



**OPEN
DATA
SUPPORT**

Module de formation 2.2

La qualité des Données et Métadonnées Ouvertes



PwC firms help organisations and individuals create the value they're looking for. We're a network of firms in 158 countries with close to 180,000 people who are committed to delivering quality in assurance, tax and advisory services. Tell us what matters to you and find out more by visiting us at www.pwc.com. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

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

This presentation has been created by PwC

Authors:

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

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.

Les objectifs de cette formation

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

- Ce qu'on entend par la qualité des données (ouvertes).
- Les déterminants et les critères de qualité pour les données ouvertes.
- Les bonnes pratiques pour la publication des données ouvertes (liées) de haute qualité.

Contenu

Ce module contient ...

- Une définition pour la « qualité des données »;
- Une vue d'ensemble des dimensions pour la qualité des données et des métadonnées;
- Une sélection des meilleures pratiques pour publier des données et des métadonnées de bonne qualité.

Qu'est-ce que la qualité des données (et métadonnées)?

Les données sont de haute qualité « si elles sont aptes pour leurs utilisations prévues dans les opérations, la prise de décision et la planification. »

Ou plus précisément:

« Des données de qualité sont précises, disponibles, complètes, conformes, cohérentes, crédibles, traitables, pertinentes et ponctuelles »

Les métadonnées sont des données sur des données

“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 sur des données ou information sur l’information”

-- National Information Standards Organization

- Nous observons que les métadonnées sont un type de données.
- Les mêmes considérations s'appliquent à la qualité des métadonnées qu'à la qualité des données.

Les dimensions de la qualité des données

Quelles sont les principales dimensions à prendre en compte pour fournir des (méta)données de bonne qualité?

Les dimensions de la qualité des données.

- **Précision:** Les données représentent-elles correctement l'entité ou l'événement du monde réel?
- **Cohérence:** Les données ne contiennent-elles pas de contradictions?
- **Disponibilité:** Les données sont-elles accessibles maintenant et au fil du temps?
- **Exhaustivité:** Est-ce que les données comprennent tous les éléments de données représentant l'entité ou d'un événement?
- **Conformité:** Les données suivent-elles les normes en vigueur?
- **Crédibilité:** Les données proviennent-elles de sources dignes de confiance?
- **Traitabilité:** Les données sont-elles lisibles par machine?
- **Pertinence:** Est-ce que les données comprennent une quantité appropriée de données?
- **Ponctualité :** Les données sont-elles représentatives de la situation actuelle et sont-elles publiées assez vite?

Précision

La précision des données est la mesure dans laquelle elles représentent correctement les caractéristiques de l'objet, une situation ou un événement dans le monde réel.

Par exemple:

- Une mesure correcte des conditions météorologiques (température, précipitations).
- Bonne indication des conditions de réutilisation des données.

Recommandations:

- **Équilibrer l'exactitude** de vos données par rapport au **coût** dans le contexte de l'application; les données doivent être assez **bonnes pour l'usage prévu**.
- Assurez-vous qu'il existe un **engagement organisationnel** et de **l'investissement dans les procédures et les outils** pour maintenir la précision

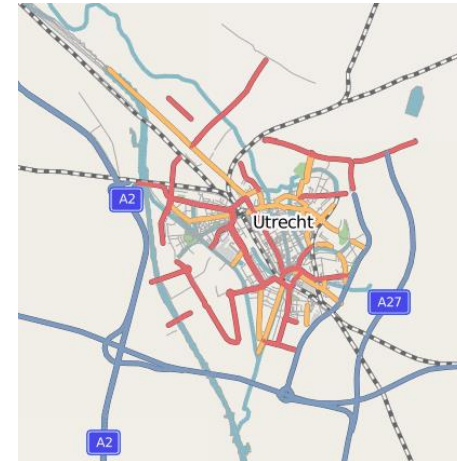
Exemple de précision

Forte précision

OpenStreetMap, Ville d'Utrecht, Pays-Bas (2011 vs. 2007)



Faible précision



Cohérence

La cohérence des données est la mesure dans laquelle elle ne contiennent pas de contradictions qui rendent leur utilisation difficile ou impossible.

