Wolfgang Kresse
kresse@hs-nb.de
Delegate of EuroSDR to ISO/TC 211

18.8.2015

AGILE / OGC / JRC / ELF / EuroSDR
Workshop: Data Modelling and Model driven Implementation of Data Distribution
Copenhagen, Denmark, January 28 – 30, 2015

Analysis regarding update-requirements of the ISO standards

The workshop in Copenhagen gathered experts for modelling of geographic information from many parts of Europe and focussed on the experiences made while implementing the INSPIRE project and its national derivatives using the ISO-standards for geographic information.

The participants reported about numerous successful implementations. However, the workshop also aimed at the detection of shortcomings of the existing ISO-standards found during the INSPIRE-implementation in Europe.

This analysis based on about 30 presentations given at the workshop summarizes the demands for improving the standards. The statements are copied from the presentations and grouped in
1. Demands for a wider perspective of modelling – fundamentals
2. Demands for a better standardization – updating the ISO standards
3. Demands for better rules for working with models – implementation issues
4. Outreach

Part 1 puts the fundamental question whether the existing approach of the ISO standards of the ISO 19100 series developed by the ISO/TC 211 “Geographic information / Geomatics” fully meets the modern requirements. The core element of the ISO 19100 series is the “General Feature Model” according to the ISO 19109. However, this model does not so well model dynamic structures such as linked data where the user defines the data sources at the moment of calling the map components that meet his requirements. One could say in other words that the ISO-model is good for a producer-driven map-production while modern technology also allows for user-driven definition of maps.

Part 2 addresses a number of update requests for existing ISO standards. Many of the comments deal with the problem of ISO-standards being abstract by their nature on one side and a lot more detailed definitions required by the implementation process on the other side. It has also turned out that the model-based data exchange is a wider task than perceived during the standard’s development and addressed by the ISO 19118 “Encoding”. Another new challenge is the a lot more dynamic nature of geospatial data and structures today compared to the present tenor of the ISO 19100 series of standards.

Part 3 is a collection of demands regarding the implementation procedures. The topics range from a simplification of models and an extended registry-creation to recommendations for state-of-the-art software products.

Part 4 calls for a better education of gi-experts in modelling because this knowledge is essential for a successful integration of the relevant systems such as INSPIRE with local and / or thematic datasets.

Annex A

AGILE / OGC / JRC / ELF / EuroSDR
Workshop: Data Modelling and Model driven Implementation of Data Distribution

Classification of the relevant statements

1. Demands for a wider perspective of modelling – fundamentals

1.1 Classification (General Feature Model)

Explicit and documented
Relational or tree structured (XML)
Can match the way communities think
Can be adjusted to new knowledge
Can perform and scale … not (necessarily) closed, in either sense

1.2 Linked data (Graph data models)

Linked data helps to integrate and re-use
URI-strategy is being developed
Can be used to explore
Can be useful for combining disparate data
Hierarchical (tree) models
Fit some kinds of things

1.3 Cartographic (vector)

1.4 Raster, grid, coverage

2. Demands for a better standardization – updating the ISO standards

2.1 Abstract nature for ISO-standards

UML leaves a lot of freedom → different modelling styles and approaches
Conceptual models may not always be directly suitable for generating data encodings

2.2 Mapping between standards

Model mappings may be required between
- Conceptual and implementation models
- Exchange data models and existing standards (e.g. GeoSciML, CityGML)
- Exchange data models and data models of source data sets

Work is still needed on Model mappings

2.3 Unification of standards

Multiple standards
- INSPIRE (static road data)
- DATEX II (real-time road status and traffic information

Common concepts (network, location references)
Common data modelling principles, patterns and practices
Common data types

2.4 Standard for handling URLs

How can users use the URL's they get?

2.5 Dynamic code lists

Who has the responsibility for creating and maintaining an infrastructure for code lists?
History of code lists

2.6 Dynamic IDs

ID management and support for linking: How to best make Object IDs dereferenceable in a dynamic environment?

