porte sur... (2/3)

- La **conformité** de vos métadonnées aux normes acceptées - est ce que les métadonnées sont conformes à une norme spécifique de métadonnées ou à un profil applicatif?
  - *par ex. la description d'un ensemble de données est conforme au DCAT-AP.*
- La **cohérence** de vos métadonnées - les données ne contiennent elles pas de contradictions?
  - *par ex. ne pas avoir de multiples licences contradictoires pour la donnée.*
- La **crédibilité** et la **provenance** de vos métadonnées - sont elles basées sur des sources dignes de confiance?
  - *par ex. lier aux données de référence publiées et gérées par une organisation stable(par ex. l'Office des publications de l'UE)*

## *La qualité des métadonnées porte sur... (3/3)*

- **La capacité de traitement** des métadonnées - les métadonnées sont elles correctement lisibles par machine?
  - *par ex. en rendant disponible les métadonnées en RDF et/ou XML, et non en texte libre.*
- La **pertinence** des métadonnées – est-ce que les métadonnées contiennent la bonne quantité d'information pour la tâche à accomplir?
  - *par ex. limiter l'information pour servir de façon optimale les besoins des utilisateurs.*
- **La ponctualité** de vos métadonnées – correspondent les métadonnées aux caractéristiques réelles (actuelles) de la ressource et sont elles publiées assez vite?
  - *par ex. En indiquant la date de dernière modification de la ressource, afin d'assurer la fraîcheur des métadonnées et que les utilisateurs voient l'information la plus récente.*

# Échanger les métadonnées d'ensembles de données

*Établir une correspondance entre vos métadonnées et un vocabulaire de métadonnées commun, et échanger les métadonnées sur toutes les plateformes.*

# Homogénéisation des métadonnées

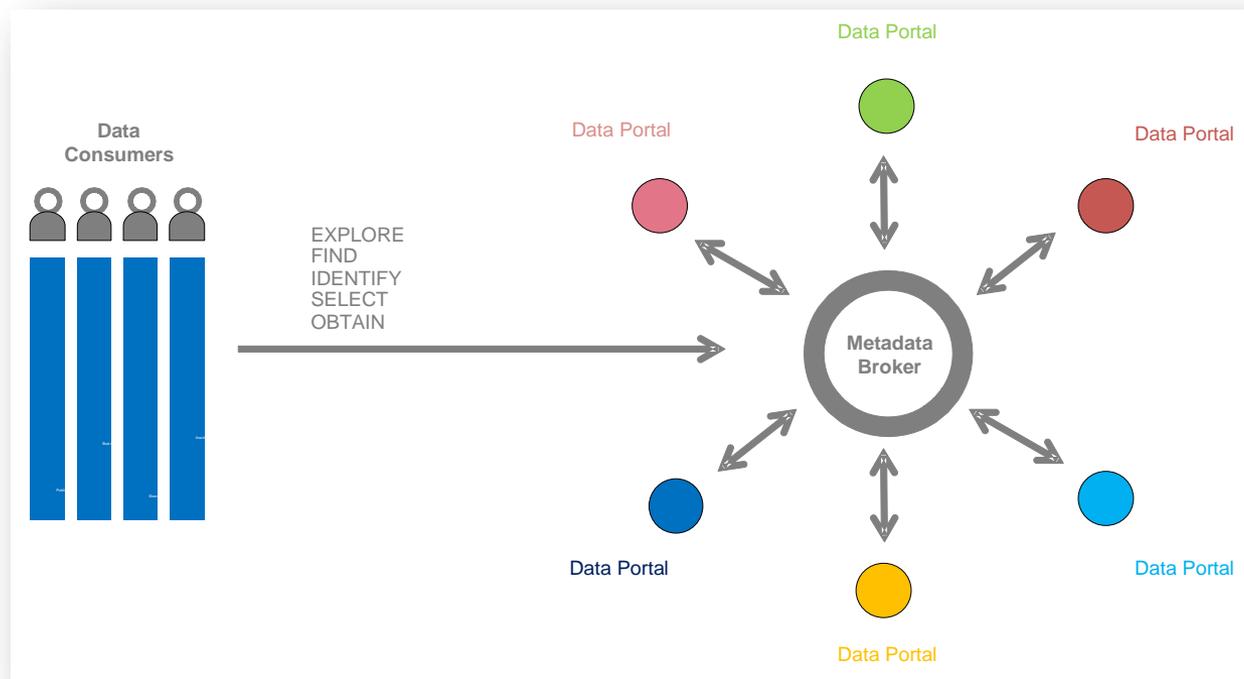
*Lorsqu'elles sont échangées entre les systèmes, les métadonnées doivent correspondre à un modèle commun de sorte que l'expéditeur et le destinataire partagent une compréhension commune de la signification des métadonnées.*