Par exemple:

- Un ensemble de données qui combine les données provenant de différentes sources qui ont été traitées pour détecter les déclarations contradictoires qui ont été résolues.
- Une description d'un ensemble de données qui ne contient pas plusieurs déclarations de licence ou lorsque la date de la dernière modification n'est pas avant la date de création.

Recommandations:

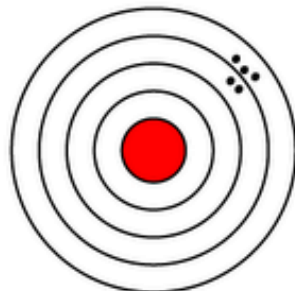
- **Traiter toutes les données avant leur publication** afin de détecter des déclarations contradictoires et d'autres erreurs (en particulier si les données sont agrégées à partir de différentes sources).

Exemple de cohérence

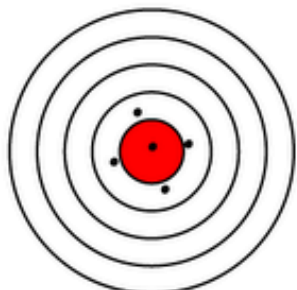
Forte cohérence



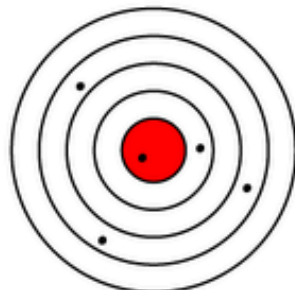
Accurate & Consistent



Not Accurate
but Consistent



Accurate but
Not Consistent



Not Accurate &
Not Consistent

```
:weather1-7 a dcat:Dataset ;
dct:title "Measurements from weather stations 1-7" ;
dct:description "Data from seven weather stations
                showing temparture, humidity,
                wind direction and wind speed" ;
dct:issued "2013-01-01T00:00:00+01:00" ;
dct:modified "2013-07-01T19:20:30+01:00" ;
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/CC0> ;|
dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```

Faible cohérence

```
:weather1-7 a dcat:Dataset ;
dct:title "Measurements from weather stations 1-7" ;|
dct:description "Data from seven weather stations
                showing temparture, humidity,
                wind direction and wind speed" ;
dct:issued "2014-01-01T00:00:00+01:00" ;
dct:modified "2013-07-01T19:20:30+01:00" ;
dct:publisher <http://myweather.com/id/myweather> ;
dcat:keyword "weather" ;
dcat:landingpage <http://myweather.com/stations1-7.html> ;
dcat:distribution :weatherdata-xlsx
.

ERROR INCONSISTENT DATA: Issue date is after modification date

:weatherdata1-7-xlsx a dcat:Distribution ;
dct:format <http://publications.europa.eu/resource/authority/file-type/XLSX> ;
dct:licence <http://creativecommons.org/licenses/CC0> ;
dct:licence <http://creativecommons.org/licenses/BY/3.0> ;
dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.

