S100 WG2 9.3B

Paper for Consideration by S-100 WG

Interest of semantic web standards for S-100 developments

Submitted by:	France (Shom)
Executive Summary:	This paper proposes a possible evolution of the S-100 GI Registry toward semantic web technologies and RDF-based data model, in order to increase interoperability, to facilitate information sharing and cross-domain applications. We propose to develop a technical cooperation with the relevant department of the European Commission (Joint Research Center – JRC) regarding the implementation of semantic web standards for INSPIRE. Considering these contributions, a road map could be part of the input to the upcoming meeting HSSC 9.
Related Documents:	Action HSSC 8/18; S-100; S-99.
Related Projects:	

Introduction / Background

S-100, adopted in 2010, is the framework geospatial standard for hydrographic and related data. It is compliant with the ISO 19100 series of geographic standards, in order to improve data interoperability. S-100 is underpinned by a Registry, composed of current Registers based on ISO 19135 (Feature Concept Dictionary; Portrayal; Metadata; Product Specification; Data Producer Code). The IHO owns and manages the Registry (http://registry.iho.int).

The IHO Geospatial Information (GI) Registry is supporting the harmonization effort between the different S-100 geospatial products. The information contained in the IHO GI Registry and its component Registers is authentic, authoritative and form the cornerstone of e-navigation services. There is therefore a great need to make it available for wider reuse, separately or in combination with base registries from other sectors.

The diagram below illustrates the concept of e-navigation. The main component is the Maritime Cloud framework that enables efficient, secure and reliable information exchange in and around the maritime sector. This framework relies on a service registry, where authenticated service providers (registered in the Identity Registry) give users access to services, for example navigational warnings, weather forecast, etc. In order to be compliant with such an infrastructure, the information contained in the IHO GI Registry should be accessible in a standardized way so as to build the future services.

From http://efficiensea2.org/beta-release-of-the-maritime-cloud-and-balticweb/

In order to achieve this aim, some amendments to the current GI Registry and its component Registers are proposed in this paper. They rely on semantic web technologies, so the first part of this paper reviews the characteristics of semantic web standards and especially the RDF (Resource Description Framework) standard, adopted as a W3C recommendation in 1999 (last edition (1.1) published in 2014). The use of this technology is then illustrated by a practical example of implementation for the INSPIRE Registry and Registers. This paper then highlights the best practices and the benefits of building an RDF-based data model, focusing on the IHO S-100 GI Registry and Registers.

Analysis/Discussion

What is "semantic web"?

Semantic web or "Web of data" is an extension of the web that "provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries (Tim Berners-Lee)". While classic Web, mainly based on the HyperText Markup Langage standard (HTML) is understandable by a human reader, semantic Web is designed to provide online data that can be used by computer programs, not only for display, but for automation, integration and reuse across various applications; existing information can be automatically processed to "make sense" and produce knowledge.

First, information has to be structured in a knowledge database consisting of a network that represents semantic relations between concepts (vertices represent concepts and arcs the relations that bind the concepts). Semantic networks are knowledge graphs that allow reasoning on knowledge. These formal representations are used for expressing ontologies – i.e. explicit specification of conceptualization (ontology describes a domain; knowledge database (based on an ontology) describes a particular state of affairs).

The architecture of Semantic Web is based on web protocols and standard languages:

- HyperText Transfer Protocol (HTTP) for communication between the browser and the server;
- Uniform Resource Identifiers (URI) - unique identifier of a resource on a network;
- XML language.

