Data quality control
Add value to Inspire services
Presentation

Johan Esko

Project manager ISO 19157
Inspire
Data management
Presentation
Anna Halvarsson

Project manager SDI
Inspire
Data management
Agenda

Introduction
Standards
Data quality
Example
Summary
Introduction
A year ago

Inspire conference Krakow 2010

- Inspire download service prototype
Today

Krakow 2010 - Future development

• Integrate data quality control (QC)
• Integrate with OGC CSW Metadata
Standards
Why use standards?

- Cornerstone of a SDI
- Inspire specifications
- Legible information
- Own development is expensive
From product specification to metadata

• **Data product specification**
  ISO 19131, INSPIRE Data specifications

• **Data quality**
  ISO 19113/19114/19138
  (ISO 19157 will replace the standards above)

• **Metadata**
  ISO 19115/19119/19139

Recipe: 4 eggs
4 tsp water
salt

Analyse of the process: 3 eggs
6 tsp water
salt

Declaration of contents: 3 eggs
6 tsp water
salt
Data quality
Definition of quality

“The degree to which a set of inherent characteristics fulfills requirements”

*ISO 9000:2000*
Best quality?
“Short wheelbase and minimal overhang at the front and rear, combined with high ground clearance makes it extremely well suited to continue the drive where the road ends. Indeed the Niva outperform the Porsche Cayenne, which in comparison is clumsy as an elephant with a backpack in a supermarket”.  

*(Teknikens värld, 2006)*

Comment:

“If it is the terrain capabilities that is evaluated, the Niva fulfils the quality requirements better than the Porsche.”
Best quality?

Which one of these view services delivers data with the highest positional accuracy?
Geographic information - Data Quality

Data quality can be described by:

• Data quality elements
  • Completeness
  • Topological consistency
  • Positional accuracy
  • Temporal consistency
  • Thematic accuracy

• Non-quantitative overview statements (usage, purpose, lineage)
Geographic information - Data Quality

ISO 19157 CD

Some news

- Metaquality
- Usability element
- New data quality measures
- etc

- Schedule ISO 19157 CD
  - Text sent to ISO for DIS
  - DIS 2011-06
  - FDIS 2012-06
  - IS 2013-01
Quality of data used in Inspire services

- To evaluate if the data provided by an Inspire service is of sufficient quality
- Lack of quality information:
  - can make data “useless”
  - can cause misleading results
- Inspire data specifications specifies quality elements and data quality measures
Metria’s data warehouse stores more than 2 Tb geographical data

Automatic data management processes are important
Input and refining processes

Product Specification

Quality Control

Metadata

ISO 19131
INSPIRE Data specifications

ISO 19113, 19114, 19138 (ISO 19157)

ISO 19115, 19139
Example – Data management
INSPIRE Protected Sites
Data management – Step 1

Product Specification

Quality Control

Metadata
**Requirement – Completeness commission**

### 7.1.1 Commission

Commission should be documented using the rate of excess items.

<table>
<thead>
<tr>
<th>Name</th>
<th>Rate of excess items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative name</td>
<td>–</td>
</tr>
<tr>
<td>Data quality element</td>
<td>Completeness</td>
</tr>
<tr>
<td>Data quality sub-element</td>
<td>Commission</td>
</tr>
<tr>
<td>Data quality basic measure</td>
<td>Error rate</td>
</tr>
<tr>
<td>Definition</td>
<td>Number of excess items in the dataset in relation to the number of items that should have been present.</td>
</tr>
<tr>
<td>Description</td>
<td>–</td>
</tr>
<tr>
<td>Parameter</td>
<td>–</td>
</tr>
<tr>
<td>Data quality value type</td>
<td>Real, percentage, ratio (example: 0.0189; 98.11%; 11:582)</td>
</tr>
<tr>
<td>Data quality value structure</td>
<td>–</td>
</tr>
<tr>
<td>Source reference</td>
<td>–</td>
</tr>
<tr>
<td>Example</td>
<td>–</td>
</tr>
<tr>
<td>Measure identifier</td>
<td>3 (ISO 19138)</td>
</tr>
</tbody>
</table>

**Conformance level = 0**
**Requirement – Completeness omission**

### 7.1.2 Omission

Omission should be documented using the rate of missing items.

