Validation of INSPIRE metadata, data and services

INSPIRE conference – 30 September 2016
Context

- Workshop about validation of INSPIRE technical components
  - Paris – 02-03 June 2016
  - Co-organised by
    - EuroGeographics (INSPIRE KEN)
    - EuroSDR
    - European Commission

Presentations, videos, minutes available on:
http://www.eurogeographics.org/content/validation-workshop-organised-eurogeographics-jrc-and-eurosdr
Why?
Why validation?

• To check if the technical requirements are met => to ensure that the infrastructure will work
  − binary results
    • Passed or not
    • Not conform => not usable
  − for implementers

Validation does not guarantee easy use; still remaining issues (lack of client applications, only the structure is checked not the content, e.g. voidable attributes not filled)
Why validation?

• To measure the conformance to INSPIRE

=> to assess the progress

  − Indicator
    • Percentage of success
    • Rate
    − For deciders: encourage to do better

• Example: Commission metadata validator: from boolean (hardly no one conform) to indicator (significant progress)
Against what?
Against what?

Implementing rules

Technical guidelines

Choice of MI-WP5
- logical: ensure technical interoperability
- feasible: to control only one implementation (and not all possible ones)

National profiles, INSPIRE extensions
Who?
Who?

• Data producers
  – To get confident in what they provide
  – To fill the “conformity” element of metadata
  – To have their metadata, data and services accepted by …

• Brokers, integrators
  – European commission
    • Metadata validator (INSPIRE geoportal)
  – ELF project
    • Validation of data and services (ELF cascading services)
Who?

• Standardisation body
  – European Commission
    • Work under progress (MIWP-5)
    • Objective: to develop validation tools in agreement with MS
  – OGC
    – Services

• Users
  – No experience reported during workshop
  – Not yet the case?
When?
When?

• As soon as possible, integrate validation in the production process:
  – Metadata editor
  – Data transformation
    • Snowflake
    • HALE (integrate the Web validation of eENV+ project)

• At the end of the process
  – Validation tool to be independent from the production ones
How?
How?

• General method
  – Automatically (tools)
  – Manually (inspections, ...)

There are controls that can’t be done automatically; it does not mean they are not executable
How?

• Wide variety of tools
  – Metadata
    • European / national validators
  – Data
    – XML validators (XML Spy, FME, ...)
    – Other (Schematron for rules, CRS validator, ...)
  – Services
    – OGC CITE
    – ETF (from ELF project)
    – Spatineo (performances)

What is done so far

Duplications, inconsistencies
How?  

**INSPIRE Maintenance Implementation Group**

<table>
<thead>
<tr>
<th>MIWP-5: Validation and Conformity test</th>
<th>2016.3 Action: Validation and conformity testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/2013</td>
<td>2015 - 2016</td>
</tr>
</tbody>
</table>

- Develop a commonly agreed European validator for data, metadata and network services (incl. performance testing) - the validation rules should be made explicit so that data providers in Members States know what is validated upon exactly and how is validated;
- Establish rule that all new TG need to ATS and executable tests;
- Discuss the possibilities for setting up a compliance certification facility and process similar to the OGC;
How?

[+350 comments received during the MS consultation]

ATS

Metadata

Discovery

View

Download

Datasets

ETS

INSPIRE MIWP-5

ARE3NA ISA action

Existing validation tools and solutions

Software developer >

Existing validation tools and solutions
Which main difficulties?
Main difficulties

- Some tools are not mature enough
- Lack of knowledge
  - We are pioneers in validating big SDI as INSPIRE

<table>
<thead>
<tr>
<th>Service category</th>
<th>Status indicator</th>
<th>Experiences, remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELF Cascading View services WMS and WMTS</td>
<td></td>
<td>Tools ok, simple service complexity. Only 2 services!</td>
</tr>
<tr>
<td>ELF Cascaded Direct download services WFS2.0, GML3.2.1</td>
<td></td>
<td>Only 2 services (will be only 1 when Inspire v3 expires). Problems inherited from National services.</td>
</tr>
<tr>
<td>ELF National Direct download services WFS 2.0, GML3.2.1</td>
<td></td>
<td>Validation and test tools premature. Complex services, WFS service software issues, lack of skills</td>
</tr>
<tr>
<td>National view services WMS 1.3</td>
<td></td>
<td>Tools ok, easy/moderate service complexity</td>
</tr>
</tbody>
</table>

Main issues on download services (ELF project)
Main difficulties

• Analysis of errors
  – Takes time to analyse a new type of error
  – Reports not informative enough, error messages useful only for nerds
    • Difficult to find source of error (ETF)
  – However, once identified, errors may be easy to correct
Main difficulties

And also

- Different options when interpreting Technical Guidelines
- Ensure conformity over time (if update in Technical Guidelines)
  - Ex: European Commission metadata validator
    - Nice policy for validator
    - But editor is not working on same version
- Assess what is reasonable level of conformity
- Need for cross-component validation
- Issue with protected services (authentication, security)
What to be improved?
Research Conclusions

- **INSPIRE Test Framework** – Stage of Construction
  - Challenging + Impressive what has been achieved so far

- **Abstract Test Suite** - Comprehensive topic
  - Alignment with technological developments + User requirements matching

- **Executive Test Suite** – More straightforward topic - Still lots to do
  - Alignment with ATS, Re-use existing test suites, Development new suites

- **Metadata Validation** – Advancement stage / Stage of refinement
  - Automated metadata changes, multilingual issues, statistics tools

- **Services Validation** – Rather in an infancy stage
  - Immature tools (functionalities, error reporting), authentication, certification

- **Data validation** – Full attention / Already lots has been achieved
  - Schema validity, ‘incorrect’ files tests, usability, transformation workflow

- Most research issues are operational oriented and less conceptual
Potential improvements

• Data providers
  – Many errors due to broken links => resource can’t be found
  – Most of these links (between services and data) in metadata
  – Keep metadata updated

• Software providers
  – More user friendly tools
    • Ex: messages written by domain experts
Potential improvements

• **From formal compliance to data usability**
  - Check content (not only conformity to schema)
  - “voidable” issue => provide statistics about what is filled

• Better knowledge exchange in whole INSPIRE community
  - FAQ, frequent errors
  - Registry for testing, monitoring
Not valid?