ISA Action 1.17: A Reusable INSPIRE Reference Platform (ARE3NA)

Expert contract supporting the
Study on RDF and PIDs for INSPIRE
Deliverable D.EC.1.2

INSPIRE RDF vocabularies – open issues

Clemens Portele

Joint
Research
Centre

This publication is a Deliverable of Action 1.17 of the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Union, A Reusable INSPIRE Reference Platform (ARE3NA), managed by the Joint Research Centre, the European Commission's in-house science service.

Disclaimer

The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Copyright notice

Table of contents

1 Scope 3
2 Known open issues and obstacles 3
3 Transformation of data to RDF triples 5
4 Implications to other INSPIRE components 6
4.1 Code list register 6
4.2 RDF vocabularies 6
4.3 Media type register 6
4.4 Metadata 6
4.5 Discovery services 6
4.6 Download services 6
4.7 Access control 6
4.8 INSPIRE Geoportal 7
4.9 Validators and testing tools 7
4.10 Data policy 7
4.11 Monitoring & reporting 7

1 Scope

The previous report[1] documented and analyzed a potential methodology for transforming INSPIRE application schemas in UML into RDF vocabularies.

This report

- lists known open issues or potential obstacles to the application of the potential methodology to the INSPIRE application schemas;
- briefly describes how INSPIRE-related source data (in their original format and schema) as well as INSPIRE-compliant data (in GML) can be transformed to the generated RDF vocabularies;
- outlines known implications that using RDF as an encoding would have for other INSPIRE components.

2 Known open issues and potential obstacles

The proposed methodology is based on ISO/DIS 19150-2 with a significant number of changes. The proposed methodology is neither complete nor final. It requires broader review and discussion as well as testing in applications.

To be more specific, the first report and subsequent discussion between the experts and in the public webinar have already identified the following specific open issues and obstacles:

- ISO/DIS 19150-2 is not finalized and technical comments have been submitted to ISO/TC 211 as part of the DIS vote. These will be discussed during an ISO/TC 211 meeting in June 2014. Some of the comments overlap with issues raised in the

[1] INSPIRE-RDF-vocabularies-CP-20140425.docx with appendicies

reports, but others do not. At this stage in the ISO process it is not foreseen to raise any new comments, so it is likely that RDF vocabularies for the INSPIRE application schemas would not conform to the final version of ISO 19150-2, if the general direction of the methodology used in this document is used. Nevertheless, the disposition of the DIS comments should be taken into account in the process in INSPIRE.

- In general, the proposed changes to the rules from ISO/DIS 19150-2 should be reviewed.
- The URI scheme for the RDF namespaces for the INSPIRE application schemas needs to be fixed.
- When properties have identical names, but inconsistent ranges or inconsistent semantics, the definition of the conflated property needs to be updated or separate properties with appropriate names have to be created. This requires review and editing of the RDF vocabulary. Similarly, properties should be reviewed, if distinct properties should be conflated into a single property.
- Decide, if properties that reference an open code list should use rdfs:resource or skos:Concept as range.
- Clarify to which upper ontologies for features the INSPIRE spatial object types should be linked (e.g. GeoSPARQL, Core Location, ISO 19150-2, etc.).
- Identify existing RDF types or alternatively define an INSPIRE RDF vocabulary for base types from ISO 19103 that are used by INSPIRE, but which are not covered by ISO 19150-2. Measure is an example.
- Decide about the geometry vocabulary. In the proposed methodology, GeoSPARQL has been used.
- Identify RDF types and properties for the INSPIRE foundation schemas, in particular ISO 19123 and ISO 19156. In this context, it also needs to be taken into account that not all INSPIRE application schemas follow the ISO 19109 rules for application schemas. The lack of proven RDF vocabularies for concepts from INSPIRE foundation schemas is a key obstacle for using RDF for INSPIRE data.
- For several feature attributes and classes in INSPIRE application schemas, commonly used properties and classes from existing RDF vocabularies should be reused. Whenever the semantics of such properties matches that of a feature attribute, the existing property should be used instead. The same applies for classes. This requires review to ensure that the use of items from other vocabularies is appropriate.
    - Example: prov:generatedAtTime / prov:invalidatedAtTime from PROV-O are good candidates for beginLifespanVersion / endLifespanVersion from the INSPIRE application schemas.
- The schema conversion rule for union data types in ISO/DIS 19150-2 is insufficient as it does not handle cases where values are a mix of object or datatypes, or the same value type is used by more than one option.
- The ISO/DIS 19150-2 schema conversion rule maps constraints from the UML model. However, including OCL in the RDF vocabulary is questionable. Probably the most reasonable way would be to include only the documentation of a constraint.
- Association classes are currently not supported by the schema conversion rules in ISO/DIS 19150-2, but are requried for INSPIRE application schemas. The approach taken in GML 3.3 (equivalent transformation of the UML) can be taken here, too.
- Representation of at least the following base types in the INSPIRE RDF vocabularies requires more discussion:
    - SpatialDataSet

