SKOS Core: Simple Knowledge Organisation for the Web

Alistair Miles
CCLRC Rutherford Appleton Laboratory
Tel: +44 1235 445440
Fax:+44 1234 445831
a.j.miles@rl.ac.uk

Brian Matthews
CCLRC Rutherford Appleton Laboratory
Tel: +1235 446648
Fax:+44 1234 445831
b.m.matthews@rl.ac.uk

Michael Wilson
CCLRC Rutherford Appleton Laboratory
Tel: +44 1235 446619
Fax:+ 44 1234 445831
m.d.wilson@rl.ac.uk

Dan Brickley
World Wide Web Consortium
danbri@w3.org

Abstract:
This paper introduces SKOS Core, an RDF vocabulary for expressing the basic structure and content of concept schemes (thesauri, classification schemes, subject heading lists, taxonomies, terminologies, glossaries and other types of controlled vocabulary). SKOS Core is published and maintained by the W3C Semantic Web Best Practices and Deployment Working Group. The main purpose of this paper is to provide an initial basis for establishing clear recommendations for the use of SKOS Core and DCMI Metadata Terms in combination. Also discussed are management policies for SKOS Core and other RDF vocabularies, and the relationship between a 'SKOS concept scheme' and an 'RDFS/OWL ontology'.

Keywords:
Knowledge organization systems, KOS, taxonomies, thesauri, classification schemes, glossaries, RDF, OWL, semantic web, metadata vocabularies.

Introduction
SKOS Core (1, 2) is a model for expressing the basic structure and content of concept schemes. The term 'concept scheme' is used here to mean 'a set of concepts, optionally including semantic relationships between those concepts.' Thesauri, classification schemes, subject heading lists, taxonomies, terminologies, glossaries and other types of controlled vocabulary are all examples of concept schemes.

The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF) (3, 4, 7). It consists of a set of RDF properties and RDFS classes that can be used to express the content and structure of a concept scheme as an RDF graph. Using RDF allows data to be linked to and/or merged with other RDF data by Semantic Web applications. In practice, this means that data sources can be distributed across the web in a decentralised way, but still be meaningfully composed and integrated by applications, possibly in novel and unanticipated ways.

SKOS Core is developed and maintained by the W3C Semantic Web Best Practices and Deployment Working Group (8). Publication of the SKOS Core Guide (1) and the SKOS Core Vocabulary Specification (2) as W3C First Public Working Drafts is imminent at the time of writing. The current intention of the Working Group is that the SKOS Core Guide and the SKOS Core Vocabulary Specification will become W3C Working Group Notes (29). However the possibility that these documents should form the basis of a W3C Recommendation Track work item is being actively considered, and feedback as to the appropriate level of standardisation would be welcomed.

This paper introduces the SKOS Core Vocabulary, with some applied examples presented in the RDF/XML serialisation syntax (5). The paper goes on to discuss some further usage scenarios involving additional features of SKOS Core. The relationship between SKOS Core and RDFS (6) / OWL (9)(10)(11) is discussed. Finally the relationship between the SKOS Core Vocabulary and the DCMI Metadata
Terms (12) is explored, as an initial basis for the development of clear recommendations for combined usage.

Throughout this paper, prefixes such as 'skos:', 'dc:' and 'foaf:' are used to abbreviate URIs. The prefix 'skos:' stands for 'http://www.w3.org/2004/02/skos/core#', therefore 'skos:prefLabel' is an abbreviation of 'http://www.w3.org/2004/02/skos/core#prefLabel'. The following prefixes are also used: 'dc:' for 'http://purl.org/dc/elements/1.1/', 'dct:' for 'http://purl.org/dc/terms/', 'rdf:' for 'http://www.w3.org/1999/02/22-rdf-syntax-ns#', 'rdfs:' for 'http://www.w3.org/2000/01/rdf-schema#', 'owl:' for 'http://www.w3.org/2002/07/owl#' and 'foaf:' for 'http://xmlns.com/foaf/0.1/'. Note also that the xml:base attribute provides a URI base for relative URIs within RDF/XML examples.

A Glossary in RDF

This section discusses an example of the use of the SKOS Core Vocabulary to express the content of a glossary as an RDF graph. The example is adapted from the W3C Process Document Glossary in RDF (16).

