



OPEN DATA SUPPORT

Training Module 2.2

Open Data & Metadata Quality



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).

© 2014 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.

Learning objectives

By the end of this training module you should have an understanding of:

- What (open) data quality means.
- The open data quality determinants and criteria.
- Good practices for publishing high-quality (linked) open data.

Content

This module contains...

- A definition of data quality;
- An overview of the dimensions of data and metadata quality;
- A selection of best practices for publishing good quality data and metadata.

What is data (and metadata) quality?

Data is of high quality "if they are fit for their intended uses in operations, decision making and planning."

Or more specifically:

"High quality data are accurate, available, complete, conformant, consistent, credible, processable, relevant and timely."

Metadata is data about data...

“Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data”

-- National Information Standards Organization

- We observe that metadata is a type of data.
- The same quality considerations apply to data and metadata alike.

The data quality dimensions

What are the main dimensions to be taken into account for delivering good quality (meta)data?

Data quality dimensions

- **Accuracy:** is the data correctly representing the real-world entity or event?
- **Consistency:** Is the data not containing contradictions?
- **Availability:** Can the data be accessed now and over time?
- **Completeness:** Does the data include all data items representing the entity or event?
- **Conformance:** Is the data following accepted standards?
- **Credibility:** Is the data based on trustworthy sources?
- **Processability:** Is the data machine-readable?
- **Relevance :** Does the data include an appropriate amount of data?
- **Timeliness:** Is the data representing the actual situation and is it published soon enough?

Accuracy

The accuracy of data is the extent to which it correctly represents the characteristics of the real-world object, situation or event.

For example:

- Correct measurement of weather conditions (temperature, precipitation).
- Correct indication of re-use conditions of the dataset.

Recommendations:

- **Balance** the **accuracy** of your data against the **cost** in the context of the application; it needs to be **good enough for the intended use**.
- Make sure that there is **organisational commitment** and **investment in procedures and tools** to maintain accuracy .

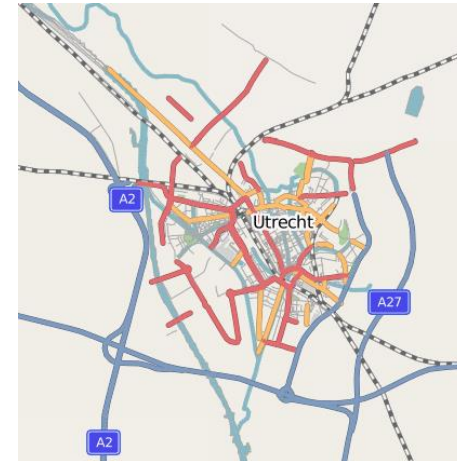
Accuracy by example

Higher accuracy

OpenStreetMap, City of Utrecht, The Netherlands (2011 vs. 2007)



Less accuracy



Consistency

The consistency of data is the extent to which it does not contain contradictions that would make its use difficult or impossible.

For example:

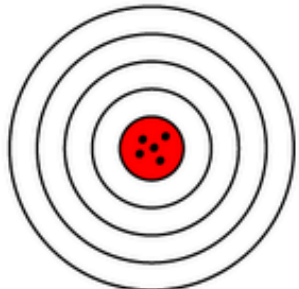
- A dataset that combines data from different sources that has been processed to detect conflicting statements which have been resolved.
- A description of a dataset that does not contain multiple licence statements or where the data of last modification is not before the creation date.

Recommendations:

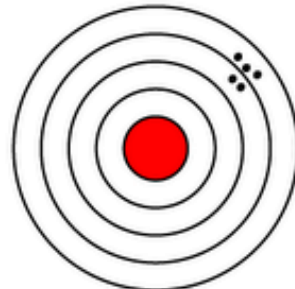
- **Process all data before publication** to detect conflicting statements and other errors (in particular if data is aggregated from different sources).