2.7 Versioning

Versioning of models, conceptual schemas and implementation schemas: Many different approaches

3. Demands for better rules for working with models – implementation issues

3.1 General

In some cases overly complex models (inheritance, abstraction, tagged values, levels of nesting)
What about our approach of using « association classes » for modeling core relations (extension of the association class via generalization)?
Should a data model for « management » of basic data have strong multiplicities? Should they be more loose and be managed via constraints/rules?
What about our approach of this data model for management of basic data versus products (e.g. INSPIRE, ISA)?
Missing: The modelling of concrete business use cases (public administration)
Not all things can be found in the Data Specifications and the GCM

3.2 Dependencies

Many dependencies between data models for 34 INSPIRE themes
Dependencies also reflected as imports in XML schemas
When updating individual application schemas and XML schemas, this creates a cascade of required updates in dependent schemas
Work is still needed on Management of dependencies and version

3.3 Semantics

Missing: The modelling of cross-domain concepts (terms and semantics)
No special language-dependent characters, e.g. Danish characters

3.4 GIS-recommendation

Need for automatic ETL (extract, transform, load) processes
Need for recommended GIS:

- ArcMap
- AutoCAD Map
- Gaia

- Geomedia
- MapInfo
- QuantumGIS
- U-Dig

3.5 Implementation

Processes often not being automated, which means that time-consuming hand coding is widespread and a source of error
A random selection of approach and methodology, which results in different solutions and distribution types

3.6 Code lists

Where to manage code list values during data model development?

3.7 Registries

If managed outside the UML model (e.g. in a register), how to connect the two?
We need a semantic Register: Make registers or catalogues: objects, codelists
Missing: The modelling and implementation of (a set of) registry covering: concepts, code lists and an authority register

3.8 Formats

Not all GIS software can read INSPIRE GML by default

3.9 Data management

Data management not being a developed discipline

4. Outreach

Most GI experts are no UML modelling experts
Many GI experts have difficulties with understanding UML class diagrams
Data providers have difficulties identifying relevant application schemas and feature types to map their INSPIRE-relevant data to
Work is still needed on General training & guidelines
We need to share examples and experiences
Different authorities having different degrees of knowledge
Models not easy to read and understand for people outside modelling community

Annex B:

AGILE / OGC / JRC / ELF / EuroSDR
Workshop: Data Modelling and Model driven Implementation of Data Distribution

Analysis of the presentations regarding relevance for the ISO/TC 211 – Relevant statements found in the presentations

111: Danish Geodata Agency: Thorben Hansen

112: Model-driven INSPIRE: Michel Lutz & Clemens Portele

Issues

- Most GI experts are no UML modelling experts
- UML leaves a lot of freedom → different modelling styles and approaches
- In some cases overly complex models

- Where to manage code list values during data model development?
- If managed outside the UML model (e.g. in a register), how to connect the two?
- Many GI experts have difficulties with understanding UML class diagrams
- Data providers have difficulties identifying relevant application schemas and feature types to map their INSPIRE-relevant data to
- Conceptual models may not always be directly suitable for generating data encodings, e.g. when
  - advanced modelling constructs (e.g. association classes) are used
  - Complex data types are used, e.g. GeographicalName
- Many dependencies between data models for 34 INSPIRE themes
- Dependencies also reflected as imports in XML schemas
- When updating individual application schemas and XML schemas, this creates a cascade of required updates in dependent schemas
- Model mappings may be required between
  - Conceptual and implementation models
  - Exchange data models and existing standards (e.g. GeoSciML, CityGML)
  - Exchange data models and data models of source data sets
- Mappings are often expressed using mapping tables
  - Not easy to use
  - Not easy to maintain if one of the model changes

- Work is still needed on
  - General training & guidelines
  - Model mappings
  - Management of dependencies and versioning

Guidelines and tools for generating (simple) encodings

113: Geonovum: Paul Janssen

We need a semantic Register

- Make registers or catalogues: objects, codelists
- Linked data helps to integrate and re-use
- URI-strategy is being developed
- Extending models using ADE
- What about our General Feature Model?
- What about UML modelling?

114: OS Great Britain: Peter Parslow

Pros & cons – my thoughts: GFM-based models
- with explicit, documented, classification!
- Relational or tree structured (XML)
- Most people classify things
  - Can match the way communities think
- Can be adjusted to new knowledge
- Can perform and scale
- …not (necessarily) closed, in either sense

✔ Performance
✘ Effort to extend

Pros & cons – my thoughts: Graph data models
- ‘linked data’
- Can be used to explore
- Can be useful for combining disparate data
- Hierarchical (tree) models
- Fit some kinds of things

✔ Easy to extend
✘ Performance

Other conceptual models exist:
- cartographic
- raster / grid / coverage

What do you think?
- Can GFM-based implementations be open and extensible? and still perform?
- Can we ever know enough about a domain to completely model it?
- Are generic data models useful?
- Do extensible code lists break interoperability, and even software?
- Do you have an (organisational) approach to estimating how much you don’t know?