The example illustrates the use of the skos:Concept class. This class may be used to assert that a resource is a conceptual resource, i.e. is a concept. The skos:prefLabel property is used to assert the preferred lexical label for a resource. The skos:definition property, one of a family of 'documentation properties' in the SKOS Core Vocabulary, is used to assert a definition for the meaning of the given resource.

A Taxonomy in RDF

This section discusses an example of the use of the SKOS Core Vocabulary to express the content and structure of a simple taxonomy. The example is adapted from Morten Frederiksen's SKOS Concept Scheme (17) which is used to categorise weblog entries according to their subject.

The example illustrates the use of the skos:broader and skos:narrower properties. These properties are part of a family of 'semantic relation properties' which are used to assert relationships of meaning between concepts. The properties skos:broader and skos:narrower are each others inverse, and may be used to assert a generalisation/specialisation relationship between two concepts, where the meaning of the narrower concept falls completely within the scope of the broader concept.

The example also illustrates the use of the skos:ConceptScheme class to identify a concept scheme, and the skos:inScheme property to assert that a concept participates in a concept scheme.

A Thesaurus in RDF

This section discusses an example of the use of

```xml
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:skos="http://www.w3.org/2004/02/skos/core#"
 xml:base="http://www.w3.org/2003/03/glossary-project/data/glossaries/"
>
 <skos:Concept rdf:about="#w3CRecommendationREC">
  <skos:prefLabel xml:lang="en">W3C Recommendation (REC)</skos:prefLabel>
  <skos:definition xml:lang="en">A W3C Recommendation is a specification or set of guidelines that, after extensive consensus-building, has received the endorsement of W3C Members and the Director. W3C recommends the wide deployment of its Recommendations. Note: W3C Recommendations are similar to the standards published by other organizations.</skos:definition>
 </skos:Concept>

 <skos:Concept rdf:about="#workingGroupNote">
  <skos:prefLabel xml:lang="en">Working Group Note</skos:prefLabel>
  <skos:definition xml:lang="en">A Working Group Note is published by a chartered Working Group to indicate that work has ended on a particular topic. A Working Group MAY publish a Working Group Note with or without its prior publication as a Working Draft. W3C MAY also publish "Interest Group Notes" and "Coordination Group Notes" for similar publications by those types of groups. Interest Groups and Coordination Groups do not create technical reports that advance toward Recommendation.</skos:definition>
 </skos:Concept>

</rdf>
```
the SKOS Core Vocabulary to express some content from a thesaurus. The example is adapted from the UK Archival Thesaurus (18).

The example introduces the skos:altLabel property, which may be used any number of times to assert alternative lexical labels for a concept. The skos:related property is also introduced, which is another semantic relation property used to assert an associative relationship between two concepts (19). Finally, the property skos:scopeNote is similar to skos:definition in that it is a 'documentation property' used to explain the meaning of a concept. However, whereas a definition should attempt to completely explain the meaning of a concept, a scope note may consist of partial information about the meaning of the concept, in terms of what is in or out of 'scope' for the concept.

A Classification Scheme in RDF

This section discusses an example of the use of SKOS Core for expressing a 'classification scheme' in RDF. The example is adapted from (25) and (23).

Note the use of the skos:hasTopConcept property to assert a relationship between a concept scheme and a concept which is a top-level concept in that scheme. This gives applications an efficient way of locating the top-level (i.e. broadest) concepts for a given scheme in an open world context.

Most 'classification schemes' are appropriately modeled as a broader/narrower generalisation hierarchy of concepts, and not as a subsumption hierarchy of classes. Therefore SKOS Core is often a more suitable vehicle for expressing a classification scheme as an RDF graph than RDFS or OWL would be.

More SKOS Core Features

The sections above introduce some of the basic features of the SKOS Core Vocabulary. Some additional features are mentioned here, for a full description and guide to recommended usage at the time of writing see (1) and (2).

