Metadata workshop

New rules and processes for metadata

Marcin Grudzień

29th of May 2019
Agenda

• Introduction
• Survey results
• Rationale behind metadata changes
• Activities related to metadata
  • Fixing links between data sets and services
  • Monitoring and reporting
  • New INSPIRE Geoportal harvesting console
  • Other relevant activities
• Summary
Introduction (1)

• The goal of this presentation is to inform about recently implemented, current and planned changes related to metadata

• The source of information:
  • Geoportal workshop in January 2019, Ispra (Italy)
  • Maintenance and Interoperability Technical Group (MIG-T) meeting in April 2019, Ispra (Italy)
Introduction (2)

• Geoportal workshop
  • 23-24.1.2019
  • 45 experts from 28 European countries
  • The workshop goal was to:
    • Improve Member States (MS) data sets accessibility via INSPIRE Geoportal by:
      • Presenting & clarifying Geoportal back-end system
      • Sharing JRC assessment of the MS feedback related to the description of the MS infrastructures in use
      • Presenting a new system for harvesting MS resources
    • Improving MS data sets usability
    • Plan future steps (operational aspects) and the way of collaborating with MS technical experts (Helpdesk, Forum, Virtual meetings..)
  • https://webgate.ec.europa.eu/fpfis/wikis/display/InspireMIG/INSPIRE+Geoportal+Workshop
Introduction (3)

• MIG-T meeting
  • 3-4.4.2019
  • The meeting goal was to discuss and agree on technical aspects related to INSPIRE implementation by MS
  • This presentation covers INSPIRE geoportal developments that were discussed during the meeting and are strongly related to metadata
MS Survey on the INSPIRE Geoportal workshop

• Launched in December 2018
• 31 answers received
• 1 reply per country
• Complete survey results available to download as a [presentation](#) and [individual MS responses](#)
Survey: MS Metadata organisation – harvesting mode

Do you harvest automatically metadata of data sets and services from your data providers or from regional catalogues?

- Yes: 20
- No: 11
Survey: Metadata organisation –
For those who do harvest automatically

How often do you harvest metadata from your regional and data providers’ catalogues?

- Daily or mostly daily: 13
- Monthly or mostly monthly: 3
- According to data providers needs: 4

Do you use pull or push method for harvesting metadata?

- Pull: 17
- Push: 3

Have you implemented incremental updates to harvest metadata?

- Yes: 11
- No: 9

Provide a description of your implemented harvesting solution

- Many to 1 node: 12
- 1 to 1 node: 2
- No data: 6
Survey: Metadata organisation – For those who do NOT harvest

- Reasons for manually creating MD centrally:
  - “almost all the metadata are created by us (geoportal administrators) because structure of metadata is quite complex for data providers and they often make mistakes, etc. So it is easier for use to create metadata for them.”
  - “Harvesting process was done only once with the [***] agency for environment and nature but since their metadata profile has some errors and is not completely compliant with INSPIRE implementing rules we had to manually correct the metadata.”
Survey: Software used for the management of metadata

Some countries are planning:
- to move to Geonetwork
- upgrade their current Geonetwork’s version
Survey: Metadata software versions used & main issues

<table>
<thead>
<tr>
<th>Geonetwork Version</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.X</td>
<td>5</td>
</tr>
<tr>
<td>2.6.4</td>
<td>1</td>
</tr>
<tr>
<td>2.10.1</td>
<td>1</td>
</tr>
<tr>
<td>2.10.4</td>
<td>2</td>
</tr>
<tr>
<td>2.11</td>
<td>1</td>
</tr>
<tr>
<td>3.X</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3.0.2.0</td>
<td>2</td>
</tr>
<tr>
<td>3.2</td>
<td>3</td>
</tr>
<tr>
<td>3.2.2</td>
<td>1</td>
</tr>
<tr>
<td>3.4.1</td>
<td>1</td>
</tr>
<tr>
<td>3.4</td>
<td>1</td>
</tr>
<tr>
<td>3.4.2</td>
<td>2</td>
</tr>
<tr>
<td>3.4.3</td>
<td>2</td>
</tr>
<tr>
<td>Grand Total</td>
<td>19</td>
</tr>
</tbody>
</table>

**Geonetwork**
- **Index issues** - duplicated metadata
- Download errors with **stored query**
- Requires still **too much manual operation on raw XML** *(Tagging of priority datasets, resource linking)*
- Doesn’t drop off decimals of **coordinates** without rounding up
- **Opensearch** document is not complete (missing information)
- **Metadata editor doesn't produce 100% INSPIRE compliant metadata.**