Consistency by example

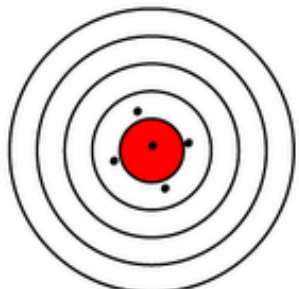
High consistency



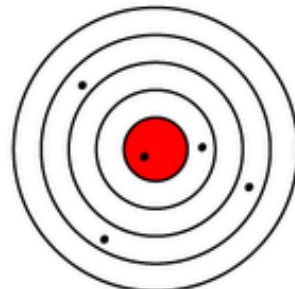
Accurate & Consistent



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>
.
```

Less consistency

```
: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
```

Availability

The availability of data is the extent to which it can be accessed; this also includes the long-term persistence of data.

For example:

- A Dataset that is identified by a http: URI that resolves persistently to the right resource (and does not give back 404 Not found).
- A description of the dataset that is included in the search engine of a data portal.

Recommendations:

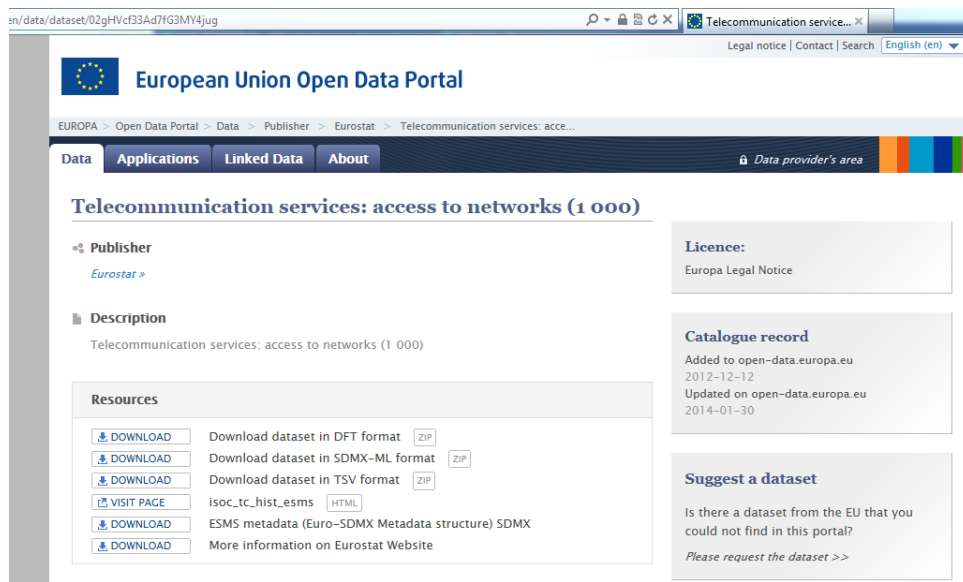
- Follow **best practices** for the assignment and maintenance of URIs.
- Make sure **that responsibility** for the maintenance of data is **clearly assigned** in the organisation.

See also:

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

Availability by example

High availability



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 >>](#)

Less availability

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](#)

Completeness

The completeness of data is the extent to which it includes the data items or data points that are necessary to support the application for which it is intended.

For example:

- A Dataset that includes spending data for all ministries enables a complete overview of government spending.
- A description of data that is generated in real time that includes the date and time of last modification.

Recommendations:

- **Design the capture and publication process** to include the necessary data points.
- **Monitor** the update mechanisms on a continuous basis.

Completeness by example

High completeness

```
: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>
.
```

Less completeness

```
: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

Conformance

The conformance of data is the extent to which it follows a set of explicit rules or standards for capture, publication and description

For example:

- A Dataset that expresses coordinates in WGS84 and statistics in SDMX.
- A description of a dataset according to the DCAT Application Profile.

Recommendations:

- **Apply the most used standards** in the domain that is most relevant for the data or metadata.
- **Define local vocabularies if no standard is available**, but publish your vocabularies according to best practice (e.g. dereferenceable URIs).

