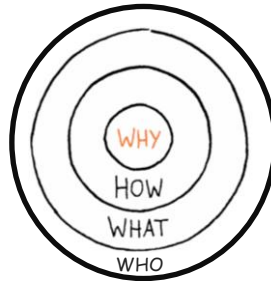


New common method for declaring data quality in Denmark

Lars Erik Storgaard (laers@sdfe.dk)
Jan Hjelmager (jnh@sdfe.dk)

Agenda

- Recap from Malta
- New strategy in our agency
- Implementing the strategy
- Questions



Why?

Need for authorities to talk same language when assessing the quality of data from widely different domains. The goal is to make better use of data across the public sector.

How?

By using four quality elements, the new data quality declaration should enable a better and more secure way to have a dialogue about data quality across the public sector.

What?

A common public declaration template and undelaying vocabulary.

Who?

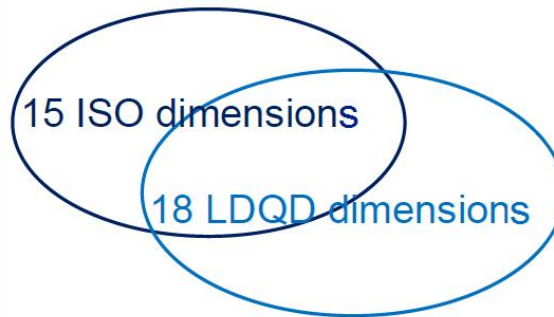
Public authorities and private companies are on voluntary basis welcome to use the declaration method.

Data on the Web Best Practices: Data Quality Vocabulary



W3C Working Group Note 15 December 2016

<https://www.w3.org/TR/vocab-dqv>





The fundament:

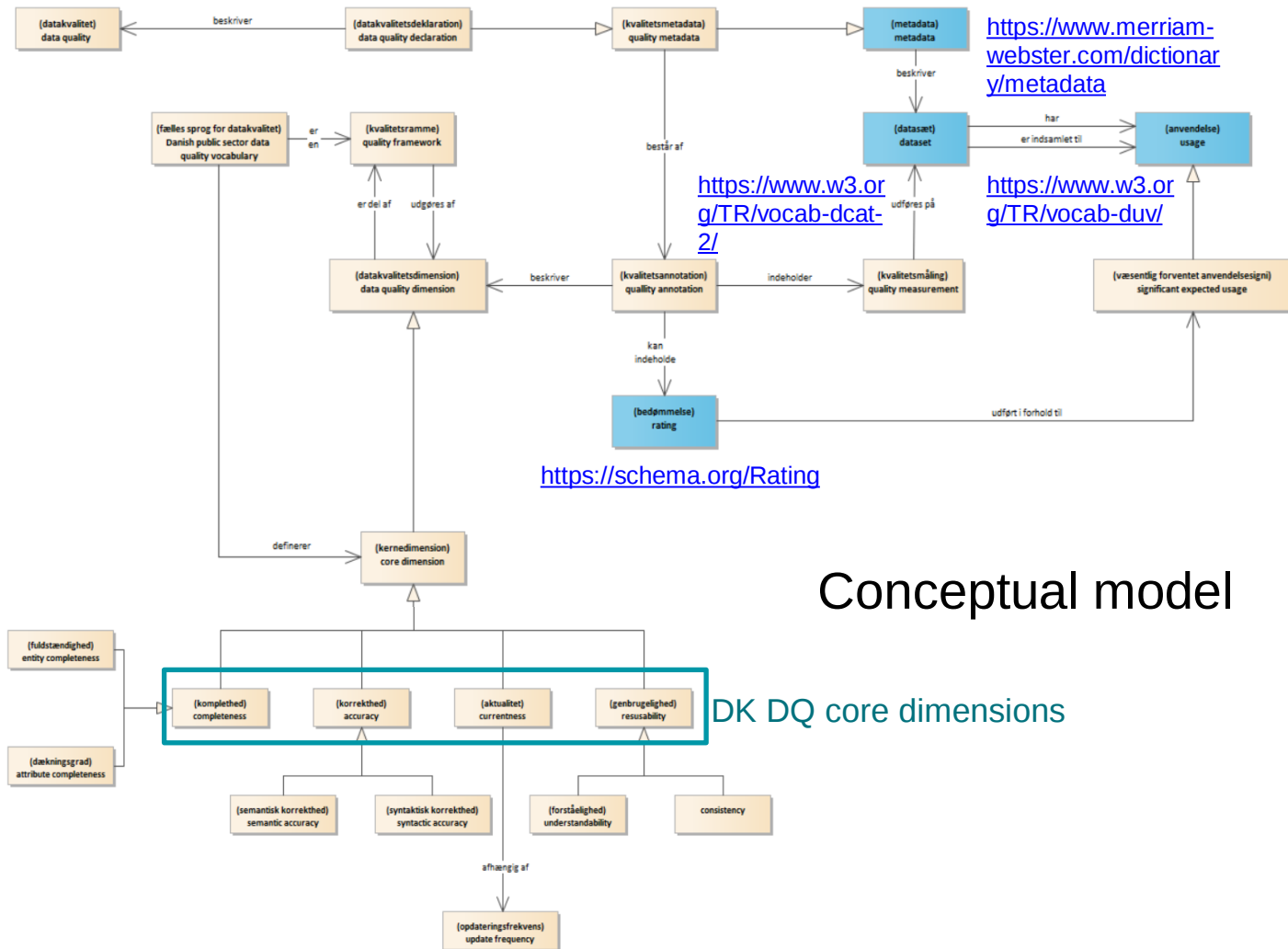
ISO 25012 og Linked Data Quality Dimensions – this is the two sets of DQ dimensions from W3Cs Data Quality Vocabulary)

