Use of SKOS for encoding GeoSciML Concept Vocabularies.

Version 0.1

Title:
Use of SKOS for encoding GeoSciML Concept Vocabularies

Specification URLs:
This version:
Latest approved version:
No approved version

Creator:
Stephen M Richard, CGI IWG Concept Definition Task Group

Creation date:
11/27/09

Last revision date:
12/18/2009 8:55 AM

Status:
V. 0.1 Draft for public comment

Publisher:
CGI IWG Concept Definition Task Group

Description:
This document describes usage of SKOS RDF schema for encoding vocabularies for use in populating content in GeoSciML GML-XML documents.

Rights:
Public document

Identifier:
urn:cgi:document:CGI:GeoSciML_SKOS_vocabulary:text:application%2Fmsword:2003

Table of Contents

Introduction

The GeoSciML GML application schema includes numerous attributes that are specified using terminological values. The use of terms from a shared, controlled vocabulary to populate these attributes in interchange documents enables true semantic interoperability. In order to facilitate interoperability, the GeoSciML development team formed the Concept Definition Task Group to develop a collection of vocabularies that can be used to populate terminological attributes in GeoSciML documents. The vocabularies are provided for use with test bed service demonstrations, and as working examples of the kind of vocabularies necessary for interoperability.

Vocabularies for GeoSciML are concept vocabularies, designed by analyzing the property the vocabulary will specify to determine the component conceptual facets/dimensions, and the scope/extent of its concept space (Gardenfors, 2000). The concept space is then partitioned into regions each of which is labeled with a term in the vocabulary. The essence of the vocabulary is the definition of the concept space is covers, and the of the boundaries of the partitions in that space to which labels/names/terms are given. The purpose of terms in the vocabulary is to represent regions in a specific concept space.

Because concepts are inherently abstract mental constructs, their ‘embodiment’ in an information system is established by an identifier (a unique string that corresponds to the concept) and a text definition that conveys the semantics of the concept to a human reader. Future work may allow encoding of the definitions using a formal system (e.g. OWL) such that some machine-reasoning may be used for testing logical coherence and for query expansion within a vocabulary, and for concept similarity measurement and semantic mediation between different vocabularies. The move to using RDF-based SKOS is the first step in this direction, with the ‘isDefinedBy’ property providing a path for linkage between the language-based definitions we currently use and formal, computable definitions that are under development.

Focus on the concepts as the first order members of the vocabulary contrasts with the thesaurus/library indexing approach to vocabularies that make the terms/words first order members, with semantics loosely defined by scope notes and associations between words. The concept vocabularies defined here are meant to be quantifiers for property values, and their meaning is required to be immutable. If the definition of a concept changes, this may impact the scope of other concepts in the vocabulary. Thus, these vocabularies are versioned as a whole. This has practical implications for how various parts of the RDF schemas are used, as will be discussed below.

Concepts are identified by language-neutral tokens (currently URN's defined in the CGI URN namespace). These tokens can then be matched with words in any language for use by communities of users, which are encoded as labels for the concept with a language tag.

Simple Knowledge Organization System (SKOS)

SKOS is an RDF application for encoding vocabularies. The normative documents defining the schema are W3C SKOS reference (http://www.w3.org/TR/2009/REC-skos-reference-20090818/) and SKOS primer (http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/). SKOS was developed from the perspective of thesaurus developers. "Using SKOS, concepts can be identified using URIs, labeled with lexical strings in one or more natural languages, assigned notations (lexical codes), documented with various types of note, linked to other concepts and organized into informal hierarchies and association networks, aggregated into concept schemes, grouped into labeled and/or ordered collections, and mapped to concepts in other schemes." (SKOS reference, Aug 18, 2009, http://www.w3.org/TR/2009/REC-skos-reference-20090818/)

RDF is a graph-based knowledge representation scheme consisting of nodes that represent resources and links that represent semantic relationships between the resources. Resource is defined in the context of the semantic web for which RDF was designed. In this context a resource is an identifiable thing that serves some purpose. In the RDF encoding of these graphs, a resource may be represented by an identifier, or may be a literal value (string or number). Identifiers in RDF instance documents are strings that have some known mapping to the resource of interest.

A SKOS resource may have one or more labels, which are words in some language that people use to name the resource. The SKOS model allows for labels in different languages, and multiple labels in a single language. At most one label in each language may be specified as the preferred label for that language. A SKOS resource may have associated annotation elements, and 'notation' attributes which are identifiers for that resource in some scope. The UML model for SKOS (Figure 1, Figure 2, Figure 3, Figure 4) explicitly models the resource identifier that would be used in RDF encoding as well (identifier in *AbstractSKOSElement*). This is meant to be a globally unique identifier defined in an IETF identification scheme indicated by a prefix like *urn:* or *http:*. There should be some mechanism for resolving these identifiers to locate the identified resource.