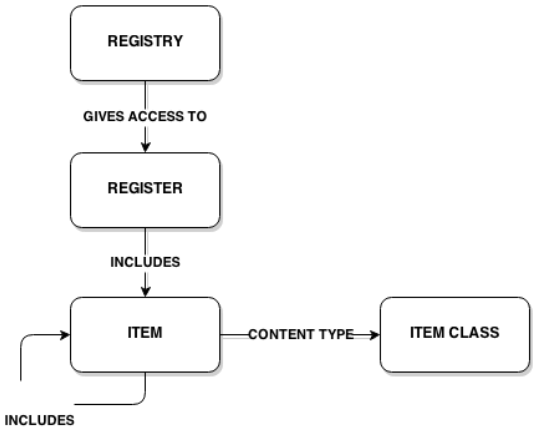
with the addition of the following particular standards:

- RDF (Resource Description Format) that is both a model and a language to express RDF triple: subject (the resource appointed by its URI), predicate (property applicable to the resource, it belongs to a vocabulary from a namespace), object (a property value which can be another resource or a literal);
- RDF Schema: language designed to create vocabularies;
- OWL: language designed to create ontologies;
- SPARQL: query language for RDF data manipulation.

An example of implementation: the INSPIRE Registry and Registers

The INSPIRE infrastructure involves a number of items, which require clear descriptions and the possibility to be referenced through unique identifiers. Examples for such items include INSPIRE themes, code lists, application schemas or discovery services. Registers provide a mean to assign identifiers to items and their labels, definitions and descriptions (in different languages). The INSPIRE Registry provides a central access point to a number of centrally managed INSPIRE Registers. The content of these Registers is based on the INSPIRE Directive, Implementing Rules and Technical Guidelines. (From http://inspire.ec.europa.eu/registry/).

The data schema of the content of the Registry ("Re3gistry") consists of 4 main entities, depicted in the following diagram:



The *Registry* is the service, whereas entities *Register* and *Item* denote how the content served by the *Registry* is organised. As shown in the diagram, *Items* may include other *Items* (referred to as *child items*). This is, e.g. the case of code lists, which are modeled as *Items*, and code values, which are child *items* of code lists.

Entity *Item class* is meant to denote the *type* of content corresponding to an *Item*. Version 4 of the INSPIRE Registry includes the following ones: application schemas, (metadata) code lists, documents, feature concepts, glossary items, themes.

The proposed approach is as follows:

- The W3C Data Catalog vocabulary (DCAT) is used to model entities *registry* (dcat:Catalog) and *register* (dcat:Dataset) – DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the web.
- The Provenance Vocabulary is used to model entity *item* (prv:DataItem) – The Provenance Vocabulary Core Ontology (prv) provides the main classes and properties to describe provenance of data on the web.
- The W3C Simple Knowledge Organization System (SKOS) to model entity *item class* (skos:Concept, or skos:ConceptScheme when an item has child items), with the only exception of the item class concerning reference documents, where the Friend of a Friend vocabulary (FOAF) is used instead (foaf:Document).

Source : https://ies-svn.jrc.ec.europa.eu/projects/registry-development/wiki/RDF_format

The Registry contains 9 Registers:

INSPIRE application schema register	INSPIRE feature concept dictionary	INSPIRE metadata code list register
INSPIRE code list register	INSPIRE glossary	INSPIRE reference document register
INSPIRE enumeration register	INSPIRE layer register	INSPIRE theme register

Every item of the Registry is defined by a unique reference that is the URL of the webpage describing the item. It goes from the items of the Registers to the elements describing the items. For example the statuses (valid, invalid, retired...) have their own webpage explaining what they are.

