European Commission

Study and guidelines on Geospatial Linked Data as part of ISA Action 1.17

Resource Description Framework

DG Joint Research Center

6th of May 2014

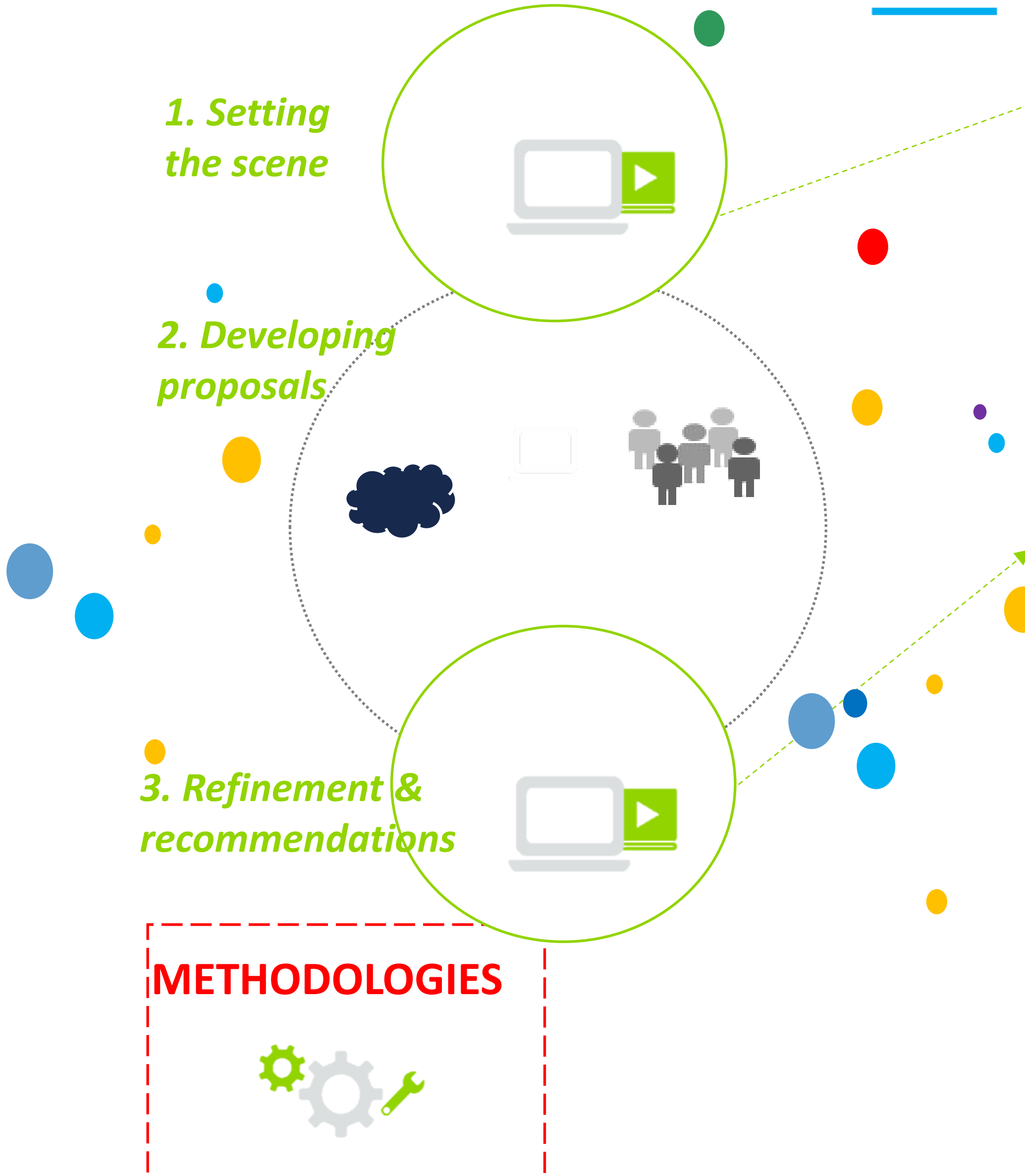
Danny Vandenbroucke
Diederik Tirry

Agenda

European Commission

1. Introduction
2. Context
3. Development of a methodology for INSPIRE in RDF
4. Preliminary results
5. Conclusions and next steps

Introduction



State-of-Play webinar (13th of March)

We provided an outline of the study and our work so far.
This webinar was also an opportunity to provide feedback and to exchange experiences: **where could INSPIRE RDF be used in e-government?**