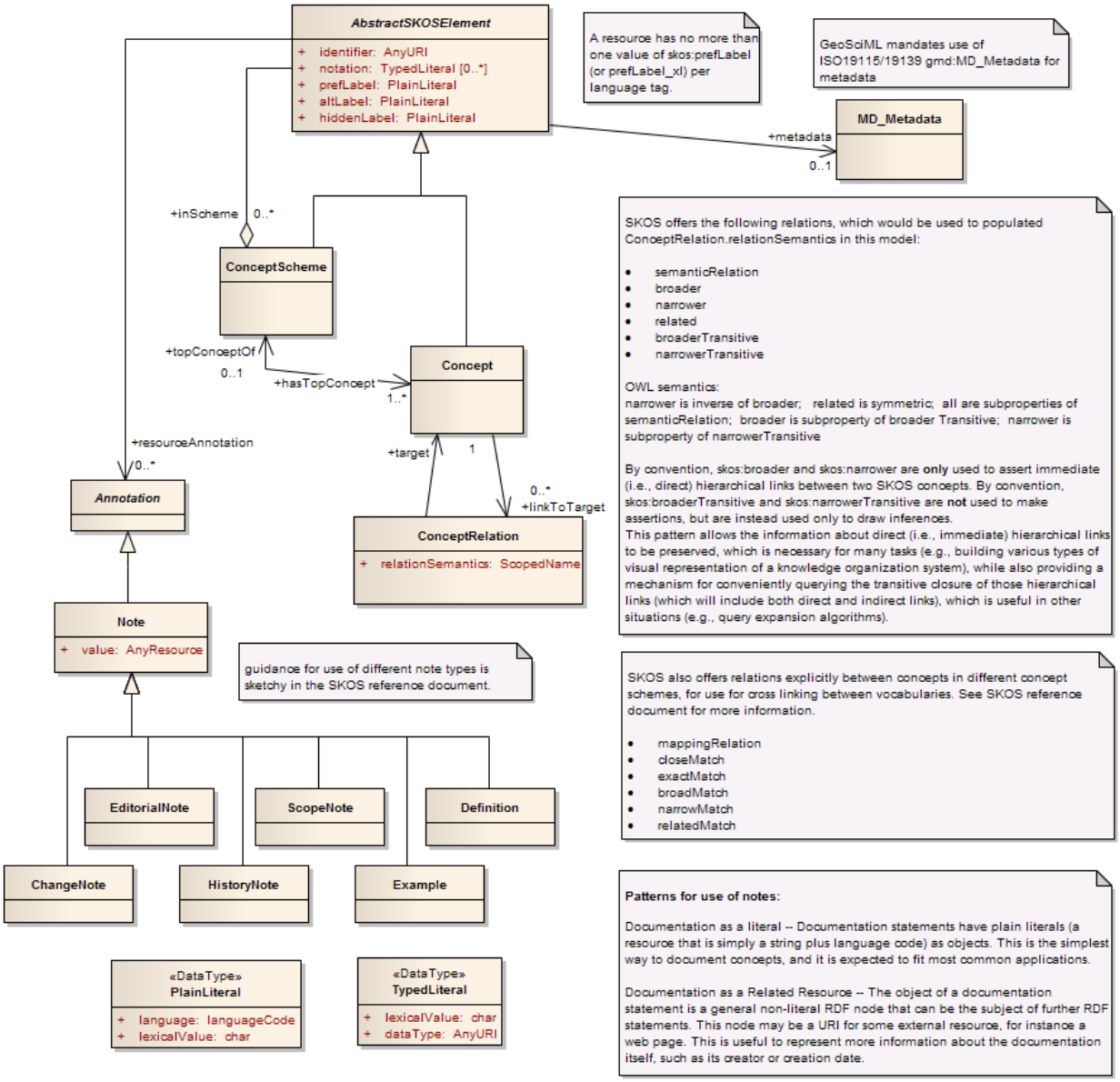


Figure 1. UML conceptual model for SKOS core elements.

skos:Concept is the core element in the representation of a vocabulary. A skos:Concept must have at least one preferred label. skos:Concepts may be grouped into skos:ConceptSchemes.

In the CGI vocabulary scheme, a skos:ConceptScheme is a collection of skos:Concepts defined within the same concept space. Operationally, this would mean that definitions of the concepts in the scheme are differentiated based on variation in the value of the same properties. In order to enable mapping from ConceptSchemes to formal ontology for use in reasoning software, the properties that distinguish members of a ConceptScheme should be defined in the definition of the ConceptScheme.