Conformance by example

High conformance

```
: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>
.
```

Less conformance

```
: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

See also:

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

Credibility

The credibility of data is the extent to which it is based on trustworthy sources or delivered by trusted organisations.

For example:

- A dataset that contains data from processes that can be independently verified, e.g. election results or parliamentary proceedings.
- A description of a dataset that is published by a government agency.

Recommendations:

- **Base data on sources that can be trusted** or on explicit Service Level Agreements where possible and appropriate.
- **Make appropriate attributions** so that re-users can determine whether or not they can trust the data.

Credibility by example

High credibility

Data coming from the **Publications Office of the EU:**

```
<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:pr
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](#).

Less credibility

Data coming from **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 Inggreh</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.

Lingvoj/Lexvo data may not be of less quality than Publications Office data, but the Publications Office is an authoritative source, while Linvoj and Lexvo are initiatives of individuals.

Processability

The processability of data is the extent to which it can be understood and handled by automated processes.

For example:

- A dataset that contains coded information based on publicly available controlled vocabularies and code lists.
- A description of a dataset that expresses dates in W3C Date and Time Format (e.g. 2013-06-01) rather than as text (e.g. 1 June 2013).

Recommendations:

- **Identify the source of terminology and codes** used in the data in machine-readable manner.
- **Apply recommendations for syntax** of data given in common standards and application profiles.

Processability by example

Higher processability

```
▼<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>
```

Less processability

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.

Relevance

The relevance of data is the extent to which it contains the necessary information to support the application.

For example:

- A Dataset that contains temperature measurements rounded to degrees Celsius for climate calculations; a dataset with precision of a thousandth of a degree for chemical reactions.
- A description of a dataset that only contains temporal coverage data if necessary for its processing .

Recommendations:

- **Match coverage and granularity** of data to its intended use within constraints of available time and money.
- However, also **consider potential future usages** of the data.

Relevance by example

High relevance

	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

Less relevance

	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

Timeliness

The timeliness of data is the extent to which it correctly reflects the current state of the entity or event and the extent to which the data (in its latest version) is made available without unnecessary delay

For example:

- A dataset that contains real-time traffic data that is refreshed every few minutes.
- A description of a dataset containing annual crime statistics that is made available within days of publication of the dataset.

Recommendations:

- **Adapt the update frequency of data to the nature of the data and its intended use.**
- Make sure that **processes and tools are in place** to support the updating.

Timeliness: examples

High timeliness

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.

Less timeliness



Best practices

Best practices for publishing high-quality data and metadata.

W3C: Best practices for publishing Linked Open Data

[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.

See also:

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

Opquast: 72 Open data good practices

Some examples

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)

See also:

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

What are the common elements in the best practices

- **Provide appropriate descriptions** of data (i.e. metadata).
- **Use standard vocabularies** for metadata and data whenever such vocabularies exist.
- **Specify the license** under which the data may be re-used.
- **Adhere to legal requirements** concerning protection of personal and other sensitive data.
- **Represent** metadata and data **according to the Linked Data principles** using **persistent URIs** for identifying things.
- **Provide information about the source** of the data.

Maintenance of metadata and data is critical!

See also:

<http://www.slideshare.net/OpenDataSupport/introduction-to-metadata-management>

Conclusions

- The quality of data is determined by its fitness for (re-)use by data consumers.
- Metadata is “data about data”, i.e. metadata is a type of data.
 - The same quality considerations apply to data and metadata alike.
- Data quality has multiple dimensions and is about more than the correctness of data.
 - Accuracy, availability, completeness, conformance, consistency, credibility, processability, relevance, timeliness.

Group questions



<http://www.visualpharm.com>

In your opinion, which factors contribute the most to metadata quality?



<http://www.visualpharm.com>

Improving quality can require time and resources. To which extent would your organisation be willing to invest in metadata quality?



<http://www.visualpharm.com>

Could you provide an example of high quality metadata description for a dataset using the DCAT Application Profile?

Take also the online test [here!](#)

Thank you!
...and now YOUR questions?

References

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>

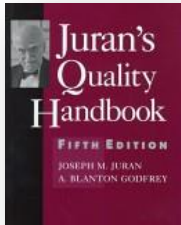
Slide 29:

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

Further reading



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



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

Related projects and initiatives



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