Today's webinar

Presentation of the preliminary results of the guidelines for transforming INSPIRE application schemas in UML into RDF vocabularies: **opportunities for community-led improvements and next steps towards an official INSPIRE encoding.**

Resource Description Framework (RDF)

Agenda

1 Introduction

2 Context

3 Development of a methodology for INSPIRE in RDF

4 Preliminary results

5 Conclusions and next steps

Context

This study has been prepared in the context of the Interoperability for European Public Administrations (ISA) Programme and, in particular **A Reusable INSPIRE Reference Platform** (ARE3NA, ISA Action 1.17)

This study should provide:

1. Shared evidence about the current status in Europe of linked (geospatial) data related to INSPIRE.
2. An initial common/agreed methodology and guidelines towards **RDF encodings for INSPIRE**
3. Recommendations for how location **PIDs** could be governed for INSPIRE and other relevant activities.

Context

European Commission

INSPIRE

Interoperability of geospatial data sets and services through harmonised data models and encodings for the exchange of data related to one of the 34 spatial data themes

1. Data models using UML at conceptual level
2. Encoding using GML based on encoding rules

Several European project and national initiatives are publishing geospatial data as Linked Data

☞ Using the Resource Description Framework (RDF)

However:

☞ No agreed rules or guidelines on how to create such RDF vocabularies from the UML models

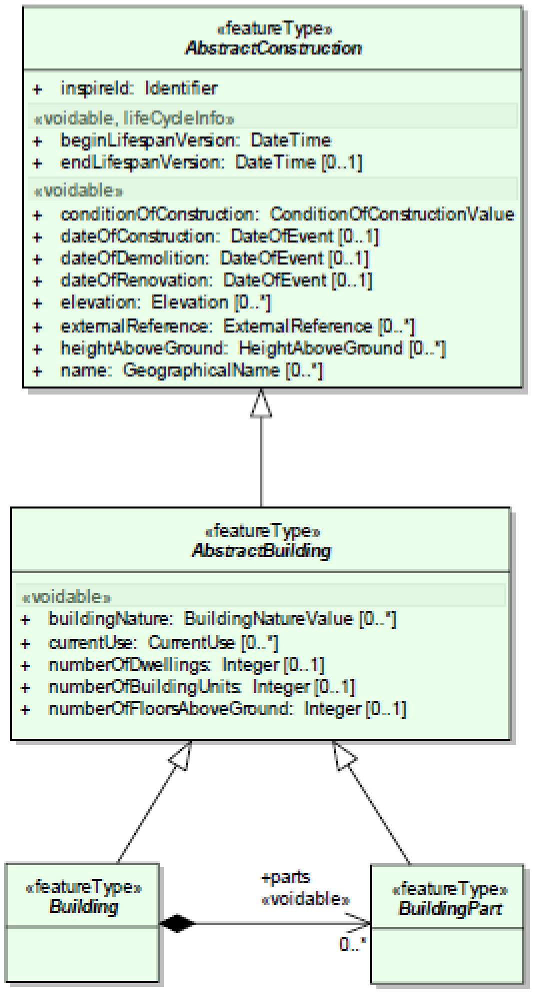


Figure 13: Feature types of Buildings Base application schema

Agenda

European Commission

1 Introduction

2 Context

3 Development of a methodology for INSPIRE in RDF

4 Preliminary results

5 Conclusions and next steps

Development of a methodology

With the support of 3 experts

Who?

- ☞ Linda van den Brink (Geonovum - NL)
- ☞ Stuart Williams (Epimorphics - UK)
- ☞ Clemens Portele (Interactive Instruments - DE)

Why these experts?

- ✓ Their expertise in INSPIRE and Linked Data domain
- ✓ Their involvement in past or ongoing projects related to this topic

How?

