Implementing Quality in SDIs –
Introducing Building Blocks of Quality Web Services

Antti Jakobsson

J. Mäkelä, R. Henriksson, J. Marttinen, L. Tsoulos, I. Kavadas, E. Onstein, M. Beare, F. Williams, J. Walsh, I. Persson, L. De Meulenaer, J. Brennecke

4th Inspire conference 2010, Krakow 25th June 2010
ESDIN project info (www.esdin.eu)

• Project partially funded by eContentplus programme
• Started in September 2008 and will run for 30 months until March 2011
• Coordinated by EuroGeographics with 20 project partners
The challenge

• "The next challenge in INSPIRE is to populate it with quality data" – Hugo de Groof at INSPIRE plenary Krakow
• "EEA will concentrate on quality assurance of environmental data” – Jacqueline McGlade, Executive Director, Eea at INSPIRE plenary Krakow
• “Data quality is a very difficult topic” – Katalin Toth, JRC at INSPIRE WS Krakow

• ESDIN message “Most of the quality assurance processes needed in SDIs can be automated bringing significant savings to the producers and improving quality for users” – and we have done it
• Well, nearly...
Data flow in SDIs

**Producer**
- Data Specification
- Conformance levels
- Production/quality control

**MS**
- Inspire specifications
- NSDI conformance levels for reference information

**INSPIRE**
- INSPIRE quality evaluation metadata requirements for reference data
- INSPIRE conformance Levels (e.g. logical consistency)

**User communities**
- User Specification
- User community conformance levels

**Data**
- Quality evaluation
- Conformance testing

**Metadata**
- Transformation
- Quality Evaluation
- Conformance testing

**Data /Metadata**
- Transformation
- Quality Evaluation
- Conformance testing

**Data /Metadata**
Where quality terms apply in relation to quality assurance activities

Quality assurance is anything to do with quality anywhere in the production process

Accreditation / Certification could apply to the process, the person or the production system

A quality evaluation can be carried out at any point in the flowline

Editor → Production process → Quality control → Data Store → Product
ESDIN approach
ESDIN approach to quality

Quality Model

- Defines Measures

Data Quality Web Service

- Evaluates Data
  - Automatic Service
    - Radius Studio
  - Semi-Manual Service

Metadata

- Documents Metrics
- Describes Quality

Data User Web Service

- Publishes Quality (as provided)
- Assesses Quality (as needed)

Usability Model

- Defines Needs

ESDIN In Scope
Quality Evaluation process according to ESDIN Quality Model (D8.1)

Conformance levels described by error rate (AQLs), error count and CE95

Select from tables quality measures for feature types and attributes to be evaluated

Test completeness, thematic accuracy and positional accuracy by sampling

Check conformance with AQLs if not acceptable then correct

Test logical consistence and temporal accuracy with full inspection

Report error count=0

Report error rate as DQL/ and CE95 based on requirements

Completeness
Thematic accuracy
Positional accuracy

Metadata
Scope dataset

Scope: dataset

Scope: subset (sample)

Test report

Report error rate/ and CE95

feedback
Quality Model content (D8.1)

A. Identification of the objectives and requirements of quality at feature type level
   - Analysis and identification of quality requirements
   - Selection of quality parameters requires the (prior) identification of:
     • the data quality elements and subelements (ISO 19113)
     • the quality measures (ISO 19138)
     • the acceptance criteria and conformity levels of quality. These conformity levels may be set as declared quality levels (DQLs) which then are reported in metadata.

B. Evaluation of the spatial data
   - The evaluation procedures according to ISO 19114.
   - The production and metadata recording process (ISO 19115 or quality report).
<table>
<thead>
<tr>
<th>FEATURE TYPE &amp; Attributes</th>
<th>COMPLETENESS</th>
<th>LOGICAL CONSISTENCY</th>
<th>POSITIONAL ACCURACY</th>
<th>TEMPORAL ACCURACY</th>
<th>THEMATIC ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administraive Boundary</td>
<td>DQ BM error rate id 2</td>
<td>DQ BM error count id 12</td>
<td>DQ BM error count id 19</td>
<td>DQ BM error count id 12</td>
<td>DQ BM LE 95, 139</td>
</tr>
<tr>
<td>geometry (GM_Curve)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>id</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>country</td>
<td>DQ BM error rate id 7</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error rate id 51</td>
<td>DQ BM error rate id 51</td>
</tr>
<tr>
<td>boundaryLevel</td>
<td>DQ BM error rate id 7</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error rate id 51</td>
<td>DQ BM error rate id 51</td>
</tr>
<tr>
<td>meaningOfLine</td>
<td>DQ BM error rate id 7</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error rate id 51</td>
<td>DQ BM error rate id 51</td>
</tr>
<tr>
<td>boundaryStatus</td>
<td>DQ BM error rate id 7</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error rate id 51</td>
<td>DQ BM error rate id 51</td>
</tr>
<tr>
<td>validFrom</td>
<td>DQ BM error rate id 7</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error rate id 51</td>
<td>DQ BM error rate id 51</td>
</tr>
<tr>
<td>validTo</td>
<td>DQ BM error rate id 7</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error count id 16</td>
<td>DQ BM error rate id 51</td>
<td>DQ BM error rate id 51</td>
</tr>
<tr>
<td>beginLifespanVersion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>endLifespanVersion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdministrativeUnit</td>
<td>DQ BM error rate id 3</td>
<td>DQ BM error count id 10</td>
<td>DQ BM error count id 19</td>
<td>DQ BM error count id 12</td>
<td>DQ BM LE 95, 139</td>
</tr>
<tr>
<td>geometry (GM_MultiSurface)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How to utilize the quality model