121: ELF: Morten Borrebaek, Anja Hopfstock, Heidi Vanparys

122: Statens vegwesen, Norway: Knut Jetlund

123: Agile Design and Testing, Thorsten Reitz & Simon Templer

- Model-driven
  - No Schema Language Mismatches between steps
- Data-driven
  - Inform every step with real-world data
- Usage-driven
  - Instantly validate data fitness-for-purpose
- Collaboration
  - Versioning, Forking
  - Comments, Tasks, Notes
- **Accessible**
- Touch-friendly View
  - Based on UML Model
- Discover the relevant model elements
  - Visual Hints
- One Fixed Layout
  - Directions have meaning

- Supports subset of UML/XSD concepts
  - Classes, Choices, Enumerations
  - Interhitance, Aggregation, Reference

HALE

124: Linked data infrastructure, FGI: Esa Tiainen

211: ArcGIS for INSPIRE, ESRI: Paul Hardy & Roberto Lucchi

212: Environment in ArcGIS, SYKE: Lena Hallin-Pihlatie et al.

●A confusing mismatch between the ESRI LC attributes and the DS/GCM/ISO attributes
●A confusing mismatch between the ESRI data structure and the DS data structure
●All things cannot be found in the Data Specifications and the GCM
  ○ISO standards
●The DS is not in always consistent. CorineClassValue = 111 or 1.1.1?

●Is INSPIRE compatibility possible without sharing vector data as GML-files in your Atom feeds or your WFS services?
●“IRs do not oblige the usage of a specific encoding”
●“This data specification proposes the use of GML as the default encoding”
●If we use restructured/flattened versions of the INSPIRE data models how can INSPIRE conformance be ensured?
●Not all GIS software can read INSPIRE GML by default
●If we want to produce INSPIRE compliant WFS and/or Atom feeds and also fulfil the technical requirements of our users, it seems that we need to choose a software that does not only support the sharing of GML but also other formats

●To ensure resource-efficiency and timely consistency
  ○Need for automatic ETL processes, probably in FME
  ○Transformation services?
  ○We need to share examples and experiences

213: ShapeChange: Johannes Echterhoff & Clemens Portele

214: ShapeChange, EA: Kartverket

215: ELF-Codelists: Heidi Vanparys

216: Generalisation, GenIE, OS: Debbie Wilson

221: MDA in Germany: Clemens Portele & Markus Seifert

Issues for discussion

• ID management and support for linking
– How to best make Object IDs dereferenceable in a dynamic environment?
• Versioning of models, conceptual schemas and implementation schemas
– Many different approaches
• Use of INSPIRE tools (PID, registries, EA subversion, ISO schemas)

222: Flandern: Ziggy Vanlishout & Jef Vanbockryck

Challenges / questions to forum

- *What about our approach of using « association classes » for modeling core relations (extension of the association class via generalization)?*
- *Should a data model for « management » of basic data have strong multiplicities? Should they be more loose and be managed via constraints/rules?*
- *What about our approach of this data model for management of basic data versus products (e.g. INSPIRE, ISA)?*

223: swisstopo: Christine Najar, Pasquale Di Donato, Patrick Gamma

224a: Danish Geodata Agency, e-government: Fleming Nissen

Further Work
- Some high level modelling issues are missing:
- The modelling of cross-domain concepts (terms and semantics)
- The modelling of concrete business use cases (public administration)
- The modelling and implementation of (a set of) registry covering: concepts, code lists and an authority register

224b: Danish Geodata Agency, Basic Data infrastructure: Heidi Vanparys et al.

Context
Principles of the Data Distribution department for the Digital Map Supply:
- Flat structure
- No special Danish characters
- GML 3.1.1 SFP

Vælg GIS program:

- ○ ArcMap
- ○ AutoCad Map
- ○ Gaia
- ○ Geomedia
- ○ MapInfo
- ◉ QGIS
- ○ U-Dig

Code lists
- No infrastructure
- Who has the responsibility for creating and maintaining the infrastructure?
- History of code lists
- How can users use the URL's they get?

Enumerations
«Values of enumerated types shall follow the naming conventions for regular attribute names and should be mnemonic, unless an established name for the concept exists.» (ISO/DIS 19103)

(from GST-udstillingsmodeller:: Udstillingsmodeller:: Stednavne)

(from GST-udstillingsmodeller:: Udstillingsmodeller:: GeoDanmark::Trafik)

224c: distribution of gi

Current issues
At present, the handling and distribution of geographic information is characterised by:
· Each authority using its own methods, techniques, tools and infrastructure
· Processes often not being automated, which means that time-consuming hand coding is widespread and a source of error
· Different models and terms making cross-authority matching difficult
· A random selection of approach and methodology, which results in different solutions and distribution types
· Data management not being a developed discipline
· Different authorities having different degrees of knowledge