A ConceptScheme has one or more top concepts, which are concepts that do not have a broader (subsuming) concept included in the scheme.

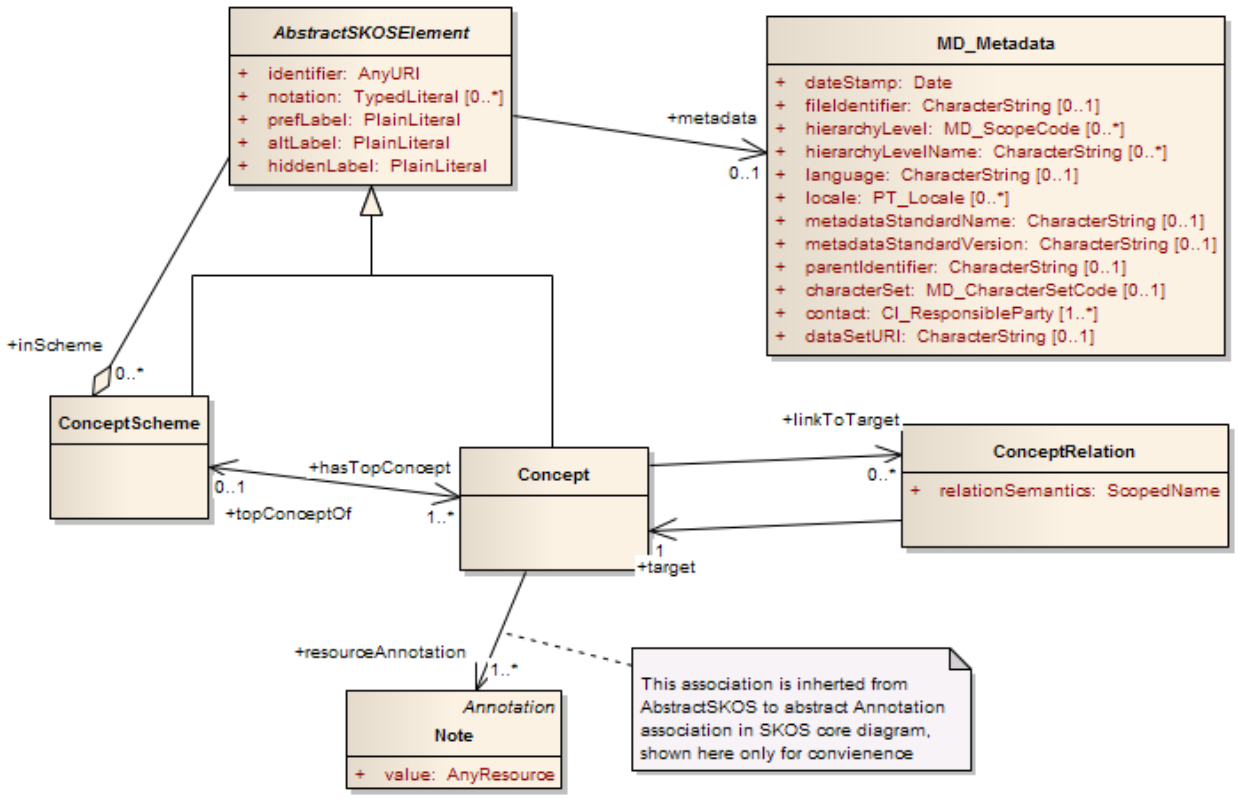


Figure 2. UML model context for SKOS Concept and ConceptScheme

Concepts may have relationships to other concepts. SKOS provides a collection of defined relationship types between concepts in the same ConceptScheme, based on practices in the thesaurus community. The most general is semanticRelation, which should not be used directly, but as a super-property for all properties denoting a relationship of meaning between concepts. Other relations between concepts defined in the SKOS reference (Aug 18, 2009) are broader, narrower, related, broaderTransitive, and narrowerTransitive. SKOS provides a parallel collection of Concept relationships for use to specify relationships between concepts in different ConceptSchemes. These include closeMatch, exactMatch, broadMatch, narrowMatch, and relatedMatch. The SKOS core RDF schema defines several other relations in addition to these; see Table 1for a summary of relations defined in the schema.

Table 1. Relations between concepts defined in the SKOS RDF schema.