- Au **niveau du schéma**, des métadonnées provenant de différentes sources peuvent être basées sur **différents schémas de métadonnées**, par ex. DCAT, schema.org, CERIF, modèle propre interne ...
- Au **niveau des données** (valeurs), les propriétés des métadonnées doivent recevoir **les valeurs de différents vocabulaires ou syntaxes contrôlés**, par ex.:
  - Langue: L'Anglais peut être exprimée comme:  
<http://publications.europa.eu/resource/authority/language/ENG> ou comme:  
<http://id.loc.gov/vocabulary/iso639-1/en>
  - Dates: ISO8601 (“20130101”) contre W3C DTF (“2013-01-01”)

## Exemple: Homogénéisation des métadonnées

### Le profil applicatif DCAT pour les portails de données en Europe

DCAT-AP peut être utilisé comme modèle commun pour l'échange de métadonnées avec les plateformes de données ouvertes à travers l'Europe et / ou avec une plateforme d'échange de données (par ex. The Open Data Interoperability Platform - ODIP).



**Voir aussi:**

[http://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/home](http://joinup.ec.europa.eu/asset/dcat_application_profile/home)

# Exemple de correspondance – 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.



dct:description

Source agency: Scottish Government

Designation: National Statistics

Language: English **dct:language**

Alternative title: Scottish Road Accident Statistics

Licence **dct:license**

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

Data Resources **2**

Key statistics for 2007

2007 Volume

**dct:title (Distribution)**

**Dcat:accessURL**

Details Download ▾  
Details Download ▾

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

## Additional Information

Openness score	☆☆☆☆☆
Geographic coverage	Scotland <b>dct:spatial</b>
National statistic	yes
ONS Category	Travel and Transport <b>dct:theme</b>
Temporal coverage	No value <b>dct:temporal</b>
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
- Modified on data.gov.uk: 10/06/2013
- History of changes
- JSON, API and URI for developers