Some thesauri use a structural feature known as 'arrays' with 'node labels' (19). A node label is essentially a label for a grouping construct, which is inserted into the main term hierarchy to make browsing easier. There is

```xml
<rdf:RDF
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xml:base="http://www.wasab.dk/morten/blog/archives/author/mortenf/skos.rdf">
  <skos:ConceptScheme rdf:about="#scheme">
    <dc:title>Morten Frederiksen's Categories</dc:title>
    <dc:description>Concepts from the weblog "Binary Relations" based on category usage by Morten Frederiksen.</dc:description>
    <dc:creator>Morten Frederiksen</dc:creator>
  </skos:ConceptScheme>

  <skos:Concept rdf:about="#c1">
    <skos:prefLabel>General</skos:prefLabel>
    <skos:narrower rdf:resource="#c23"/>
    <skos:narrower rdf:resource="#c30"/>
    <skos:inScheme rdf:resource="#scheme"/>
  </skos:Concept>

  <skos:Concept rdf:about="#c23">
    <skos:prefLabel>Travelling</skos:prefLabel>
    <skos:broader rdf:resource="#c1"/>
    <skos:inScheme rdf:resource="#scheme"/>
  </skos:Concept>

  <skos:Concept rdf:about="#c30">
    <skos:prefLabel>Politics</skos:prefLabel>
    <skos:broader rdf:resource="#c1"/>
    <skos:inScheme rdf:resource="#scheme"/>
  </skos:Concept>
</rdf:RDF>
```
consensus that node labels do not represent labels for concepts in their own right, therefore correctly modeling these constructs requires care. SKOS Core includes a number of grouping constructs (known as 'collections') for handling arrays and node labels. These constructs allow the generation of hierarchical displays that include node labels, while still maintaining integrity within the underlying network of semantic relationships between concepts.

Being an RDF vocabulary, SKOS Core may obviously be used in combination with other RDF vocabularies such as DCMI Metadata Terms (12) and FOAF (26). This is a feature that enhances the utility of SKOS Core as a potential standard representation framework for controlled vocabularies, because where SKOS Core does not go far enough to cover specific local requirements, additional classes/properties can be defined and used in combination with the parts of SKOS Core that are appropriate. Thus SKOS Core has a natural flexibility, which allows it to be used as a basis for maximising interoperability in situations where controlled vocabularies have been developed to different specifications.

Another feature that SKOS Core inherits from its basis in RDF is the natural extensibility mechanism provided by the sub-class and sub-property tools in RDF schema (6). Thus 'extensions' ('refinements') to SKOS Core may be defined by declaring and publishing classes and/or properties as sub-classes/sub-properties of SKOS Core classes/properties. To support this kind of extension/refinement, the properties of the SKOS Core Vocabulary are grouped into families: lexical labeling properties, symbolic labeling properties, documentation properties, semantic relation properties. Properties within these families are arranged in a property hierarchy, so you can extend at the appropriate level of semantics. So for example, if you have a particular requirement for more precise semantic relationships between concepts, declare properties that are sub-properties of the appropriate SKOS Core semantic relation property (skos:semanticRelation being the hierarchy root, skos:broaden skos:narrower and skos:related being the next level down). Again, this simple extensibility mechanism means that SKOS Core can be used as a basis for interoperability beyond a small set of extremely similar use cases.

**SKOS Core Management Policies**

Development of the SKOS Core Vocabulary was initiated by the Semantic Web Advanced Development for Europe (SWAD-Europe) project (22), an EU-IST project in the 5th Framework Programme. Previous work on RDF vocabularies for representing thesaurus content (27, 28) formed the basis for the SKOS Core work, as well as other work reviewed in (24). The initial scope of SKOS Core was the RDF expression of thesauri that conform to ISO 2788-1986 (20), however it became apparent that the scope could reasonably be expanded to cover less 'standard' thesauri, as well as other types of controlled vocabulary with some underlying conceptual basis.

In September 2004 the W3C Semantic
Web Best Practices and Deployment Working Group (8) assumed responsibility for the publication, development and maintenance of SKOS Core. At the time of writing the Working Group is about to publish First Public Working Draft editions of the SKOS Core Guide (1) and the SKOS Core Vocabulary Specification (2), which are currently 'editor's drafts' and as such have no status within the W3C Process (29). Further publication iterations are planned, to allow for continued development in response to a growing base of feedback and deployment experience.

