The Inevitability of Calibration in VGI Quality Assessment

Franz-Benjamin Mocnik
Heidelberg University
• Where do data quality issues originate from? 
   (approaching the foundations of data quality)

• How does data quality and fitness for purpose relate?

• Why is calibration in data quality assessment inevitable?
Data

Mental Representation

Perception of the Environment

Action in the Environment

Processed Data
context

How was perceived?

affordances

What does the data mean?

grounding

How can the data be interpreted?

purpose

How does the data relate to the environment?
Data can …
… afford to be interpreted and used … in a given context … for a given purpose.
Data can ...

... afford to be interpreted and used ...

... in a given context

... for a given purpose.
Fitness for purpose is whether …
… a suitable **context** exists, such that data
… **afford** to be interpreted and used
… for a given **purpose**.
Fitness for purpose is whether …
… a suitable context exists, such that data … afford to be interpreted and used … for a given purpose.

The same dataset can afford to be interpreted in different ways, and for different purposes.

- fitness for purpose cannot solely depend on the data!
- fitness for purpose depends on the data, the interpretation, and the purpose.
How can we understand fitness for purpose and data quality by referring to the same terms?

(This relates fitness for purpose and data quality.)
Approach of a Definition

Data quality is the discrepancy between

• the fitness for purpose of optimal data, and

• the fitness for purpose of the actual data, aggregated for all possible purposes.
Approach

data with maximal fitness for purpose but the same scope as the actual data

Data quality is the discrepancy between
• the fitness for purpose of optimal data, and
• the fitness for purpose of the actual data, aggregated for all possible purposes.

\[ DQ(\text{data}) = \bigcup_{\text{purpose } p} (\text{FFP}_p(\text{data}_{\text{optimal}}) - \text{FFP}_p(\text{data})) \]
When we assess data quality, we compare the data to other ‘data’:

- reference data,
- data from the same dataset,
- principles,
- etc.
When we assess data quality, we compare the data to other ‘data’:

- reference data,
- data from the same dataset,
- principles,
- etc.

We assume that no data is perfect*

- also the ‘data’ we compare with is not perfect
- need/inevitability to calibrate

* There exist even good (philosophical) reasons to assume that complex geographic data is always subject to imperfection [Couclelis 2003]
Three Examples

- Context of Principles
- Historic Context
- Spatial Context
Context of Principles

Kiruna, Sweden
Historic Context

Stockholm, Sweden, 2007
Historic Context

Stockholm, Sweden, 2017
Kiruna, Sweden
Spatial Context

Kiruna, Sweden
Where do data quality issues originate from?  
(approaching the foundations of data quality)

How does data quality and fitness for purpose relate?

Why is calibration in data quality assessment inevitable?
• What does the inevitability to calibrate mean in practice?

• How can we evaluate the findings?

• What is special about the findings in the context of VGI?