ERROR INCONSISTENT DATA: Licence element repeated
```

Disponibilité

La disponibilité des données est la mesure dans laquelle elles peuvent être consultées, ce qui inclut également la persistance à long terme des données.

Par exemple:

- Un ensemble de données qui est identifié par un http:URI qui renvoie la bonne ressource (et ne redonne pas 404 Not found).
- Une description de l'ensemble de données qui est inclus dans le moteur de recherche d'un portail de données.

Recommandations :

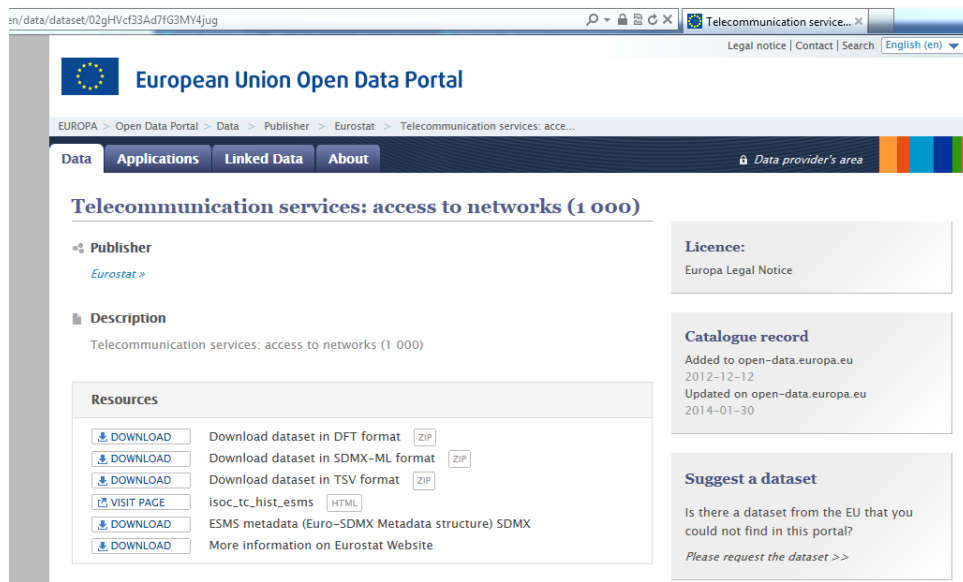
- Suivez les **meilleures pratiques** pour l'affectation et la maintenance des URIs.
- Assurez-vous que la responsabilité de la maintenance des données est clairement attribuée dans l'organisation.

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/de-sign-and-manage-persitent-uris>

Exemple de disponibilité

Très grande disponibilité



European Union Open Data Portal

EUROPA > Open Data Portal > Data > Publisher > Eurostat > Telecommunication services: access to networks (1 000)

Telecommunication services: access to networks (1 000)

Publisher
Eurostat >

Description
Telecommunication services: access to networks (1 000)

Resources

- Download dataset in DFT format
- Download dataset in SDMX-ML format
- Download dataset in TSV format
- isoc_tc_hist_esms
- ESMS metadata (Euro-SDMX Metadata structure) SDMX
- More information on Eurostat Website

Licence:
Europa Legal Notice

Catalogue record
Added to open-data.europa.eu 2012-12-12
Updated on open-data.europa.eu 2014-01-30

Suggest a dataset
Is there a dataset from the EU that you could not find in this portal?
[Please request the dataset >>](#)

Moindre disponibilité

The page cannot be found

The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

Please try the following:

- If you typed the page address in the Address bar, make sure that it is spelled correctly.
- Open the www.shawnandrews.ca home page, and then look for links to the information you want.
- Click the [Back](#) button to try another link.

HTTP 404 - File not found
Internet Information Services

Technical Information (for support personnel)

- More information:
[Microsoft Support](#)

Exhaustivité

L'exhaustivité des données est la mesure dans laquelle elles incluent les éléments de données ou points de données qui sont nécessaires pour soutenir l'application pour laquelle elles sont destinées.

Par exemple:

- Un ensemble de données qui comprend des données sur les dépenses pour tous les ministères permet d'avoir une vue complète des dépenses publiques.
- Une description des données qui sont générées en temps réel qui comprend la date et l'heure de dernière modification.

Recommandations:

- **Concevoir le processus de captage et de publication** pour inclure les points de données nécessaires.
- **Surveiller** les mécanismes de mise à jour sur une base continue.

Exemple d'exhaustivité

Grande exhaustivité

```
:weather1-7 a dcat:Dataset ;
  dct:title "Measurements from weather stations 1-7" ;
  dct:description "Data from seven weather stations
    showing temparture, humidity,
    wind direction and wind speed" ;
  dct:modified "2013-07-01T19:20:30+01:00" ;
  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/CC0> ;
  dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```

Moindre exhaustivité

```
:weather1-7 a dcat:Dataset ;
  dct:title "Measurements from weather stations 1-7" ;
  dct:description "Data from seven weather stations
    showing temparture, humidity,
    wind direction and wind speed" ;
  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/CC0> ;
  dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```

ERROR: MISSING DATA dct:modified

Conformité

La conformité des données est la mesure dans laquelle elle suivent un ensemble de règles ou de normes explicites pour le captage, la publication et la description.

Par exemple:

- Un ensemble de données qui exprime des coordonnées en WGS84 et des statistiques en SDMX.
- Une description d'un ensemble de données suivant le « DCAT Application Profile ».

Recommandations:

- **Appliquer les normes les plus** utilisées dans le domaine qui est le plus pertinent pour les données ou métadonnées.
- **Définir des vocabulaires locaux si aucune norme** n'est disponible, mais publier vos vocabulaires conformément aux meilleures pratiques (par exemple avec des URIs déréférençables).

Exemple de conformité

Forte conformité

