Integrating spatial data integration:
An Architecture for Complex Transformation Services

INSPIRE Conference 2010, Krakow, Poland
24th of June 2010

Thorsten Reitz, Daniel Fitzner, Marian de Vries, Ana Belen Antón Gironés
Spatial Data Integration

Combination of different geospatial data sets, across
- spatial,
- temporal and
- thematic
boundaries, opens up opportunities for policy and business improvement

Hurdles to combination include:
- Required data integration is often performed only to a minimum degree
- ROI on data integration is often bad since bespoke, non-reusable or mostly manual solutions are employed
- Adoption to ever-growing amounts of data and to real-time integration is required
Spatial Data Harmonisation

High-Quality Integration of Services and Data requires data harmonisation

- "the possibility to combine data from heterogeneous sources into integrated, consistent and unambiguous information products, in a way that is of no concern to the end-user"

Data harmonisation has many aspects:

- Organisational and procedural harmonisation;
  - Agreeing on common terminology, conceptual schemas, maintenance processes, data quality requirements

- Dataset harmonisation:
  - Processing of the data to conform to the rules as formalized in organisational and procedural harmonisation
## Aspects of Geodata harmonisation and dependencies

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Description</th>
<th>Related Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Application Schema</strong></td>
<td>Transformation of data from one application schema to another, or enrichment of data within the schema</td>
<td>Data (and schema) Format</td>
</tr>
<tr>
<td><strong>Spatial Aspects</strong></td>
<td>Conversion of different geometry types</td>
<td>Multiple representations</td>
</tr>
<tr>
<td><strong>Coordinate Referencing</strong></td>
<td>Reprojection</td>
<td>Multiple representations</td>
</tr>
<tr>
<td><strong>Units of Measurement</strong></td>
<td>Unit Conversion</td>
<td>Application schema transformation</td>
</tr>
<tr>
<td><strong>Object References</strong></td>
<td>Normalisation or denormalisation of data sets and their interdependencies; changing format of references</td>
<td>Application schema transformation</td>
</tr>
<tr>
<td><strong>Identifier</strong></td>
<td>Changing formats of identifiers, using globally unique and traceable identifiers</td>
<td>Object References</td>
</tr>
<tr>
<td><strong>Metadata</strong></td>
<td>Conversion from one possible metadata schema to another; addition of Metadata from known context or from analysis of the data itself</td>
<td>Application schema transformation</td>
</tr>
<tr>
<td><strong>Data Quality</strong></td>
<td>Repairing invalid geometries, rebuilding topologies, edge matching and others</td>
<td>Spatial Type Conversion</td>
</tr>
<tr>
<td><strong>Data Format</strong></td>
<td>Conversion of exchange/storage format to another</td>
<td>Has to be handled before all others</td>
</tr>
<tr>
<td><strong>Multiple Representations</strong></td>
<td>Alignment of the (geometric) attributes of multiple representations, merging of multiple representations</td>
<td>Application schema transformation</td>
</tr>
<tr>
<td>D</td>
<td>Application schemas</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Spatial and temporal aspects</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Multilingual text</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Coordinate referencing – units of measurements</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>Object referencing model</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>Portrayal</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>Identifier management</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Metadata</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Data &amp; Information Quality</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>Data transfer</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Multiple representations</td>
<td></td>
</tr>
</tbody>
</table>

**D1** ISO 19109 General Feature Model-based AS
**D2** ISO 19126/19135 INSPIRE Feature Concept Register
**D3** ISO 19110 Feature Catalogue

**E1** ISO 19109 8.7 spatial geometry or a topology
**E2** ISO 19109 8.9 geographic identifier in a gazetteer
**E3** ISO 19123 coverage function
**E4** temporal geometry or a temporal topology
**E5** properties with a value Date, DateTime and Time

**F1** Coordinate reference systems
**F2** Temporal reference systems
**F3** Units of measurement, such as SI units
**F4** Geographical grid systems (needs to be clarified)

**H1** SE documents and rendering

**K1** Metadata Schema and Values

**M1** Model and Values

**O1** Delivery Medium (Protocol, such as HTTP)
**O2** Delivery Format (such as GML via WFS)

**P1** Multiple Levels of Detail
**P2** Multiple Representations in different AS
A Model for Atomic Processing Services

- Handle each aspect using a generic, well-documented and repeatable approach
  - Provide processing lineage metadata according to a common standard

- Handle each aspect isolated from others
  - Annotate the service so that resolved data integration aspects are clearly identifiable
  - Provide information on affected aspects (i.e. changes which could lead to side effects in other transformation steps) of the transformed data

- Combine atomic services in optimal ordering
  - To avoid/minimize side-effects between different aspects processing services
Assembling Complex Transformation Services

- A **Workflow Repository Service** can use the annotations to the atomic services to create workflows in optimal order.
- Workflows are created manually by a user or automatically on the basis of a user-given product specification with the following elements:
  - Thematic constraint (application schema, spatial object types)
  - Spatial and temporal constraints
  - Format and protocol constraints
  - Quality and Metadata constraints
Executing Complex Transformation Services

- Modular construction of complex transformation services
- Dynamic selection of façade, processing and data access modules based on request
Integrating Complex Transformation Services

- WFS Italy
- WFS Portugal
- WPS Fraunhofer
- PA Application Logica
  - OpenLayers Standard Client
  - IGS + Workflow Service
  - Mediator Service
  - Context Service
- WPS Fraunhofer
  - Scale Harmonization WPS
  - Coordinate Transformation WPS
  - Multiple Representations WPS
- WFS Italy
- WFS Portugal

Desktop Toolkit
- Toolkit
  - HALE
  - GeoModel Editor
- Model Repository
Evaluation & Conclusion

To perform data harmonisation based on the combination of atomic processing services, harmonisation aspects needs to be modeled and understood well.

Evaluation of the model and architecture in HUMBOLDT scenarios
- Clear identification of relevant harmonisation issues

Main conceptual issues
- Fully automated service assembly for harmonisation

Main issues in implementation
- Varying interpretations of existing standards, especially WFS and GML, make automated service composition very hard
- Current standard geospatial interfaces such as WPS not optimized for asynchronous processing and large data amounts
Thanks for your interest! Any questions?

To get more information of the presented work, please visit

http://www.esdi-humboldt.eu

http://community.esdi-humboldt.eu

Thorsten Reitz
Fraunhofer-Institut für Graphische Datenverarbeitung IGD
Fraunhoferstraße 5
64283 Darmstadt

Tel  +49 6151 155 – 416 | Fax – 444
thorsten.reitz@igd.fraunhofer.de
www.igd.fraunhofer.de