- Elaborate a methodology to transform INSPIRE application schemas into RDF
- Participate in **one-day workshop** to compare and discuss
- Apply the methodology to **three INSPIRE Annex Themes**
- Describe potential tools to be used for the transformation of INSPIRE-related data (source as well as INSPIRE-compliant data) to the generated RDF vocabularies.
- Outlining open issues or potential obstacles to the application of the proposed methodology

Development of a methodology

Tested INSPIRE themes

Expert 1	Expert 2	Expert 3
Land Cover	Buildings	Area Management Zones
Transport Networks	Statistical Units	Hydrography
Environmental Monitoring Facilities	Environmental Monitoring Facilities	Environmental Monitoring Facilities

Agenda

1 Introduction

2 Context

3 Development of a methodology for INSPIRE in RDF

4 Preliminary results

5 Conclusions and next steps

Outline key elements

Key elements to be covered in the guidelines

General context

- Scope of the transformation to RDF
- Relevant standards and specifications
- Spatial objects vs real-world things

Specific conversion rules

- Properties and integrity constraints (domain and ranges, cardinality)
- Representing features in RDF
- Codelists
- Versioning of features
- Voidability
- Lifecycle information and other metadata
- Foundation schemas
- External vocabularies

General context

Scope of the transformation to RDF

INSPIRE

- A basis for standardizing and harmonizing spatial objects in thematic domains.
- Characterized by a service based dissemination of (mostly) GML structured data.
- Data specifications provide clear definitions of semantics in predefined domains and use cases.

⇒ The semantics are defined within information domains.
⇒ INSPIRE data could be of use outside its original domain as well

Scope of transformation

☞ Intended use of the RDF vocabularies is to publish structured data that might be linked to data from other domains and that allows other data providers to refer their own data to INSPIRE data.

☞ Support for semantic inferencing or reasoning is out-of-scope as is data validation.

General context

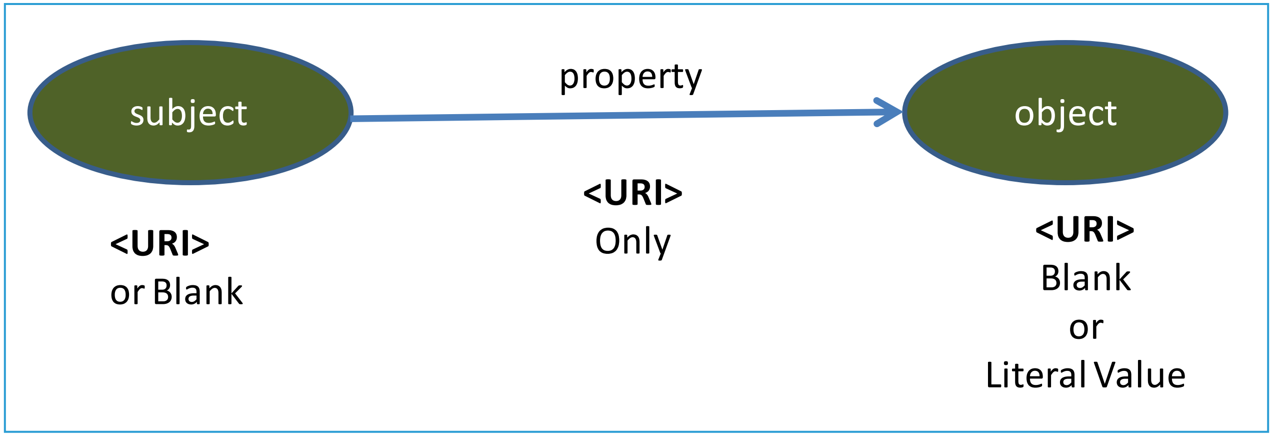
Relevant standards and specifications

- **ISO/DIS 19150-2 - Geographic information — Ontology — Part 2: Rules for developing ontologies in the Web Ontology Language (OWL)**
  - Schema conversion rules as **starting point**
  - ISO/DIS 19150-2 is not finalized and technical comments have been submitted to ISO/TC 211 as part of the DIS vote
  - Show strongly the UML roots and not really reflect common practice in the linked open data world
  - Changes are proposed, but require broader review, discussion and testing

- **Other relevant standards**
  - ISO/DIS 19103:2013 (Conceptual Schema Language), ISO/DIS 19109:2013 (Rules for application schema), ISO 19118:2010 (Encoding)
  - Cool URIs for the Semantic Web, W3C Interest Group Note, 03.12.2008, http://www.w3.org/TR/cooluris/
  - INSPIRE Generic Conceptual Model, INSPIRE Guidelines for the encoding of spatial data
  - GeoSPARQL, NeoGEO, FOAF, ORG, Location Core, Person Core……

General context

Spatial objects vs real-world things



INSPIRE

- No requirement for identifiers (URIs) for real-world phenomena
- Object identifier to identify the abstraction rather than the abstracted thing

The **GFM** does not make a separation between properties where the subject is the real-world object and where the subject is the spatial object

Linked Data / Semantic web

- Separate between real-world thing and an abstraction
- Subject identifier explicitly intended to designate the real-world thing about which statements are being made