```
:weather1-7 a dcat:Dataset ;
  dct:title "Measurements from weather stations 1-7" ;
  dct:description "Data from seven weather stations
    showing temparture, humidity,
    wind direction and wind speed" ;
  dct:modified "2013-07-01T19:20:30+01:00" ;
  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/CC0> ;
  dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```

Faible conformité

```
:weather1-7 a dcat:Dataset ;
  dct:description "Data from seven weather stations
    showing temparture, humidity,
    wind direction and wind speed" ;
  dct:modified "2013-07-01T19:20:30+01:00" ;
  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/CC0> ;
  dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```

ERROR MISSING MANDATORY ELEMENT dct:title

Voir aussi:

https://joinup.ec.europa.eu/asset/adms_foss/news/just-released-admssw-validator-verify-and-visualise-rdf-software-metadata

Crédibilité

La crédibilité des données est la mesure dans laquelle elles sont basées sur des sources fiables ou sont fournies par des organismes fiables.

Par exemple:

- Un ensemble de données qui contient des données de processus qui peuvent être vérifiées de façon indépendante, par exemple les résultats des élections ou des procédures parlementaires.
- Une description d'un ensemble de données qui est publiée par une agence gouvernementale.

Recommandations:

- **Les données devraient être basées sur des sources auxquelles on peut faire confiance** ou sur des accords de niveau de service explicites lorsque cela est possible et approprié.
- **Créer des attributions appropriées** pour que les réutilisateurs puissent déterminer si oui ou non ils peuvent faire confiance aux données.

Exemple de crédibilité

Très crédible

Les données provenant de l'Office des publications de l'UE:

```
<skos:ConceptScheme at:table.version.number="2013-05-29 14:01:09" at:table.id="language"
rdf:about="http://publications.europa.eu/resource/authority/language">
  <rdfs:label>Languages Authority Table</rdfs:label>
  <at:prefLabel xml:lang="en">Languages Authority Table</at:prefLabel>
</skos:ConceptScheme>
<skos:Concept rdf:about="http://publications.europa.eu/resource/authority/language/ENG" at:p
skos:inScheme rdf:resource="http://publications.europa.eu/resource/authority/language"/>
  <at:authority-code>ENG</at:authority-code>
  <at:op-code>ENG</at:op-code>
  <atold:op-code>ENG</atold:op-code>
  <dc:identifier>ENG</dc:identifier>
  <at:start.use>1950-05-09</at:start.use>
  <skos:prefLabel xml:lang="bg">английски</skos:prefLabel>
  <skos:prefLabel xml:lang="cs">angličtina</skos:prefLabel>
  <skos:prefLabel xml:lang="da">engelsk</skos:prefLabel>
  <skos:prefLabel xml:lang="de">Englisch</skos:prefLabel>
  <skos:prefLabel xml:lang="el">αγγλικά</skos:prefLabel>
```

The Metadata Registry is maintained by the [Publications Office of the EU](#).

Peu crédible

Donnée provenant de Lexvo:

