Serving vocabularies in the semantic web

Simon J D Cox, Terry Rankine
CSIRO Earth Science & Resource Engineering, Perth, WA, Australia, {simon.cox}{terry.rankine}@csiro.au

SUMMARY
SISSvoc3 is the latest iteration of the vocabulary service developed by CSIRO as part of the Spatial Information Services Stack. SISSvoc3 has been completely redesigned as a set of URI templates providing a SKOS API that can be deployed to complement lower-level APIs such as SPARQL. We have developed a SISSvoc3 implementation using the Linked Data API which provides useful additional functionality including content negotiation, view construction, and paging. We compare SISSvoc3 to similar vocabulary service APIs from the European Environment Agency and the UK Natural Environment Research Council.

VOCABULARY SERVICES ARE ESSENTIAL
Shared vocabularies are a basic element in interoperable systems. Terms are used as classifiers in various positions in systems, data, interfaces, etc. Most natural sciences have classification exercises at their core, and all academic disciplines require classification of various resources. The publication of classification systems is therefore a key element in the formalization of our understanding of the natural world.

In the eResearch context it is necessary for vocabularies to be available at run-time in forms compatible with various analytical systems and interfaces. Where the vocabularies are shared by a distributed community this implies the use of web technology. Given the ubiquity of vocabularies or classifiers in systems, vocabulary services are effectively the bottom of the interoperability stack.

VOCABULARY FORMALIZATION AND ACCESS
While vocabularies were traditionally conceived of as lists of terms, in contemporary knowledge organization systems the concept is central, with terms appearing as labels. SKOS formalizes this as an RDFS application, and provides a bridge from traditional vocabulary and thesaurus applications to formal logic as represented by OWL. Many existing vocabularies now have SKOS representations. These are provided to human users through a form interface. For consumption by applications access to a vocabulary at various levels of granularity may be provided by standard protocols:

1. A vocabulary as a whole, as a single resource (document), can be provided in response to HTTP GET on a URI for the vocabulary document
2. The items in the vocabulary can be provided one at a time using HTTP GET on the item URI
3. The SPARQL Protocol may be used to provide access to summaries and subsets, if the vocabulary is published at a SPARQL endpoint

SKOS-BASED VOCABULARY SERVICE
However, a vocabulary service should leverage the vocabulary model, with selection through the specific semantics provided by SKOS. SISSvoc3 [4] is a vocabulary interface defined as a set of standard URI templates. The templates support selection of a vocabulary subset based on the standard annotation properties (skos:prefLabel, skos:altLabel, rdfs:label) and semantic relations (skos:broaderTransitive, skos:broader, skos:narrower, skos:narrowerTransitive).

CSIRO has developed a SISSvoc3 implementation using the Linked Data API (LDA) [2] which binds to the vocabulary published through a SPARQL endpoint. This vocabulary service essentially provides a SKOS API to hide the complexity of the generic SPARQL interface. And by using LDA we also get access to content-negotiation, alternative views, paging, metadata and other functionality provided in a standard way.

COMPARISON WITH OTHER VOCABULARY SERVICES
A number of SKOS-based vocabulary-access APIs have been developed. In the environmental-science domain two well-known ones are the GEMET Web Service API (GEMET) from the EEA [1], and the NERC Vocabulary Service v2 (NVS2), developed by BODC. [3] Comparing the specific features of these with SISSvoc3 we find the following.

DATASOURCE
GEMET and NVS2 vocabulary service applications are tightly coupled to a local relational datastore. A SKOS representation is generated on output only.

A SISSvoc3 service binds to its content through a SPARQL endpoint. Hence, a SISSvoc3 service hosted by one organization can be provided for SKOS data published by any other organization. SISSvoc3 provides a complementary interface to content also available on the web through the lower-level interfaces described
above. SISSvoc3 encourages provision of vocabulary data through a suite of interfaces suitable for different applications.

**REST vs SOAP**

NVS2 provides access only to vocabulary subsets that have been pre-defined as SKOS Collections on the REST interface. Specific URI patterns are required within a vocabulary, and each concept can only be associated with a single collection. Queries are supported only through a SOAP interface.

SISSvoc3 and GEMET queries are formulated as URIs.

**URI DOMAINS AND PATTERNS**

All concepts served by NVS2 are in the same URI domain as the service. SISSvoc3 and GEMET may serve content whose identifiers are in any URI domain.

GEMET request URIs use verb-like tokens such as ‘getConcept’ embedded in the request URI. SISSvoc requests are encoded as URIs following the LDA patterns, so `http://{service}/concept` identifies a set of concepts, etc.

**SISSvoc DEPLOYMENTS**

SISSvoc has been deployed by CSIRO to serve a number of vocabularies required by AuScope and related services. These are used to construct user interfaces and to configure data conversion routines. SISSvoc has also been used as a generic registry interface, and for serving gazetteer information.

The exact behaviour of a SISSvoc3 service is completely dependent on how the concepts in a vocabulary are arranged into concept-schemes and collections, the particular set of semantic relations that are asserted in the vocabulary, and the logical closures that are provided by the SPARQL service underneath the SISSvoc service.

**REFERENCES**

3. NVS 2.0 http://vocab.nerc.ac.uk/

**ABOUT THE AUTHORS**

**Simon Cox** was awarded a PhD in Geophysics from Columbia University, New York in 1987. He joined CSIRO to continue work on experimental rock mechanics, spent a few years lecturing at Monash University, then rejoined CSIRO to work on web delivery of the results of the AGCRC and pmd*CRC. Recognizing both that that the technologies were immature, and that the solutions would be applicable more broadly than the earth sciences, he became involved in standardization activities through the Dublin Core Metadata Initiative, the Open Geospatial Consortium and ISO Technical Committee 211 (Geographic Information). He now specializes in information modeling, semantics and related web technologies, and is editor or co-editor of a number of international standards. He has been on many national and international committees and working groups, several of which he has chaired. He has consulted to numerous organizations in Australia and UK. Recently he spent a year seconded to the European Commission working on the INSPIRE environmental data interoperability initiative. The vision of standards-based web-hosted information and processing services is being realized through the Auscope and WIRADA projects in the Australian context for the geoscience and water informatics domains. Simon was awarded the OGC’s Gardels Medal in 2006.

**Terry Rankine** joined CSIRO in 2004 with a BSc in Chemistry from Curtin University. He now specializes in ICT applications in geoscience, and leads the Computational Geoscience research group at CSIRO Earth Science and Resource Engineering.