General context

Spatial objects vs real-world things

A thing in the world

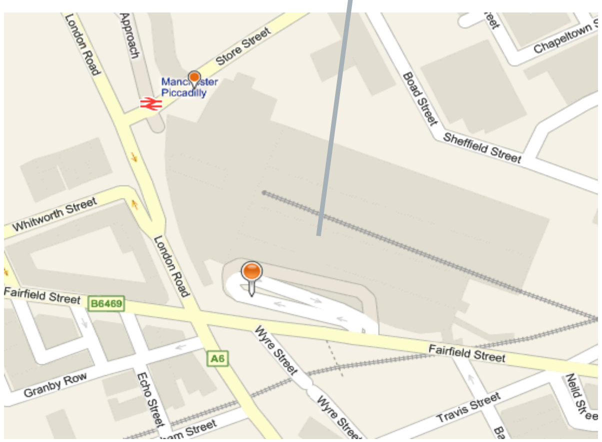
models

models

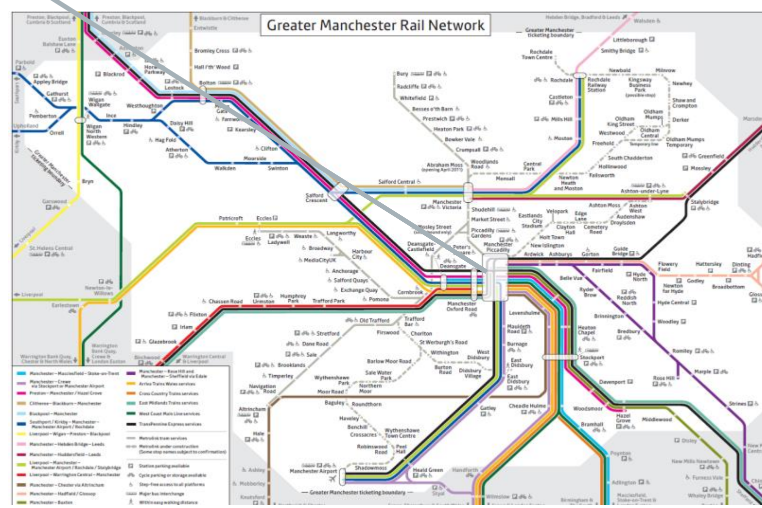
Spatial Object: abstract representation of a real-world phenomenon related to a specific location or geographical area [INSPIRE Directive]

NOTE It should be noted that the term has a different meaning in the ISO 19100 series. It is also synonymous with "(geographic) feature" as used in the ISO 19100 series.

