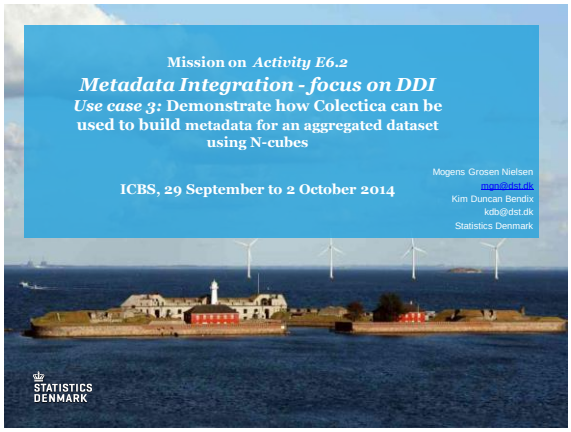


Credits, challenges on sharing data and metadata

Credits: Power Points and paper from Gesis, Leibnitz, Lars Thygesen, Statistics Denmark and Colectica reused or used for inpiration.

Challenges:

- When data and metadata are going to be shared between two or more parties using SDMX, evidently there must be an agreement on how the data need to be organised.
- There is a great degree of freedom in this as SDMX can be viewed as just a technical format that can accommodate any structure of the contents.
- Historically speaking, there has been an inclination to basing such structures on old data structures, e.g. on a paper publication that one wishes to publish, and which combines variables in a certain way that is influenced by the paper format.

Reference: *Structuring data and metadata for reusability and interoperability,* Lars Thygesen, OECD

STATISTICS DENMARK

Introduction to N-cubes (and SDMX)

However, it is important that the organisation should reflect the content of the data in a logical way. Otherwise the following problems will arise:

- it will be difficult to explain the structure to new participants in the data sharing and implement it in their systems, and misunderstandings are likely to accrue,
- it will be difficult to change the format according to data needs going beyond what was originally envisaged,
- it will be difficult to avoid duplications within the data structure and across different subject matter domains,
- it will be difficult to reuse parts of the data structure in adjacent statistical domains, and thus be able to subsequently use such data together from neighbouring domains.

Reference: *Structuring data and metadata for reusability and interoperability,* Lars Thygesen, OECD

STATISTICS DENMARK

STATISTICS DENMARK

3 Dimensional NCube

Dimensions

Measure

Gender	Is Married	Has Car	Count()
M	Yes	Yes	15
M	Yes	No	12
M	No	Yes	20
M	No	No	5
F	Yes	Yes	16
F	Yes	No	13
F	No	Yes	22
F	No	No	6

Footnotes

STATISTICS DENMARK

2 Dimensional NCube

Normal

Gender	Is Married	Count()
M	Yes	27
M	No	25
F	Yes	29
F	No	28

Tabular

	Married	Not Married
M	27	25
F	29	28

STATISTICS DENMARK

Properties of an Aggregate

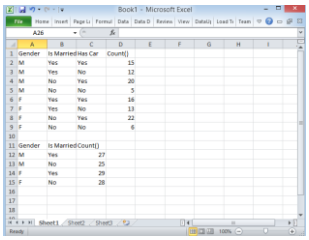
- Dimensions
- Measures
- Attributes
  - Can append footnotes to the aggregate
  - Attach to the overall structure or to individual cells or to groups of cells

STATISTICS DENMARK

STATISTICS DENMARK

Data Structure

- In a data file, only one NCube per record



Gender	Is Married	Has Car	Count()
M	Yes	Yes	15
M	Yes	No	12
M	No	Yes	20
M	No	No	5
F	Yes	Yes	16
F	Yes	No	13
F	No	Yes	22
F	No	No	6

Gender	Is Married	Count()
M	Yes	27
M	No	25
F	Yes	29
F	No	28

This isn't too well structured

STATISTICS DENMARK

Intuitive Structure

- NCube: re-usable definition of an aggregate structure
  - Dimensions
    - ordered list of Variable references
  - Measures
    - List of measures for each intersection of Dimensions
    - Variable reference
    - Type (count, %, mean, etc.)
  - Attributes
    - Attributes that are applicable to re-usable NCube definition

STATISTICS DENMARK

- DEMO

Metadata for an aggregated dataset using N-cubes

STATISTICS DENMARK 9

STATISTICS DENMARK

STATISTICS DENMARK