Four quality dimensions is chosen – named "core dimensions" – as the most important to evaluate the usability of a data set including new purposes

DQ dimension	Definition
completeness	the degree to which the data set contains the data elements that are expected based on the specification of the data set
accuracy	the degree to which data values corresponds to actual values
currentness	indicating to which degree data is current
Reusability	the degree to which data is understandable and without difficulties can be used by others

Mapping to ISO 25012 and Linked DQ

"DK" DQ dimension	ISO 25012 dimension	Linked Data Quality dimension
completeness	skos:narrowMatch	skos:exactMatch
accuracy	skos:exactMatch	skos:narrowMatch
currentness	skos:exactMatch	skos:exactMatch
reusability	skos:narrowMatch	skos:narrowMatch



Conceptual model

DK DQ core dimensions

The declaration template

The template has two fields for each quality dimension:

- 1) Measurements and descriptions and
- 2) Assessment.

Measurement and description is meant to provide objective information describing the data set in relation to the relevant dimension.

In **Assessment**, the data owner provides assessment of how good the quality of the data set is in relation to the relevant dimension when it is assessed in relation to the main use case (application).

☆☆☆☆☆	Rated unsuitable for the given application
★☆☆☆☆	Is judged to be poorly used
★★☆☆☆	Can be used subject to a greater number of errors / unintended results
★★★☆☆	Can be used subject to errors / unintended results
★★★★☆	Can be used subject to individual errors / unintended results
★★★★★	Can be used immediately and unconditionally

The declaration template

DQ Dimension	Measurement and description <i>Describe the quality of the dataset in relation to each dimension.</i>	Assessment in relation to use <i>Fill in the appropriate number of stars.</i>
completeness		☆☆☆☆☆
accuracy		☆☆☆☆☆
currentness		☆☆☆☆☆
reusability		☆☆☆☆☆