```
- <rdf:Description rdf:about="http://lexvo.org/id/iso639-3/eng">
  <rdf:type rdf:resource="lvont:Language"/>
  <rdfs:comment xml:lang="en" rdf:datatype="xsd:string">English is a West
  Germanic language that arose in the Anglo-Saxon kingdoms of England and
  spread into what was to become south-east Scotland under the influence of
  the Anglian medieval kingdom of Northumbria. Following the extensive
  influence of Great Britain and the United Kingdom from the 18th century, via
  the British Empire, and of the United States since the mid-20th century, it
  has been widely dispersed around the world, becoming the leading language
  of international discourse and the lingua franca in many regions. It is widely
  learned as a second language and used as an official language of the
  European Union and many Commonwealth countries, as well as in many
  world organisations. It is the third most natively spoken language in the
  world, after Mandarin Chinese and Spanish.</rdfs:comment>
  <rdfs:label xml:lang="aa" rdf:datatype="xsd:string">English</rdfs:label>
  <rdfs:label xml:lang="ace" rdf:datatype="xsd:string">Bahsa Inggris</rdfs:label>
  <rdfs:label xml:lang="af" rdf:datatype="xsd:string">Engels</rdfs:label>
  <rdfs:label xml:lang="agq" rdf:datatype="xsd:string">Kingele</rdfs:label>
  <rdfs:label xml:lang="aii" rdf:datatype="xsd:string">ܐܢܓܠܝܫܐ</rdfs:label>
  <rdfs:label xml:lang="ak" rdf:datatype="xsd:string">Borfo</rdfs:label>
  <rdfs:label xml:lang="ak" rdf:datatype="xsd:string">English</rdfs:label>
```

Rights: Lexvo.org is Copyright © 2008-2012 [Gerard de Melo](#). All rights reserved.

Liability for Contents: We make every reasonable effort to ensure that the content of Lexvo.org is accurate and up-to-date. Nevertheless, the possibility of errors and inaccuracies cannot be ruled out. We do not give any warranty with respect to the information provided from Lexvo.org being accurate, up-to-date, or complete. We disclaim all liability for material or non-material loss or damage arising directly or indirectly from the use of our services.

Données Lingvoj / Lexvo peuvent ne pas être de moins bonne qualité que les données de l'Office des Publications, mais ce dernier est une source officiel de référence, alors que Linvoj et Lexvo sont des initiatives individuelles.

Traitabilité

La traitabilité de données est la mesure dans laquelle elles peuvent être comprises et traitées par des procédés automatisés.

Par exemple:

- Un ensemble de données qui contient des informations codées sur la base de vocabulaires contrôlés et de listes de codes disponibles publiquement.
- Une description d'un ensemble de données qui exprime les dates dans le format; « Date » et « Time » de W3C (par exemple 2013-06-01) plutôt que sous forme de texte (par exemple 1 Juin 2013).

Recommandations:

- **Identifier la source de la terminologie et des codes utilisés** dans les données de manière à être lisible par machine.
- **Appliquer les recommandations** pour la syntaxe des données contenues dans des normes communes et des profils applicatifs.

Exemple de traitabilité

Grande traitabilité