<table>
<thead>
<tr>
<th>Geoportal Server version</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2.5</td>
<td>2</td>
</tr>
<tr>
<td>1.2.7</td>
<td>1</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>No version indicated</td>
<td>4</td>
</tr>
<tr>
<td>Grand Total</td>
<td>8</td>
</tr>
</tbody>
</table>

**Geoportal Server**
- Not possible to **import new thesaurus, code lists**
- **Metadata anchor element not supported** *(manually corrected metadata, is not valid for ESRI validator the file cannot be uploaded)*
- Have to remain on an older version of Geoportal software as **newer versions of CSW not allowed under INSPIRE technical guidelines**

<table>
<thead>
<tr>
<th>MDE Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mix of issues with MD editor and harvesting process</td>
</tr>
</tbody>
</table>
Rationale behind metadata changes

Usability improvement

• First of all for end users
• But also for data providers
INSPIRE Geoportal

- Launched on 18th of September 2018 during INSPIRE conference in Antwerp
- Provides easier, more intuitive access to a spatial data
- Top-down search approach
- Focusses on finding data sets rather than services
New INSPIRE Geoportal

- Top-down approach
  - Priority data sets
    - Country overview
    - Environmental domains
  - Environmental legislation
- INSPIRE themes
  - Country overview
  - INSPIRE Data themes
Discrepancies between number of datasets and number of network services

* Data acquired on 7th of May 2019
Causes of discrepancies between number of datasets and number of network services

It can be caused by many reasons:

• Many data sets published by one service
• Data sets are not published -> lack of services publishing data sets
• **Broken links between data sets and network services**

The broken links issue was identified at INSPIRE KEN validation workshop in 2016
Broken links

• Links between the data set and the service metadata harvested from the national discovery services as well as on the service metadata that can be obtained from the network services through their Get View/Download Service Metadata
What has been done to solve broken links issue?

- JRC provided:
  - The background document [Geoportal workflow for establishing links between data sets and network services](#)
  - Individual report per country with resource linking issues + examples of good metadata – services linking
  - January workshop was organised to further explain activities in this area
  - „Softening” data-service harvesting checker
  - JRC developed Resource Linking Testing Tool
    - Allows data providers to check whether the necessary linkages between MD – Services (View – Download) resources are provided correctly / consistently
    - Eliminates/reduces the use of the Geoportal harvesting as a testing environment.
  - Goes through internal security checks – should be available till 31st of May
Survey: Data set - service linking

**Additional information:**
- New national projects to improve (4)
- National workshops/webinars upcoming (3)
- Correct links only for harmonized INSPIRE datasets (2)
- There is still a value in providing only MD for datasets without services (2)
- Due to the WFS problems – ATOM feed only (2)
- Technical issues (missing extended capability for WFS, WFS 2.0 generation, WFS Stored queries..) related to the use of various SW packages

The majority of MSs INSPIRE coordinators know - how to set up the correct links between data and services. The controlling and updating of MD of National “resources” is on-going activity.
Plans for further simplification of data set and service linking

- **Pros**
  - Less metadata to maintain
  - Less complex implementation

- **Cons**
  - Necessity to update mainly metadata returned by view/download services

Diagram:
- Data set metadata
- Service metadata
- Metadata returned by services (GetCapabilities)
Plans for further simplification of data set and service linking

- **Draft document** describing the change proposal has been developed by JRC
- The change proposal was endorsed by MIG-T
- Technical details are being developed
- Volunteered countries (CY, CZ, EL, ES, FR, IT, LT, NL, PL, SE, SK) work on example implementations
- Initial plan was to finish the activity by the end of 2019
- MS will have a transition period (2 years?) to implement the changes
Monitoring and Reporting

• Dedicated INSPIRE KEN webinar

• From 2020 metadata driven approach
  • Information collected from the metadata (for monitoring and part of reporting)
  • Plan is to harvest metadata once a year and calculate indicators automatically

• New metadata tagging
  • priority data sets
  • spatial scope
Metadata tagging - priority data sets

• Main goal was to develop a list of datasets related to environmental reporting
• INSPIRE Registry was updated with priority data sets controlled vocabulary http://inspire.ec.europa.eu/metadata-codelist/PriorityDataset
• Encoding guidelines have been developed for Metadata TG 2.0 and 1.3
• Consequently, properly tagged priority data sets can be identified
• European Commission opened infringement procedure against 4 EU countries over lack of priority data sets tags in metadata
Metadata tagging – spatial scope