From:
INSPIRE Generic Conceptual Model D2.5 v3.3



An Area on a Map



A Node in a Transport Network

Source: Stuart Williams

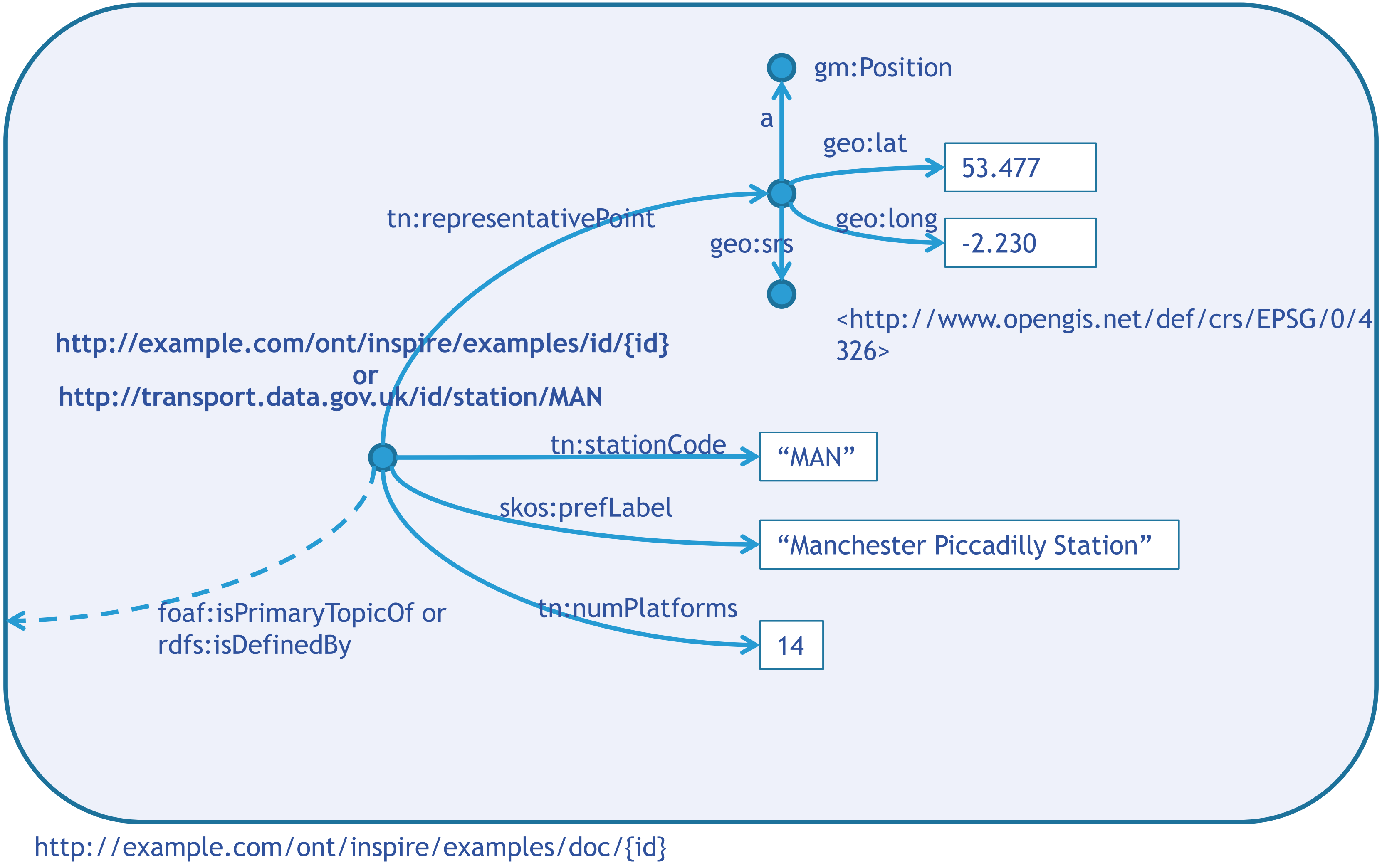
General context

Spatial objects vs real-world things

- Common practice in **Linked Data** is to be clear about 'subjects' and avoid conflation of object with real-world thing i.e. two persistent URIs for each feature
  1. URI for the INSPIRE feature document
     (collection of triples with the real-world phenomenon or the feature as subject)
  2. URI for the real-world phenomenon
- Conclusions and good practices have been documented in the W3C document "Cool URIs for the Semantic Web".
- No requirement that only a single URI is used for the real-world phenomenon - it is fine to use different URIs if there are no reference sets around.
- Related to other key issue 'Lifecycle information and other metadata'

General context

Spatial objects as graphs and nodes



Outline key elements

Key elements to be covered in the guidelines

General context

- Scope of the transformation to RDF
- Relevant standards and specifications
- Spatial objects vs real-world things

Specific conversion rules

- Properties and integrity constraints (domain and ranges, cardinality)
- Representing features in RDF
- Codelists
- Versioning of features
- Voidability
- Lifecycle information and other metadata
- Foundation schemas
- External vocabularies

Specific conversion rules

Properties and integrity constraints

UML	Linked Data
• Closed-world Assumption • UML properties (attribute and associations or more specifically association roles) are scoped to the UML Classes in which they are defined and inherited by subclasses thereof • Properties cannot exist independently	• Open-world Assumption • RDF properties are first class entities that can exist independently • strong cultural imperatives to reuse and share existing terminology ("*skos:prefLabel*", "*skos:altLabel*" and "*rdfs:label*") • may be 'bound' by the use of '*rdfs:domain*' and 'rdfs:range' statements

Specific conversion rules

Properties and integrity constraints

- ISO/DIS 19150-2
  - Range is mandatory
  - Domain may be used but no guidance when to include or not
  - Different options to make ‘domain’ open ('*rdfs:Resource*' , '*owl:Thing*‘, ‘*owl:Class*’….)