```
▼<recipe>
  <script/>
  <script/>
  <title>Hippie Pancakes</title>
  ▼<recipeinfo>
    <blurb>Socially conscious breakfast food.</blurb>
    <author>David Horton</author>
    <yield>12 to 16 small pancakes, enough for two hippies</yield>
    <preptime>10 minutes</preptime>
  </recipeinfo>
  ▼<ingredientlist>
    ▼<ingredient>
      <quantity>1</quantity>
      <unit>C.</unit>
      <fooditem>unbleached wheat blend flour</fooditem>
    </ingredient>
    ▼<ingredient>
      <quantity>2</quantity>
      <unit>tsp.</unit>
      <fooditem>baking powder</fooditem>
    </ingredient>
    ▼<ingredient>
      <quantity>1</quantity>
      <unit>tsp.</unit>
      <fooditem>unrefined sugar</fooditem>
    </ingredient>
    ▼<ingredient>
      <quantity>1/4</quantity>
      <unit>tsp.</unit>
      <fooditem>coarse kosher salt</fooditem>
    </ingredient>
    ▼<ingredient>
      <quantity>1</quantity>
      free-range egg
      ..
  </ingredientlist>
```

Moindre traitabilité

Hippie Pancakes

Socially conscious breakfast food.
Recipe by: David Horton
Yield: 12 to 16 small pancakes, enough for two hippies
Preptime: 10 minutes

Ingredients

1 C. unbleached wheat blend flour
2 tsp. baking powder
1 tsp. unrefined sugar
1/4 tsp. coarse kosher salt
1 free-range egg
1 1/4 C. hormone-free milk
1 tsp. organic vegetable oil

Preparation Instructions

Pre-heat griddle over medium heat. Combine dry ingredients in a mixing bowl. Stir in egg, milk and oil. Use a large spoon or gravy ladle to transfer pancake batter to the griddle. Pancakes are ready to flip when large bubbles can be seen on top.

Serving Instructions

Top with fruit and berries or serve with traditional maple syrup.

Pertinence

La pertinence des données est la mesure dans laquelle elles contiennent l'information nécessaire pour appuyer l'application.

Par exemple:

- Un ensemble de données qui contient des mesures de température arrondis aux degrés Celsius pour les calculs climatiques; un ensemble de données avec une précision d'un millième de degré pour les réactions chimiques.
- Une description d'un ensemble de données qui contient uniquement les données temporelles nécessaires à leur traitement.

Recommandations:

- **Atteindre la couverture et la granularité** des données pour l'usage prévu compte tenu des contraintes de temps et d'argent.
- Cependant, considérez également **les possibilités d'usages futurs** des données.

Exemple de pertinence

Très pertinent

	Engine (cm3)	Fuel type	CO2 (g/km)	Tax (%)
Car Type 1	900	Gasoline	90	0
Car Type 2	1.100	Gasoline	120	5
Car Type 3	1.300	Gasoline	125	5
Car Type 4	1.400	Gasoline	150	5
Car Type 5	1.800	Diesel	180	10
Car Type 6	2.200	Diesel	190	10
Car Type 7	2.500	Gasoline	210	15

Peu pertinent

	Engine (cm3)	Fuel type	CO2 (g/km)	Color	Tax (%)
Car Type 1	900	Gasoline	90	Red	0
Car Type 2	1.100	Gasoline	120	Silver	5
Car Type 3	1.300	Gasoline	125	Black	5
Car Type 4	1.400	Gasoline	150	White	5
Car Type 5	1.800	Diesel	180	Silver	10
Car Type 6	2.200	Diesel	190	Blue	10
Car Type 7	2.500	Gasoline	210	Black	15

Ponctualité

La ponctualité des données est la mesure dans laquelle elles reflètent correctement l'état actuel de l'entité ou de l'événement et la mesure dans laquelle les données (dans leur dernière version) sont mises à disposition.

Par exemple:

- Un ensemble de données qui contient les données de trafic en temps réel qui est mis à jour toutes les quelques minutes.
- Une description d'un ensemble de données contenant les statistiques annuelles de la criminalité qui est mise à disposition dans les jours suivants la publication de l'ensemble de données.

Recommandations:

- **Adapter la fréquence de mise à jour des données à la nature des données et leur utilisation prévue.**
- Assurez-vous que **des processus et des outils** sont en place pour soutenir les mises à jour.

Exemple de ponctualité

Très grande ponctualité

NOAA's National Weather Service weather.gov

Pacific Tsunami Warning Center


Home News Organization Search for: NWS All NOAA

[DOC](#) > [NOAA](#) > [NWS](#) > [PTWC](#)

Click on the tabs below to see tsunami messages relevant to each of PTWC's [areas of responsibility](#).

[All Regions](#) [Pacific Ocean](#) [Hawai'i](#) [Caribbean Sea](#) [Indian Ocean \(discontinued\)](#)

Page last loaded at: Tue, 03 Dec 2013 15:10:09 UTC.



No Current Warning, Watch, or Advisory in Effect