**dct:issued**

**dct:modified**

## Do more with this data

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



Tweet



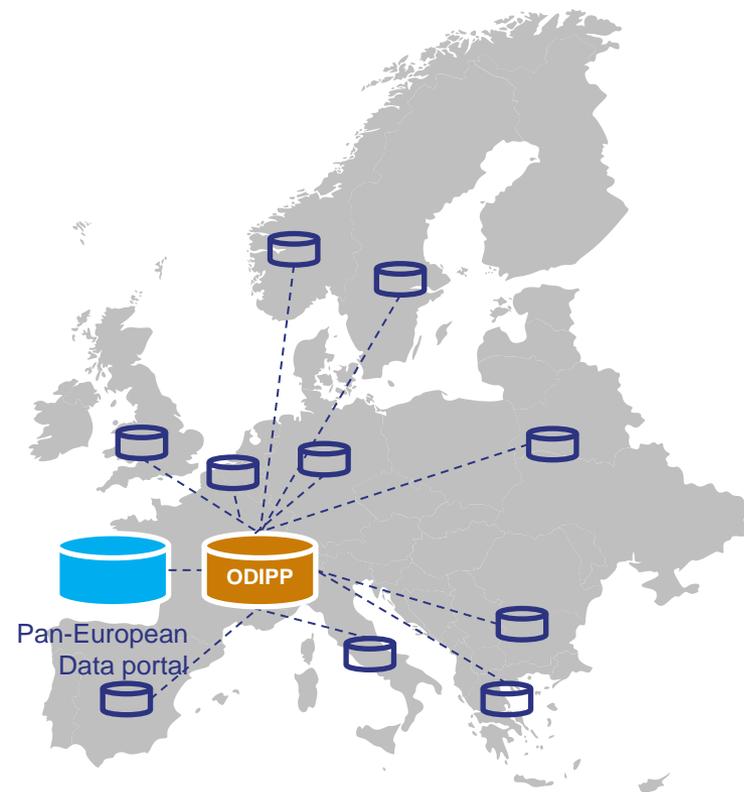
Share



Google +1

## Que peut faire la “Open Data Interoperability Platform” ?

- **Récolter** des métadonnées à partir d'un portail d'ouverture des données.
- **Transformer** les métadonnées en RDF.
- **Harmoniser** les métadonnées RDF produite dans les étapes précédentes avec DCAT-AP.
- **Valider** les métadonnées harmonisées par rapport à DCAT-AP.
- **Publier** les métadonnées de description comme des données liées.



**Voir aussi:**

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

## Conclusion

- Les métadonnées fournissent des informations sur vos données et ressources. La qualité des métadonnées influe directement sur la visibilité et la réutilisation de vos ressources.
- Une approche structurée devrait être suivie pour la gestion des métadonnées.
- Le cycle de vie de métadonnées prolonge le cycle de vie des ensembles de données (métadonnées avant publication et après suppression).
- Les métadonnées homogénéisées permettent le fonctionnement des plateformes d'échange de métadonnées, qui peuvent à leur tour réduire les barrières d'accès à vos ressources, conduisant à une meilleure visibilité, et donc augmentent leur potentiel de réutilisation.

## Questions et exercice de groupe



<http://www.visualpharm.com>

En groupe de deux, sélectionnez un ensemble de données de votre pays et décrivez-le avec le profil applicatif DCAT.



<http://www.visualpharm.com>

Est-ce que votre organisation a une liste de métadonnées requises pour la publication de données ouvertes?



<http://www.visualpharm.com>

Quelles seraient, d'après vous, les principales barrières pour la (ré)utilisation de vocabulaires contrôlés standards dans vos métadonnées?



<http://www.visualpharm.com>

Avez-vous une méthodologie de gestion des données et/ou des métadonnées au niveau de votre organisation?

**Prenez aussi le test en ligne ici!**

# Merci!

## ...et maintenant vos questions?

# *This presentation has been created by Open Data Support*

## **Disclaimers**

1. The views expressed in this presentation are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission. The European Commission does not guarantee the accuracy of the information included in this presentation, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission. All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.
2. This presentation has been carefully compiled by PwC, but no representation is made or warranty given (either express or implied) as to the completeness or accuracy of the information it contains. PwC is not liable for the information in this presentation or any decision or consequence based on the use of it. PwC will not be liable for any damages arising from the use of the information contained in this presentation. The information contained in this presentation is of a general nature and is solely for guidance on matters of general interest. This presentation is not a substitute for professional advice on any particular matter. No reader should act on the basis of any matter contained in this publication without considering appropriate professional advice.

Authors:

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

# Références

Slide 6, 7:

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

Slide 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/>

Slide 12:

- [http://gondolin.rutgers.edu/MIC/text/how/catalog\\_glossary.htm](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>

Slide 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](https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final-draf)

Slide 18:

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

Slide 19:

- Publications Office. Countries Name Authority List. <http://open-data.europa.eu/en/data/dataset/2nM4aG8LdHG6RBmumfkNzQ>

# Lecture supplémentaire



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](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>

## Initiatives apparentées



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](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/>

# *Be part of our team...*

## *Find us on*



Open Data Support

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



Open Data Support

<http://goo.gl/y9ZZI>

## *Follow us*



@OpenDataSupport

## *Join us on*



joinup

<http://www.opendatasupport.eu>

## *Contact us*

[contact@opendatasupport.eu](mailto:contact@opendatasupport.eu)