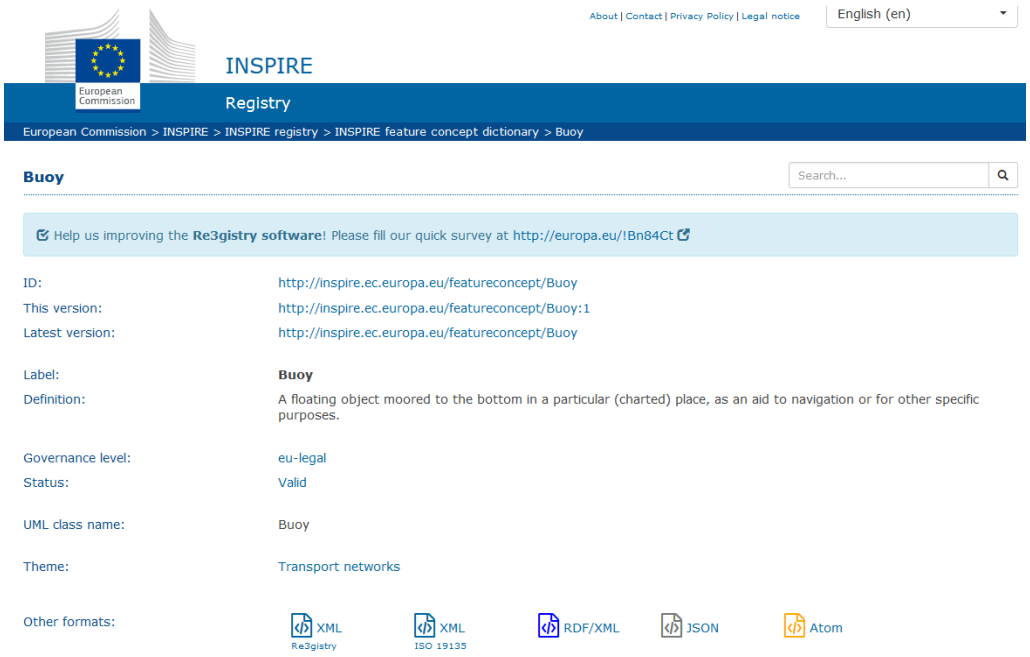
In the end, it makes Registers really clearer because everything is determined, and whatever the element it is automatically linked to its definition.

For example, INSPIRE has the feature concept *Buoy* that appears in the Transport Network theme of the Feature Concept Dictionary Register.

On the *Buoy* webpage, you get first the ID and the version, then the label and the definition (no links there). Then comes the governance and the status, clicking on it takes you to the definition of those terms. As the concept is part of the INSPIRE Transport networks theme it also appears there and there is a link to the item Transport networks from the INSPIRE theme Register.

Last but not the least you get the different formats offered to view this page. These are mostly machine to machine format, that way it can be read by an outside application that works on schema translation.

All of those relations, elements of description and links are possible thanks to the use of controlled vocabularies, RDF and unique IDs as URLs for every element.



Buoy webpage : http://inspire.ec.europa.eu/featureconcept/Buoy

Transport Network webpage : http://inspire.ec.europa.eu/theme/tn

Best practices from INSPIRE Registry

In brief, what is important to consider from the INSPIRE example is that:

- Everything is defined be it a concept, item or objects from detailing pages (domain, status, reference…) and it has a unique ID;
- Everything is available as RDF format so it can be reused by other applications even without human intervention; it's worth noting that the Registry also supports XML, JSON and Atom;
- The registry allows multilingualism.

From the INSPIRE experience, we can see that the Registry and associated Registers are built for the purpose of sharing data. This point has to be considered for the future development of the IHO GI Registry.

Conclusions

As a conclusion, the principal benefits of building an RDF-based data model and use semantic web standards can be summed up as follows:

- Increases discoverability and enable applications to consume metadata from multiple catalogs;
- Promotes interoperability across sectors; in the absence of interconnection, several base Registries may hold the same data, generating inconsistencies and uncertainty;
- Provides the advantage of machine-readable data to develop new applications (Artificial Intelligence, aggregated learning, …);
- Contributes to the build-up of a "seamless infrastructure".

All of this is consistent with the aims of S-100.

Action Required of S-100 WG

The S-100 WG is invited to:

a. Take note of this proposal.
b. Develop technical cooperation with the JRC (Joint Research Center) department of the European Commission (complement feedback from European initiatives linked to INSPIRE : see in particular the work done by ARE3NA community : https://joinup.ec.europa.eu/community/are3na/).
c. Consider the proposal to improve interoperability of the S-100 GI Registry and draw up a road map for adoption in HSSC 9.
d. Report to IHO-EU Networking Working Group (IENWG).