low [Click here to read the latest tsunami message.](#)

Tsunami Messages for All Regions (Past 30 days)

Click on the map or table below for more information.

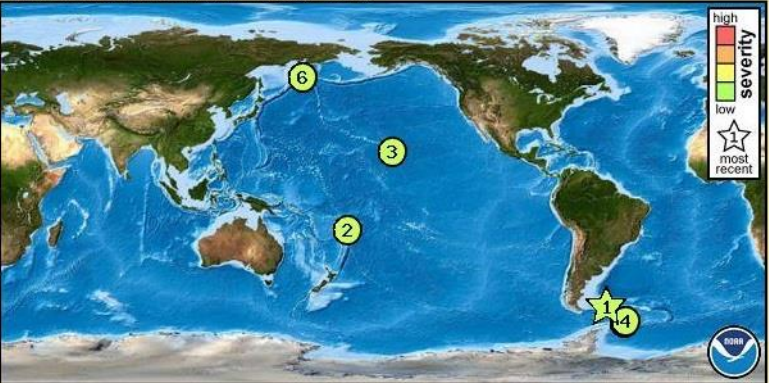


Image last created on Tue, 03 Dec 2013 15:09:23 UTC.

Moindre ponctualité



Les meilleures pratiques

*Meilleures pratiques pour la publication de données et
métadonnées de haute qualité.*

W3C: Meilleures pratiques pour la publication de Données Ouvertes Liées

[IDENTIFY](#) Identify data sets that other people may wish to re-use.

[MODEL](#) Model the data in an application-independent, objective way in terms of representation. Denormalize the data as necessary.

[METADATA](#) Provide basic metadata, including MIME type, publishing organization and/or agency, creation date, modification date, version, frequency of updates, contact email for the data steward(s).

[PII](#) Do not Publish Personally Identifiable Information as Open Data on the Web Data on the public Web can be potentially misused. Examples of personally identifiable data include: individual names, national identification number, phone number, credit card number and driver license number.

[NAME](#) Use HTTP URIs as names for your objects. Give careful consideration to the URI naming strategy. Consider how the data will change over time and name as necessary.

[STANDARD_VOCABULARIES](#) Describe objects with standard vocabularies whenever possible.

[VOCABULARY_USE](#) Use vocabularies as loosely coupled modular components.

[REPRESENTATION](#) Convert the source data into a Linked Data representation, also called an RDF serialization including Turtle, Notation-3 (N3), N-Triples, XHTML with embedded RDFa, and RDF/XML.

[HUMAN_READABLE](#) Provide human readable descriptions with your Linked Data.

[MACHINE_ACCESSIBLE](#) Provide access to the data representation via RESTful API, SPARQL endpoint(s) and RDF download.

[SPECIFY_LICENSE](#) Specify an appropriate license.

[HOST](#) Deliver open government data on authoritative domain to increase perceived trust.

[ANNOUNCE](#) Announce open government data, have a feedback mechanism and be prepared to be responsive to feedback.

[SOCIAL_CONTRACT](#) Maintenance is critical. Without a permanent identifier scheme, if you move or remove data that is published to the Web, you may break third party applications or mashups which is clearly undesirable. URI strategy and implementation are critical.

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/the-linked-open-government-data-lifecycle>

Opquast: 72 bonnes pratiques pour les données ouvertes

Quelques exemples

Metadata	1	23	Each dataset is accompanied by a descriptive record
Metadata	1	24	Each dataset includes at least a title and a description
Metadata	1	25	A creation date is given for each dataset
Metadata	1	26	A last-updated date is given for each dataset
Metadata	2	27	The datasets are categorised
Format	1	33	Each dataset includes a reference to the charset used
Format	1	34	The format of downloadable files is indicated
Format	1	35	Dates are given in documented formats
Format	2	36	Dates are available in a standard format
Format	2	37	Data is provided in at least one open format

License	1	47	The datasets are accompanied by a licence
License	1	48	The licence sets out the conditions of attribution, reuse, redistribution and commercialisation
License	2	49	Usage rights are provided for an unlimited period
License	2	50	Data producers declare their policy on releasing data
License	3	51	The datasets are accompanied by a summary and a link to the full version of the licence
Linkeddata	2	52	Any vocabularies used within the dataset are identified and documented
Linkeddata	3	53	Data adheres to the defined syntax of any specified vocabularies
Linkeddata	3	54	It is possible to query data and metadata in accordance with standards of the web of data (Linked Open Data)

Voir aussi:

<http://checklists.opquast.com/en/opendata>

Quels sont les éléments communs qu'on peut trouver dans les meilleures pratiques?

- **Fournir des descriptions appropriées** aux données (i.e. métadonnées).
- **Utiliser vocabulaires standards** pour les métadonnées et les données chaque fois que de tels vocabulaires existent.
- **Spécifier la licence** sous laquelle les données peuvent être réutilisées.