224d: Basic Data Program

224e: distribution of basic data

311: Linear Referencing: Paul Scarponcini

312: from table to GML: Knut Jetlund

313: Data models and architecture: Joergen Flensholt & Henrik Friis

Concerns

Multiple standards
- INSPIRE (static road data)
- DATEX II (real-time road status and traffic information
- Common concepts (network, location references)

Integration of standards development
Common data modelling principles, patterns and practices
Common data types

Model complexity
- Modelling concepts (inheritance, abstraction)
  Tagged values
- Many levels of nesting

Simple model design

Tool support (browsing, code generation)

Readability

- Models not easy to read and understand for people outside modelling community

User guides and manuals
Annotations
Tool support (browsing)

314: CityGML: Tadjana Kutzner & Thomas Kolbe

As long as the application schemas of geo base data do not allow to be extended by application / user data they will only live (and be used!) at the interface between the data provider and the first user / refiner. The data models (and importantly: formats) cannot be used in later steps, where data with added value is being produced and used.

316a: Geospatial Model: Jan Kolář

Do we have all the technology we need?
1. No evidence we have - no reference design available
2. Heterogeneous spectrum of data models
3. Several competing alternatives
4. Complex data models
5. Content unit is Geospatial Model
6. Unifying computational context is missing

Would technology-related law help finding a GS?

Is ability to accommodate new types of models a primary requirement?
1. If we want flexible solution >> YES
2. Solution that “survives” for longer time >> YES
GS must address COMPLEXITY and CHANGE

Conclusions
• Unifying geo computational context is missing
1. Request typed link via URI address
Specs. on the level of data model (ontology)
2. Implicit reference via points in geographic space
Specs. on the level of computation and geographic space
• Think for long-term solution
• Seek leadership
• Not ways to enforce geospatial solutions.

316b: DM Workshop, Jan Kolář

Issue 1
At the core of model-driven approach are mappings between:
Conceptual and exchange (product) data models.
Exchange (product) data models and data models of source data sets (RDBMS).
Exchange (product) data models and existing standards (e.g. GeoSciML, CityGML).

Issue 2
Transformation is hard!

Issue 3
Encodings. What is actually the intention of the INSPIRE directive – download data, but needs tools (like FME) to transform this to a format that GIS tools can use.

Issue 4
Software innovations - More and smaller models, easier to implement, or larger model covering more use cases? Needs for both.

Issue 5
Missing methods and tools for identifying and managing dependencies of data models and encodings.

Issue 6
Model aware update process - how to keep the data model updated - process?
Versioning, tracking of versions, influences on transformations.
Differing version concepts: release vs. revision, logical vs. physical versions.

Issue 7
Maintenance of data models.
Consequences on the processes, schema.
Documentation of model changes.
Update scripts / evolution scripts, SQL.

Issue 8
How to formalise, interpret and validate integrity constraints?
Use Object Constraint Language (OCL) extensions, which are available in UML?
Create templates for common constraints?

Issue 9
We use imperfect models, e.g. in terms of completeness, performance, etc.; how to make them more consistent?
Does scope of the models influence their consistency? How?
What about replacements in data models?

Issue 10
How to manage extensions. (INSPIRE extension).
Consistency among the model components.
Do extensible code lists break interoperability, and even software?
Do application domain extensions do the same?
National extensions.
Can general feature model-based implementations be open and extensible? and still perform?
Are generic data models useful?

Issue 11
We still do not know how to manage the provision of different services powered by the processing pipelines like GenIE.

Issue 12
Is INSPIRE compatibility possible without sharing vector data as GML files?
INSPIRE Requirements do not oblige the usage of a specific encoding, but proposes the use of GML as the default encoding.
If we use restructured/flattened versions of the INSPIRE data models how can INSPIRE conformance be ensured?
INSPIRE Themes use different formats, relevant to the thematic needs. Some themes have several use cases – the different cases would benefit from different encodings.

Issue 13
Some model components are complex (large aggregates, but applications and users need just part of it.)

Issue 14
How to make dynamic, online data service on algae situation INSPIRE compliant?
DATEX II (real-time road status and traffic information) is not working with INSPIRE.

Issue 15
Work with code lists and handling of updates.
Bigger code-lists.
Interaction with registries.

Issue 16
Complexity is sometimes necessary, but complex structures are expensive to ingest.
GML can handle complexity, but the lack of development on client side, to ingest complex data is a disappointment.
Vendors need assurance that the specification should not change. Challenge for INSPIRE.
Not all GIS software can read INSPIRE GML by default.
Thanks to INSPIRE more data is shared, but the provided GML is not used.

Issue 17
Agencies should go for flexibility in the use of data.

Issue 18
What is the right approach to keep pace with technological developments?

Issue 19
eGoverment is a political trend to improve productivity, contribute to a more efficient and coherent public sector with a high quality in service.