<table>
<thead>
<tr>
<th>Name</th>
<th>Rate of missing items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative name</td>
<td>–</td>
</tr>
<tr>
<td>Data quality element</td>
<td>Completeness</td>
</tr>
<tr>
<td>Data quality sub-element</td>
<td>Omission</td>
</tr>
<tr>
<td>Data quality basic measure</td>
<td>Error rate</td>
</tr>
<tr>
<td>Definition</td>
<td>Number of missing items in the dataset in relation to the number of items that should have been present.</td>
</tr>
<tr>
<td>Description</td>
<td>–</td>
</tr>
<tr>
<td>Parameter</td>
<td>–</td>
</tr>
<tr>
<td>Data quality value type</td>
<td>Real, percentage, ratio (example: 0,0189 ; 98,11% ; 11:582)</td>
</tr>
<tr>
<td>Data quality value structure</td>
<td>–</td>
</tr>
<tr>
<td>Source reference</td>
<td>–</td>
</tr>
<tr>
<td>Example</td>
<td>–</td>
</tr>
<tr>
<td>Measure identifier</td>
<td>7 (ISO 19138)</td>
</tr>
</tbody>
</table>

**Conformance level = 0**
Overall Requirement - Data product specification passed

D.7 Aggregation Measures

In a data product specification, several requirements are set up for a product to conform to the specification. The data quality measures for this element are provided in Tables D.77 to D.81.

<table>
<thead>
<tr>
<th>Line</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Name</td>
<td>data product specification passed</td>
</tr>
<tr>
<td>2</td>
<td>Alias</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>Element name</td>
<td>usability element</td>
</tr>
<tr>
<td>4</td>
<td>Description</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>Parameter</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>Value type</td>
<td>Boolean (true if all the requirements in the referred data product specification are fulfilled)</td>
</tr>
<tr>
<td>7</td>
<td>Value structure</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>Source reference</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>Example</td>
<td>—</td>
</tr>
<tr>
<td>10</td>
<td>Identifier</td>
<td>101</td>
</tr>
</tbody>
</table>

Conformance level = True
### D.2.2 Domain consistency

The data quality measures for the data quality subelement domain consistency are provided in Tables D.14 to D.18.

#### Table D.14 — Value domain nonconformance

<table>
<thead>
<tr>
<th>Line</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Name</td>
<td>value domain nonconformance</td>
</tr>
<tr>
<td>2</td>
<td>Alias</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>Data quality element</td>
<td>logical consistency</td>
</tr>
<tr>
<td>4</td>
<td>Data quality type</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>Data quality value type</td>
<td>Boolean (true indicates that an item is not in conformance with its value domain)</td>
</tr>
<tr>
<td>6</td>
<td>Data quality value structure</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>Source reference</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>Example</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>Identifier</td>
<td>14</td>
</tr>
</tbody>
</table>
Data management – Step 2

Product Specification → Quality Control → Metadata
Quality control routines

Automatic tests with FME

• Schema control
• Completeness
  • Omission
  • Comission
• Aggregation
  • Data product specification passed
  (Check if all conformance levels passed)

Quality results reported as XML in metadata
(ISO 19115, 19139)
Data management – Step 3

Product Specification → Quality Control → Metadata
Quality result reported in metadata

XML format ISO TS 19139
Update metadata in portal using FME and CSW
Summary
Summary

• Different users have different requirements
• Report data quality in metadata
• Automated quality control processes integrated with the data workflow
Questions

Visit our booth in the exhibition area

Anna Halvarsson
anna.halvarsson@metria.se

Johan Esko
Johan.esko@metria.se

http://www.metria.se