- **Respecter les exigences légales** en matière de protection des données personnelles et autres données sensibles.
- **Représenter** les métadonnées et les données **selon les principes de données liées en utilisant des URIs** persistantes pour identifier des ressources.
- **Fournir des informations sur la source** des données.

La maintenance des métadonnées et des données est essentielle!

Conclusions

- La qualité des données est déterminée par leur aptitude à être (re-)utilisées par les consommateurs de données.
- Les métadonnées sont des "données sur les données", c'est à dire que les métadonnées sont un type de données.
 - Les mêmes considérations s'appliquent à la qualité des données et des métadonnées.
- La qualité des données a des dimensions multiples et ceci s'étend au-delà de l'exactitude des données.
 - Exactitude, disponibilité, exhaustivité, conformité, cohérence, crédibilité, traitabilité, pertinence, pontualité.

Questions de groupe



<http://www.visualpharm.com>

D'après vous, quels sont les facteurs qui contribuent le plus à la qualité des métadonnées?



<http://www.visualpharm.com>

L'amélioration de la qualité peut demander du temps et des ressources. Dans quelle mesure votre organisation serait-elle prête à investir dans la qualité des métadonnées?



<http://www.visualpharm.com>

Pouvez-vous donner un exemple de métadonnées de haute qualité pour un ensemble de données utilisant le profil applicatif DCAT?

Prenez le test en ligne [ici!](#)

Merci!

...et maintenant vos questions?

Références

Slide 5:

- Juran, Joseph M. and A. Blanton Godfrey, Juran's Quality Handbook, Fifth Edition, p. 2.2, McGraw-Hill, 1999

Slide 6:

- National Information Standards Organization, <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Slide 8-26:

- Mark David Hansen. Zero Defect Data: Tackling the Corporate Data Quality Problem. 1991. <http://dspace.mit.edu/handle/1721.1/13812>
- Kevin Roebuck. Data Quality: High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors. Emereo Pty Limited, 2011. <http://bit.ly/19Qb6Ov>
- Thomas R. Bruce, Diane Hillmann. The Continuum of Metadata Quality: Defining, Expressing, Exploiting. ALA Editions, 2004. <http://www.ecommons.cornell.edu/handle/1813/7895>
- Sharon Dawes. Open data quality: a practical view. Open Data Roundtable. October 2012. <http://www.slideshare.net/cityhub/sharon-dawes-ctg>
- Joshua Tauberer. Open Government Data. Section 5.2 Data Quality: Precision, Accuracy, and Cost. June 2012. <http://opengovdata.io/2012-02/page/5-2/data-quality-precision-accuracy-and-cost>
- Stefan Urbanek. Data Quality: What is It? January 2011. <http://ckan.org/2011/01/20/data-quality-what-is-it/>
- Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, Sören Auer. Quality Assessment Methodologies for Linked Open Data. Semantic Web Journal (unpublished), 2012. <http://www.semantic-web-journal.net/content/quality-assessment-methodologies-linked-open-data>

Slide 13:

- ISA Programme. 10 Rules for Persistent URIs. <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

Slide 14:

- European Commission. Telecommunication services: Access to networks (1 000). <http://open-data.europa.eu/>

Slide 28:

- W3C. Best Practices for Publishing Linked Data. W3C Note 06 June 2013. <https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>

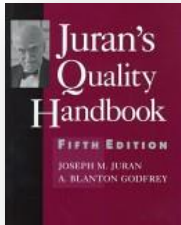
Slide 29:

- OPQUAST. 72 Open data good practices. <http://checklists.opquast.com/en/opquast>

Lecture supplémentaire



Joshua Tauberer. Open Government Data. <http://opengovdata.io/>



Juran, Joseph M. and A. Blanton Godfrey, Juran's Quality Handbook

Projets et initiatives apparentés



Best Practices for Publishing Linked Data.

<https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html>



OPQUAST. Open data good practices.

<http://checklists.opquast.com/en/opendata>



Eurostat. European Statistical System

http://epp.eurostat.ec.europa.eu/portal/page/portal/ess_eurostat/introduction

Be part of our team...

Find us on



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

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



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

<http://goo.gl/y9ZZI>

Follow us



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

Join us on



joinup

<http://www.opendatasupport.eu>

Contact us

contact@opendatasupport.eu