- According open-world view preference for **less context in the properties seems appropriate**

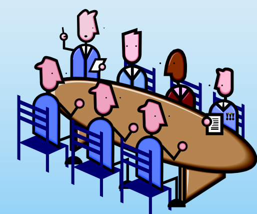
- Need to consolidate properties with conflicting names
  - Semi-automatic
  - Option to make use of unions

- Cardinality restrictions offer little value for syntactic validation in the context of RDF
  -> to be suppressed

Specific conversion rules

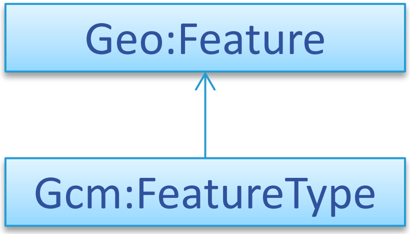
Representing features in RDF

- ISO/DIS 19150-2
  - feature type classes should be sub-classes of gfm:AnyFeature (General Feature Model vocabulary) and iso19150-2:FeatureType (ISO 19150-2 vocabulary)



Currently under discussion !)

- The OGC standard GeoSPARQL specifies with geo:Feature another class that is similar to gfm:AnyFeature.



- Other possible vocabularies are Location Core, and NeoGeo

Specific conversion rules

Codelists

Conversion depends whether they are part of the application schema

1. INSPIRE core application schema

    - Codelists are managed separately from application schemas and are managed in the INSPIRE code list register and other registers. It is therefore inappropriate to include classes and SKOS concept schemes for code lists in the RDF vocabularies.

    - Use skos:Concept as their range unless the tagged values "vocabulary" is a http or https URI.

    - The INSPIRE registry currently does not support a SKOS representation of the INSPIRE code lists.

Specific conversion rules

Codelists

Conversion depends whether they are part of the application schema

2. Embedded in application schema (e.g. extension)

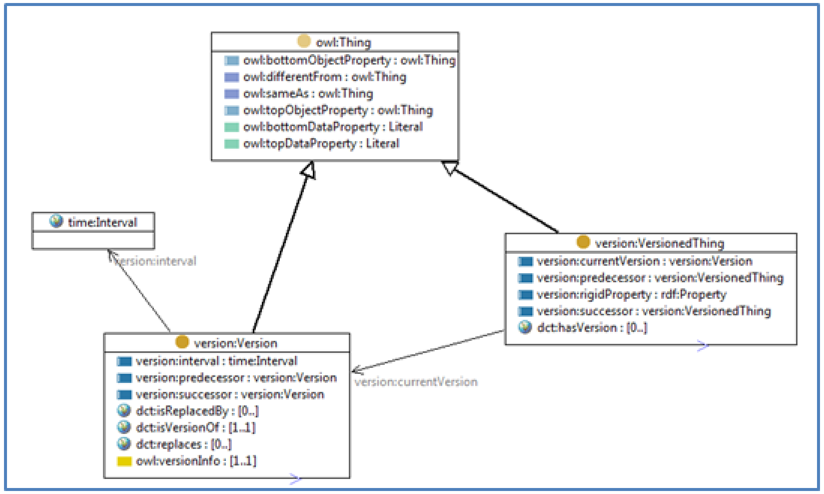
- Codelist and controlled vocabularies can be transformed into SKOS concept schemes.
- All code points are members of the scheme using skos:inScheme.
- They are also made instances of a distinguished subclass of skos:Concept which can be use to restrict the range of a property.
- An open domained property can be defined to make use of the code with an arbitrary entity.
- Linked to properties can be done using “owl:oneof” although this is more appropriate to enumerations.

Specific conversion rules

European Commission

Versioning of features

- ❑ Known open issue in general as the base standards from **ISO/TC 211 and OGC do not natively support versioning**
- ❑ RDF/OWL is not different.



Example of complex framework (source: Stuart Williams)
Enduring spatial-objects with versioned temporal parts

- ✓ As a result, one typically has to "build" its own framework on top of the existing standards and technology, which in practice is a problem.

- ✓ It is obvious that support for the history of objects adds a new level of complexity and the added complexity needs to be balanced with the requirements and priorities.

Specific conversion rules

European Commission

Voidability

Allows to state explicitly that

- something, for example the name of a road, is not known
- a road is known to have no name.

INSPIRE application schemas, although generally based on the closed-world assumption, support unknown facts.

Voidable properties present a certain amount of difficulty to RDF under the open-world assumption. Just because the value of a property may not be given does not mean that there is no value for that property that could be given elsewhere. RDF has no proper mechanism (that we are aware of) to state that a road is known to have no name.