SWAD-Europe sought to directly involve as many stakeholders and experts as possible in the development of SKOS Core, and to thereby build an interest community around SKOS Core that could provide relevant feedback, knowledge and experience. The goal was to ensure that SKOS Core represented genuine consensus within the community that were anticipated to be its primary users. To this end, all development work has been carried out in public via publicly archived mailing lists and wikis. The public-esw-thes@w3.org mailing list (30) is the primary development forum for SKOS Core, and is open to all.

There has also been a requirement to manage the development process to provide support for early implementers, without

```
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:skos="http://www.w3.org/2004/02/skos/core#"
 xmlns:dc="http://purl.org/dc/elements/1.1/"
 xml:base="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs/" >
 <skos:ConceptScheme
 rdf:about="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs">
 <dc:title>Physics and Astronomy Classification Scheme</dc:title>
 <dc:creator>American Institute of Physics</dc:creator>
 <skos:hasTopConcept rdf:resource="90."/>
 </skos:ConceptScheme>

 <skos:Concept rdf:about="90.">
 <skos:prefLabel>GEOPHYSICS, ASTRONOMY, AND ASTROPHYSICS</skos:prefLabel>
 <skos:narrower rdf:resource="91."/>
 <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs"/>
 </skos:Concept>

 <skos:Concept rdf:about="91.">
 <skos:prefLabel>Solid Earth physics</skos:prefLabel>
 <skos:broader rdf:resource="90."/>
 <skos:narrower rdf:resource="91.10.-v"/>
 <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs"/>
 </skos:Concept>

 <skos:Concept rdf:about="91.10.-v">
 <skos:prefLabel>Geodesy and gravity</skos:prefLabel>
 <skos:broader rdf:resource="91.10-Pp"/>
 <skos:narrower rdf:resource="91.10.Pp"/>
 <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs"/>
 </skos:Concept>

 <skos:Concept rdf:about="91.10.Pp">
 <skos:prefLabel>Gravimetric measurements and instruments</skos:prefLabel>
 <skos:broader rdf:resource="91.10.-v"/>
 <skos:inScheme rdf:resource="http://www.w3.org/2001/sw/Europe/reports/thes/ns/pacs"/>
 </skos:Concept>

</rdf:RDF>
```
compromising the possibility for further improvement. To establish clear expectations for potential users of the SKOS Core Vocabulary, a set of 'Policy Statements' have been drafted, for inclusion as part of the SKOS Core Vocabulary Specification (2). The current system of change management is an attempt to combine both the Friend Of A Friend (FOAF) (26) and Dublin Core (14) process/management models. The FOAF vocabulary assigns a 'term status' value to each term of the vocabulary, indicating the level of stability of that term, and therefore giving some indication as to the extent of change that may be expected. This allows different parts of the vocabulary to evolve and stabilise at different rates, a highly desirable feature. The SKOS Core change management model attempts to formalise the FOAF approach further, by relating the term status values to types of change allowed, using the classes of change outlined in (14). The definitions of the term status values, and the refinement of the change classes defined in (14), as they relate specifically to the development of RDF vocabularies, needs further study.

SKOS Core and Dublin Core

The most common usage scenario for SKOS Core is anticipated to be the description of concepts for use in subject-indexing of web documents. In this scenario a resource of type skos:Concept is used as the object of a statement involving the dc:subject property. To support this kind of usage, some extensions (refinements) to the dc:subject property have been defined as part of the SKOS Core Vocabulary.

The property skos:subject is a sub-property of dc:subject, and is intended to carry the same meaning as dc:subject, but with the range of the property restricted to resources of type skos:Concept (the range of the dc:subject property is completely unrestrained). The skos:subject property is also intended to support the inference rule [(?d skos:subject ?x) (?x skos:broadner ?y) implies (?d skos:subject ?y)]. That is, if x is a subject of d, and y is broader than x, then y is also a subject of d. This rule therefore supports a basic type of query expansion. The property skos:isSubjectOf is the inverse of skos:subject, and is included for convenience.

The property skos:primarySubject is a sub-property of skos:subject. This property allows you to assert the primary (i.e. principle, main) subject of a document or other type of resource, where the document has more than one subject. This is particularly useful for the RDF expression of metadata where resources have been categorised according to a classification scheme, as usually (but not always, so care must be taken - see (23)) the underlying meaning of a categorisation is an assertion about the primary subject of a document. The property skos:isPrimarySubjectOf is the inverse of skos:primarySubject, and is included for convenience.