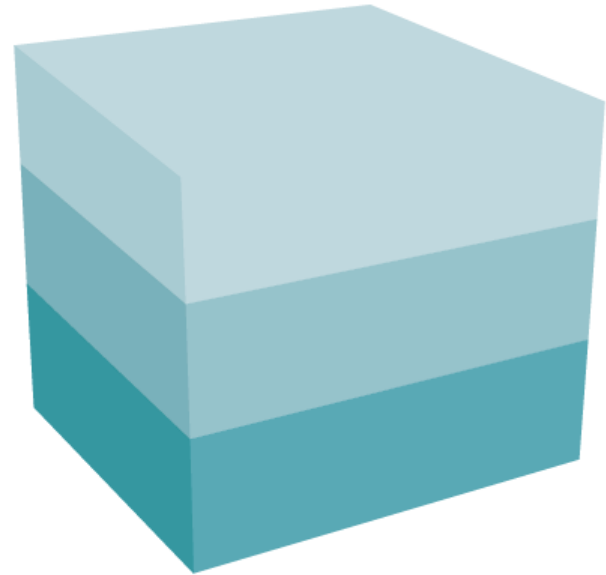
Empirical method – Data set "Basemap DK-Building"

Dimension	Measures and descriptions <i>Describe the quality of the dataset in relation to each dimension.</i>	Assessment in relation to use <i>Fill in the appropriate number of stars</i>
Completeness	<i>What proportion of the individuals' dataset describing according to its specification is currently represented in the data set? - Is this an estimate?</i> The completeness check is carried out by sampling in different numbers of samples and the number of grids. The inspection is performed as a visual inspection of the use of spring orthophoto from the year in which the data is updated (support data for centerline of rivers and lake from two additional years of spring orthophoto). SDFE, based on DS / EN ISO 19157: 2014, designates objects for the samples. 99 percent of buildings visible in spring orthophoto from 2017 are represented in the topographic basic data set building theme. The percentage is calculated from random checks carried out in August 2018 according to DS / EN ISO 19157: 2014. The GeoDanmark specification allows for the completeness of 99 percent of buildings visible in Orthophoto.	★★★★☆

New strategy for SDFE

We ensure fit for purpose data quality in registers

We prioritise **three levels of data quality** in registers so that the **currency, history and level of detail** is precisely defined and matched to its intended use. We also ensure that the appropriate datasets are ready for the technologies of the future.



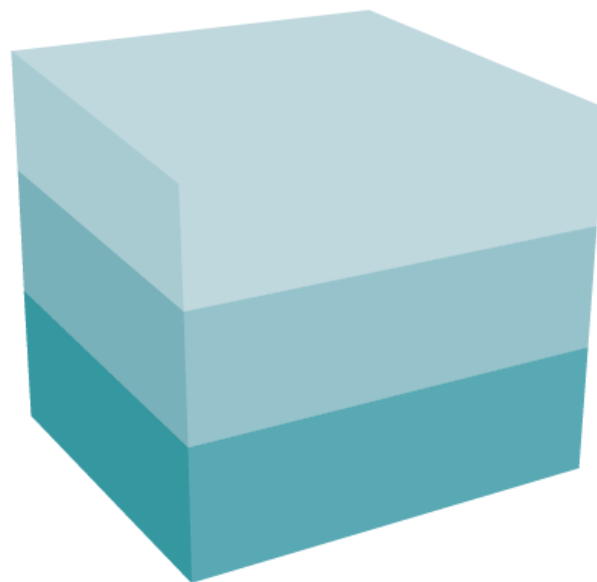
Implementing the new strategy

We ensure fit for purpose
data quality in registers

We prioritise three levels of data quality in registers so that the currency, history and level of detail is precisely defined and **matched to its intended use**. We also ensure that the appropriate datasets are ready for the technologies of the future.

Requirement:

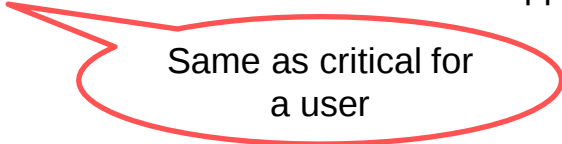
- Data product specification shall be public available



Questions

We ensure fit for purpose data quality in registers

We prioritise three levels of data quality in registers so that the currency, history and level of detail is precisely defined and **matched to its intended use**. We also ensure that the appropriate datasets are ready for the technologies of the future.



Same as critical for a user

Questions:

1. How to define criticality for intended use?
2. Should all core dataset be treated as critical to all users?
3. Can dataset have different data quality description for different feature classes? - If yes how do you produce a quality description/markings for the entire dataset?
4. How do you communicate quality in different levels to the end user?