    - ThematicIdentifier
    - all types that represent general concepts that are not really INSPIRE-specific or spatial in nature; e.g. types related to documents, citations, contacts and parties
- The representation of multiple versions in the data and potential effects on the RDF vocabularies.
- Any generated RDF vocabulary will require reviewing and additional edits.
- Good guidance and examples are needed that illustrate how feature instances should be represented in RDF as this information is not immediately accessible from the RDF vocabularies. This is in particular the case where other RDF vocabularies are reused or where the automatically generated RDF vocabularies have been edited.
- This should also include guidance about
    - the use of 303 URIs vs Hash URIs,
    - properties with the real-world phenomenon as the subject vs properties with the feature document as the subject,
    - how these two resources (real-world phenomenon and feature document) are related (e.g., whether to use rdfs:isDefinedBy, foaf:isPrimaryTopicOf, etc.),
    - how multiple resources identifying the same real-world phenomenon are related (using owl:sameAs),
    - whether the real-world phenomenon, the feature document or both are typed with an INSPIRE spatial object type classification,
    - the role of HTTP content negotiation and how it may be implemented,
    - how multiple versions should be represented.

Other recommendations and observations that have been identified in the process are:

- Feature properties that represent feature or property metadata should be clearly identified as such in the UML model, for example, by using additional stereotypes.
- The same should be cosidered for thematic identifiers.
- The published INSPIRE UML model seems to be incomplete, some associations are missing.
- The INSPIRE registry should support a SKOS representation for code lists (skos:ConceptScheme) and values in code lists (skos:Concept).

3 Transformation of data to RDF triples

For XML-based source data, including GML data, XSLT stylesheets might be used. We have used this approach to support application/rdf+xml as an additional output format in OGC Web Feature Services.

We know that others have used FME from Safe Software to create RDF from spatial data.

In any case, we are not aware of any easy-to-configure tools for such transformations and expect that either development work or configuration by experts is required for most dataset transformations.

This is in particular the case as soon a dataset needs to be incrementally updated.

4 Implications to other INSPIRE components

4.1 Code list register

The INSPIRE code list registry currently does not support SKOS. SKOS representations for code lists (skos:ConceptScheme) and values in code lists (skos:Concept) should be added to provide proper representations of code lists in the linked data context.

Other code list registries which contain extensions to the INSPIRE code lists should support SKOS, too.

4.2 RDF vocabularies

The RDF vocabularies must be available under their http URIs.

4.3 Media type register

The media types text/turtle and application/rdf+xml should be added to: http://inspire.ec.europa.eu/media-types/.

JSON-LD with geo-support would be another candidate encoding and developments in the community should be monitored.

4.4 Metadata

The ongoing work in ARE3NA to align INSPIRE metadata with DCAT-AP is important to enable that INSPIRE metadata may be provided in a representation that is used and accepted in a linked data context.

4.5 Discovery services

To fit into a linked data environment, discovery services that support representations like DCAT / DCAT-AP and RESTful APIs should be supported. CKAN is an example for a software product that may be used for this purpose (which also include some support for CSW).

4.6 Download services

A GeoSPARQL endpoint would likely be a candidate for a Direct Access Download Services that provides RDF triples. However, details would need to be analyzed as the regulation has a number of detailled requirements would need to be covered in a technical guidance document.

Although INSPIRE distinguishes only between a download of a dataset and the demanding direct access with query support, all http URIs should be required to resolve in order to meet expectations for linked data.

In addition, support for a RESTful API would be helpful for application developers. An example is the Linked Data API[2].

4.7 Access control

Access control mechanisms that are API-aware and/or content-aware (i.e., which parse and/or edit requests and/or responses) need to support .

[2] https://code.google.com/p/linked-data-api/wiki/Specification

However, while this is not a requirement, there is a general expectation that linked data is open data and URIs do not resolve to 403 responses. After all, in order to achieve at least a one star rating on the five-star linked data scale, the data must be available under an open license.

4.8 INSPIRE Geoportal

The INSPIRE Geoportal would have to be updated accordingly, at least the metadata and discovery support.

Directly supporting linked data, too, might require a redesign, although this is difficult to say without knowledge about implementation details.

For open data an integration into the EU Open Data Portal would seem to logical.

4.9 Validators and testing tools

These would need to be updated, if there is demand to support testing and validation of RDF data and metdata.

4.10 Data policy

As RDF is only a representation like GML, there is no obvious implication.

4.11 Monitoring & reporting

I am not familiar enough with the monitoring and reporting processes, workflows and tools in order to comment on implications.