- ✓ **No conversion rule needed for concept of voidability**
- ✓ **If voidreason is required a possible approach relies on the creation of a codelist for void reasons through SKOS**

Specific conversion rules

Lifecycle information and other metadata

In INSPIRE, most properties are properties that describe the real-world phenomenon. However, there are exceptions:

- Properties that represent **life-cycle information**
- Properties that have a value type from ISO 19115 are often **feature metadata**.
- Properties that require a closer review to identify them as feature metadata.

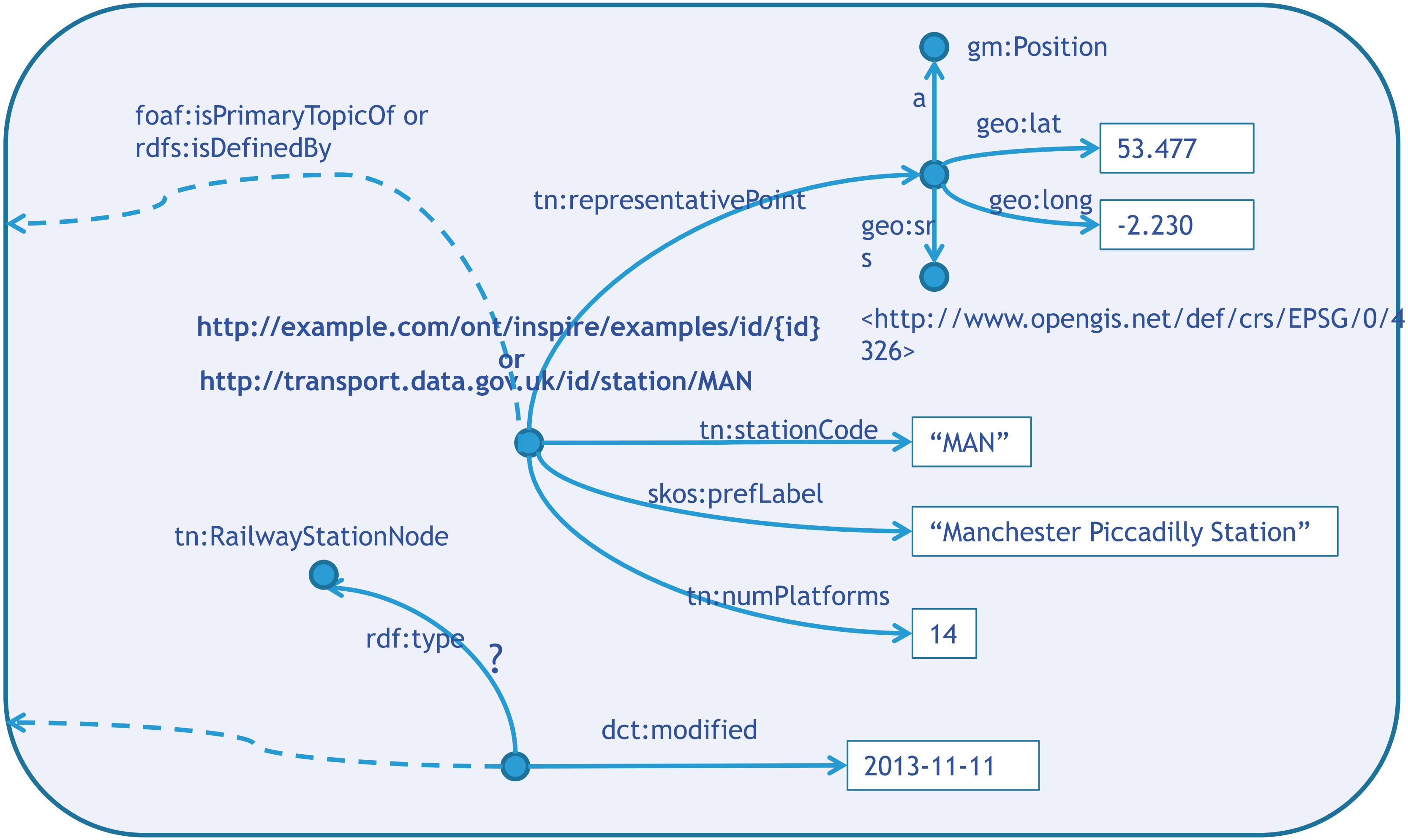
  Examples are CadastralZoning.estimatedAccuracy or CadastralZoning. originalMapScaleDenominator that are not properties of the real-world phenomenon, but of the feature.

From the perspective of the RDF vocabularies there is no distinction between the two types of properties, because the rdfs:domain is not included

Impacts how instances are represented in RDF as it is important in linked data and the semantic web to be clear about the subjects.