• Quality model will be transformed to a rule set and conformance levels
• ExM will include these for the NMCAs
• Automated tools utilizing the rule and conformance levels
ESDIN approach to quality

- Quality Model Guidelines
  - Defines Measures
  - Documents Metrics

- Metadata Guidelines
  - Describes Quality
  - Defines Needs

- Usability Model
  - ESDIN In Scope

- Data Quality Web Service
  - Evaluates Data
    - Automatic Service
    - Semi-Manual Service
      - Radius Studio

- Data User Web Service
  - Publish Quality (as provided)
  - Assess Quality (as needed)
    - Catalogue Service
    - Fit For My Purpose? Service
Where you utilize quality webservices?

• If you are a data provider for SDI
  – For quality control during production (automated) called here conformance testing (this includes edge-matching and generalization)
  – For quality evaluation after the production (semi-automated)
• If you are the SDI co-ordinator or data custodian
  – For quality audit for process accreditation or data certification doing either conformance testing and/or quality evaluation
• If you are customer or data user
  – To evaluate usability using metadata information
EBM, ERM and EGM are just specific download services from respective ExMs
EG services are ‘cascading’ to NMCA services
EG GeoRM is a ‘broker’ for NMCA GeoRM

NMCA derives both ExM Large and ExM Medium from its Master data
and generalize ExM Medium from ExM Large

This NMCA derives ExM Large from its Master data, and generalize ExM Medium from ExM Large

Edge-matching does not necessarily need to be performed twice, generalization of edge-matching should still be edgematched.
Demonstration

• Goal1: to meet INSPIRE/ESDIN quality requirements and to report conformance (producer)

• Goal2: To evaluate usability of reference data for GMES reference data services
Goal 1: Quality Evaluation

- National Land Survey of Finland is the official source of INSPIRE Annex I reference data in Finland. It will utilize the web service to evaluate if they have met the quality model requirements of INSPIRE/ESDIN before publishing data in ESDIN services.
- The same procedure may be used for GMES reference data provision when NLS is the accredited provider for GMES.

- This is partly manual process requiring sampling/testing with reality; a web service might guide a process.
- ESDIN has demonstrated that this can be done.
Goal1: Conformance testing

- Cadastral Agency in Belgium and Dutch Kadaster are providing Esdin large scale data for cadastral boundaries. Goal is to check conformance during and after the generalization and edge-matching processes.
- The same procedure may be utilized when Cadastral Agency Belgium and Dutch Kadaster are accredited provider for GMES services.
Activity 1
Identify INSPIRE Target Schema
Activity 2
Identify NMCA Source Data/Schema
Activity 3 (continued)
Author Transform Action (build CP refs)
Activity 5
Visualise ExM Data in GML Viewer
Activity 6
Author ExM Validation Rules (Conformance criteria)
Activity 7
Evaluate Data In Respect Of Conformance criteria
## Rule Conformance Results

**Class:** cp:CadastralBoundary  
**geometry**

<table>
<thead>
<tr>
<th>Rule</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (183973.156, 319844.421) max (184347.792, 319733.717)</td>
</tr>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (182354.492, 320132.757) max (182563.127, 320240.49)</td>
</tr>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (183206.072, 319348.339) max (183388.07, 319259.574)</td>
</tr>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (183956.694, 319536.516) max (184274.742, 319728.121)</td>
</tr>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (183956.403, 319530.828) max (184297.486, 319723.799)</td>
</tr>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (183796.828, 319816.527) max (183905.058, 319915.218)</td>
</tr>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (183284.401, 320223.578) max (183368.241, 320435.532)</td>
</tr>
<tr>
<td>INSPIRE Annex version 3 / Cadastral Parcels / Logical Consistency / Domain Consistency / Check Boundary has exactly one parcel on each side</td>
<td>MBR min (183065.905, 320587.92) max (183337.676, 320693.747)</td>
</tr>
</tbody>
</table>
Goal2: Evaluation usability for GMES reference data services

- Two possibilities: taking data that is not from accredited sources:
  - Usability criteria taken from GMES specifications
  - Deriving data quality criteria from usability criteria
  - Checking that minimum criteria is met from metadata
  - Running the quality web service (automated)
- If using accredited sources then just checking the conformance is met
Conclusions

- It is important that INSPIRE will give a platform for data quality information; minimum data quality conformance levels set and then ability to report other user community related conformance levels.
- Quality evaluation metadata should be available for automated conformance testing.
- Introducing a quality model which uses a same principles for all Annex I themes -> we will suggest this a guideline for INSPIRE implementation.
- Introducing conformance levels that can be evaluated using semi-automated or automated based on ISO standards.
- Automation of quality evaluation and conformance testing can be done for all transformation related workflows including schema transformation, generalization and edge matching.
- Significant saving potential in quality reporting and improvement of data.