• Main goal was to better identify spatial scope of a data set in metadata
• INSPIRE Registry was updated with spatial scope controlled vocabulary [http://inspire.ec.europa.eu/metadata-codelist/SpatialScope](http://inspire.ec.europa.eu/metadata-codelist/SpatialScope)
• Vocabulary contains 5 keywords values. However, only 2 (national, regional) are planned to become mandatory
• Discussion in some countries how to differentiate (especially between regional and local)
Previous metadata harvesting approach

Limited MS control when metadata was to be harvested by INSPIRE geoportal from national metadata catalogues

Metadata records were harvested sometimes when national infrastructure was not fully operational

Some metadata may be missing in INSPIRE geoportal
Harvesting console (1)

Main features:

- Available for MS **national catalogue admins**
- **Provides** full MS control of the harvesting process including its initiation, monitoring/reviewing and publishing of harvested MD
- **Supports** MS with indications of concrete MD records with possible errors found e.g. missing resource linkages
- **Reduces** the number of requests from MS for the ad-hoc harvesting.
Dashboard

Welcome Marcin GRUDZIEN, here you can manage the Discovery endpoints registered for your country/organization.

**Discovery services**

<table>
<thead>
<tr>
<th>Country</th>
<th>Discovery Service endpoint</th>
<th>Responsible organizations</th>
<th>Last successful execution</th>
<th>Actions</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL</td>
<td>PL_GOV - Geoportal - Polska Usługa Wyszukiwania INSPIRE</td>
<td>Centralny Ośrodek Dokumentacji Geodezyjnej i Kartograficznej</td>
<td>Fri Apr 12 2019 14:06:22 GMT+0200 (czas środkowoeuropejski letni)</td>
<td>Schedule a new harvest</td>
<td>Ready</td>
</tr>
</tbody>
</table>

Showing 1 to 1 of 1 entries

* The time and date provided are based on an automatic estimation. The calculation is based on previous elapsed times. The estimation provided is purely indicative and may not be realistic.
Dashboard

Welcome Marcin GRUDZIEN, here you can manage the Discovery endpoints registered for your country/organization.

### Discovery services

<table>
<thead>
<tr>
<th>Country</th>
<th>Discovery Service endpoint</th>
<th>Licensed organizations</th>
<th>Last successful execution</th>
<th>Actions</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL</td>
<td>PL_GOV - Geoportal - Polska Usługa Wyszukiwania INSPIRE</td>
<td>Centralny Ośrodek Dokumentacji Geodezyjnej i Kartograficznej</td>
<td>Thu May 09 2019 14:55:50 GMT+0200 (czas środkowoeuropejski letni)</td>
<td><a href="#">Check and publish</a></td>
<td><a href="#">Pending publish</a></td>
</tr>
</tbody>
</table>

Showing 1 to 1 of 1 entries

* The time and date provided are based on an automatic estimation. The calculation is based on previous elapsed times. The estimation provided is purely indicative and may not be realistic.
Other relevant info (1)

• INSPIRE reference validator developments
  • Cloud staging instance available at http://staging-inspire-validator.eu-west-1.elasticbeanstalk.com/etf-webapp/
  • Supports among others
    • Metadata v. TG 2.0
  • Tests are still ongoing
Other relevant info (2)

• Plans for new features of INSPIRE Geoportal (June/July 2019)
  • Support of metadata TG 2.0 in INSPIRE Geoportal
  • Automatic translation of metadata (already available for title and abstract)

• Plans for new features of INSPIRE Geoportal (October 2019)
  • Monitoring indicators available
  • Better integration with the reference validator

• Discussion about improving usability – make licensing more homogeneous
  • Initial studies to harmonise identification of open data and licensing information in metadata
Summary

• EC focuses on improving data sets - service links
  • In a current technical and legal framework
  • However, this framework is likely to change in order to further simplify linking

• New code lists for e-reporting

• New harvesting console for national catalogue admin

• Other relevant changes
  • Reference validator
  • INSPIRE Geoportal
  • Usability improvements
Thank you

marcin.grudzien@gugik.gov.pl
References

• Tomas Robert, Hernandez Lorena (2019) Results from the member States feedback to the survey https://webgate.ec.europa.eu/fpfis/wikis/download/attachments/324049586/GeoportalWorkshop_SurveyResults.pdf?version=1&modificationDate=1549033539587&api=v2 accessed on 2019.05.07


• Conclusions from 2016 INSPIRE KEN workshop from producer’s point of view (2016)