Specific conversion rules

Lifecycle information and other metadata



http://example.com/ont/inspire/examples/doc/{id}

Specific conversion rules

Foundation schemas

- The INSPIRE application schemas converted to RDF make use of types from ISO 10103, 19107, 19108, ISO 19111, ISO 19115, ISO 19123 and ISO 19156.
- No sufficiently mature and tested RDF vocabularies exist, which is a problem for any attempt to represent INSPIRE data in RDF at this time.
- Some are available at http://def.seegrid.csiro.au/static/isotc211/, but also differ significantly from RDF vocabularies that would be created using the ISO/DIS 19150-2 schema conversion rules.

→ For the purpose of the schema conversion of the INSPIRE application schemas, owl:Class has been proposed by default for all types without a known, more specific class.

Specific conversion rules

External vocabularies

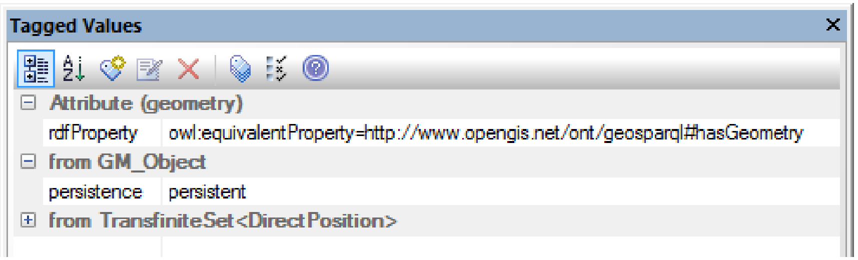
- For several feature attributes and classes in INSPIRE application schemas, commonly used properties and classes from existing RDF vocabularies should be reused.

- This **requires review** to ensure that the use of items from other vocabularies is **appropriate**.

- The RDFS, FOAF, ORG, vCard, PROV, DC, DCT and ISA Core are vocabularies that are commonly used in the linked-data community for

  - Naming (rdfs:label) e.g. GeographicalName
  - Representing people, roles and organisational structure.
  - Describing metadata and provenance

- Potential method for automated conversion is annotation via **UML tagged values**



Open issues

Known open issues

- ❑ modelling conventions in UML may result in complex UML constructions (e.g. multiple inheritance) which might be difficult to map to other representations.

➡ Modelling conventions do not make any sense in the OWL domain and could be expressed in a better way in OWL, representing the real world more closely than is possible in UML.

- ❑ The schema conversion rules in ISO/DIS 19150-2 are inappropriate or incomplete for:
  - Composition and aggregation
  - For union data types
  - Association classes
  - OCL constraints

- ❑ Generation of instance data

- ❑ vocabulary management issues, etc.…

Agenda

European Commission

Conclusions

European Commission

- ISO/DIS 19150-2 is not finalized
  - **Disposition of the DIS comments** should be taken into account
  - Series of issues from experiment that are not part of ISO/DIS19150-2

- RDF vocabularies that strongly show UML roots do not really reflect **common practice in the linked open data world**.

- Any automatically generated RDF vocabulary will require reviewing and additional edits due to their **specific context**.

- Common practice needed with respect to the use of external vocabularies e.g. geometry representation in RDF. **Dependency on other communities**.

- Mentioned issues require **broader review and discussion as well as testing** in applications.

- 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.

Conclusions

- ❑ The results of the experiment gave insight into the **challenges** of transforming INSPIRE data in RDF.
- ❑ Given the difficulties faced it is not surprising therefore that a **definitive set of guidelines** on how to transform INSPIRE to RDF **cannot be** given yet.
- ❑ Further review and discussion is needed.

Feedback

- If there is anything relevant missing from our first list of issues/questions to be tackled please say so
- What are the priorities?

contact us: are3na@jrc.ec.europa.eu

Next steps

European Commission

Guidelines on methodology

- ❑ Processing the results of the experts
- ❑ Compiling and consolidating input received
- ❑ Final document "Guidelines on methodologies for the creation of RDF vocabularies representing the INSPIRE data models and the transformation of INSPIRE data into RDF" by end of May

AND….

At any time, your feedback is very appreciated.

Next steps

Thank you for your participation!

ARe³NA

Connecting Geospatial Data

Join the collaborative platforms in other areas of Open Source & Semantic Interoperability on https://joinup.ec.europa.eu/