Using resources of type skos:Concept with the dc:subject property (or some sub-property thereof) draws to attention an area where the alignment between SKOS Core and Dublin Core needs further study. This area involves the nature of the relationship between what SKOS Core calls a 'Concept Scheme' and what Dublin Core calls a 'Vocabulary Encoding Scheme'. As defined by (15) an 'Encoding Scheme provides contextual information or parsing rules that aid in the interpretation of a term value,' and 'Vocabulary Encoding Schemes indicate that the value is a term from a controlled vocabulary, such as the value "China - History" from the Library of Congress Subject Headings.' Compare this with a 'Concept Scheme' which is currently defined by (2) as 'a set of concepts, optionally including semantic relationships between concepts.' It should be noted that this definition is not finalised.

There are two issues here. Firstly, when used as in (13) no commitment has been made as to exactly what sort of thing a resource type e.g. dct:DDC actually is. Is it a concept? Or is it just a sort of placeholder for some structured data? Initial consensus from recent discussion on the DC-ARCHITECTURE mailing list supports the former, and that it is therefore appropriate for e.g. the class dct:DDC to be a sub-class of the class skos:Concept.

Secondly, the URI of a Vocabulary Encoding Scheme identifies both the scheme itself (i.e. 'the DDC') and the class of values which comprise the scheme. SKOS Core, however, makes a distinction between a concept scheme and the class of concepts that participate in the scheme, and implicitly recommends the allocation of different URIs to these two entities (if the latter is to be formally identified at all, which is not necessary). The
nature of the relationship between the two entities can be expressed via an OWL restriction, see code example on this page.

I.e. the class of concepts in the example.com concept scheme is equivalent to the class of resources that have the example.com concept scheme as the value of the skos:inScheme property. Whether or not the current definition of a 'concept scheme' is appropriate, and the distinction between a concept scheme and the class of concepts that participate in the scheme is useful or necessary needs further study.

**SKOS Core and RDFS/OWL**

The relationship between concepts as described in RDF using the SKOS Core Vocabulary, and classes, properties or individuals in an ontology as described in RDF using RDFS and OWL, is subtle.

Consensus is emerging that there is a layer of indirection between the modeling of concepts in RDF using the SKOS Core Vocabulary and an RDFS/OWL ontology. To appreciate this point, consider the following example. Resource A is described by the following graph (assuming standard prefixes) 

```
{:A a skos:Concept; skos:prefLabel 'King Henry VIII of England'.}.
```

Resource B is described by the following `{:B a foaf:Person; foaf:name 'Henry Tudor'}.`

Now consider the statement `{:A dct:modified '2005-04-28'}.` This statement should be interpreted as saying that the concept of King Henry VIII was modified on the date given. Modifying a concept (i.e. modifying its meaning) is a common event in the management of e.g. thesauri: perhaps the scope is refined or expanded. In this example perhaps the meaning of resource A was extended to cover the idea of the man Henry Tudor throughout his whole life, whereas previously it had just covered the idea of Henry Tudor as king of England.

Compare this to the statement `{:B dct:modified '1509-06-24'}` This statement asserts that the person whose name is Henry Tudor was modified on the date given. It is unclear what the ‘modification date’ of a person might mean (1509-06-24 is the date of Henry Tudor’s coronation as king), and this is perhaps a dubious example. However the point is made that RDF statements about resource A are statements about a concept, and RDF statements about resource B are statements about a person. Therefore there is a layer of indirection between resource A and resource B. This point has important implication for the mapping and merging of RDF graphs, and requires further study.

**Summary and Conclusions**

This paper has introduced the SKOS Core Vocabulary, with examples of its use for the RDF expression of four types of 'knowledge organisation system': a glossary, a taxonomy, a...
thesaurus and a classification scheme. SKOS Core is at an initial stage of publication, and is presented here in the hope that members of the Dublin Core community may offer feedback and experience relevant to its continued development. It is also intended that this paper form an initial basis for further work to clearly establish recommended patterns of usage where SKOS Core is used in combination with DCMI Metadata Terms.

References
24. B. M. Matthews and A. J. Miles. Review


30. Mail Archives for public-esw-thes@w3.org see <http://lists.w3.org/Archives/Public/public-esw-thes/>. 