SKOS Relation	Comment	GeoSciML Vocabulary relation notes (from NISO,2005, Z39.19-2005)
semanticRelation	This property should not be used directly, but as a super-property for all properties denoting a relationship of meaning between concepts.	
broaderTransitive		
broader	Broader concepts are typically rendered as parents in a concept hierarchy (tree).	link from a source term to a superordinate (more general or parent) target term. General hierarchical relationship, more specific subtype relationships (generic, instance, and whole-part) should be used whenever possible.
broaderMatch		
broaderGeneric	An extension of the 'broader' property to specify the class subsumption (sub-class/super-class) relationship between two concepts. This property is semantically equivalent to the 'rdfs:subClassOf' property.	This relationship identifies the link between a source class term and target members or species. This type of relationship is often called 'IsA'. A simple way to apply the test for validity described above is to formulate the statement '[narrower term] is a [broader term].' This relationship is also amenable to a logical "all-and-some" test. An example that passes this test is that some members of the class succulent plants are known as cacti and that all cacti, by definition and regardless of context, are succulent plants. An example that fails the test is that some members of the class 'desert plants' are known as 'cacti', some, but not all, 'cacti' are 'desert plants'. These terms should therefore be assigned to different hierarchies in the controlled vocabulary, and both terms should be assigned to the same content object when indexing a work on "cacti as desert plants.." Inverse relationship 'narrower generic term' is not included in this

SKOS Relation	Comment	GeoSciML Vocabulary relation notes (from NISO,2005, Z39.19-2005)
		vocabulary, use broader to make searches tractable.
broaderPartitive	An extension of the 'broader' property to specify a partitive (part of) relationship between two concepts.	This relationship covers situations in which one concept (source) is inherently included in another (target), regardless of context, so that the terms can be organized into logical hierarchies, with the whole treated as a broader term. This relationship can be applied to several types of term; the three types enumerated below are not intended to be exhaustive. For time interval concepts (ordinal eras) (Allen and Ferguson, 1994), this relationship subsumes 'During', 'Interval starts', and 'Interval finishes'. Examples: brain/central nervous system/nervous system; Ottawa/Ontario/Canada; regiment/battalion/military division/army; Cambrian/Paleozoic/Phanerozoic.
narrowerTransitive		GeoSciML vocabulary does not include inverse relationships for broader because these are symmetric
narrower	Narrower concepts are typically rendered as children in a concept hierarchy (tree).	
narrowMatch		
narrowerGeneric	This property is the inverse of the 'broaderGeneric' property.	
narrowerPartitive	This property is the inverse of the 'broaderPartitive' property.	
closeMatch		"Near-synonyms are terms whose meanings are generally regarded as different, but which are treated as equivalents for the purposes of a controlled vocabulary. The extent to which terms are treated as near-synonyms depends in large measure upon the domain covered by the controlled vocabulary and its size. Near-synonyms may include antonyms or represent points on a continuum.

SKOS Relation	Comment	GeoSciML Vocabulary relation notes (from NISO,2005, Z39.19-2005)
		Examples: sea water / salt water [variant terms] meteors / meteorites / meteoroids [points on a continuum] smoothness / roughness [antonyms] For each of these sets of near synonyms, a vocabulary developer might decide to designate one of the terms as the preferred term with the understanding that it will retrieve all content described by the other terms as well. As a general rule, terms should be treated as near-synonyms only in subject areas that are peripheral to the domain of the controlled vocabulary. When concepts can be distinguished in the controlled vocabulary domain with sufficient precision to justify their representation as separate terms, they should be individually defined and retained. If two concepts cannot be consistently and reliably differentiated from each other, however, a term for one concept should be selected as the preferred term and a USE reference made from the other. ,
exactMatch		"Synonyms are terms whose meanings are regarded as the same or nearly the same in a wide range of contexts. True synonyms are rare in natural language. Although the terms are interchangeable in many circumstances, usage can vary as a result of such factors as level of formality, professional vs. lay context, or pejorative vs. neutral vs. complimentary connotation."
relatedMatch		
related		"This relationship covers associations between terms that are neither equivalent nor hierarchical, yet the terms are semantically or conceptually associated to such an extent that the link between

SKOS Relation	Comment	GeoSciML Vocabulary relation notes (from NISO,2005, Z39.19-2005)
		them should be made explicit in the controlled vocabulary, on the grounds that it may suggest additional terms for use in indexing or retrieval. " "As a general guideline, whenever one term is used, the other should always be implied within the common frames of reference shared by the users of the controlled vocabulary. Moreover, one of the terms is often a necessary component in any explanation or definition of the other; the term cells, for example, forms a necessary part of the definition of cytology."
relatedHasPart	This property is the inverse of the 'relatedPartOf' property.	
relatedPartOf	An extension of the 'related' property. Use this property to express a partitive relationship between concepts, where it is desired that such a relationship be treated as associative (i.e. linking separate branches of a hierarchy) and NOT hierarchical.	

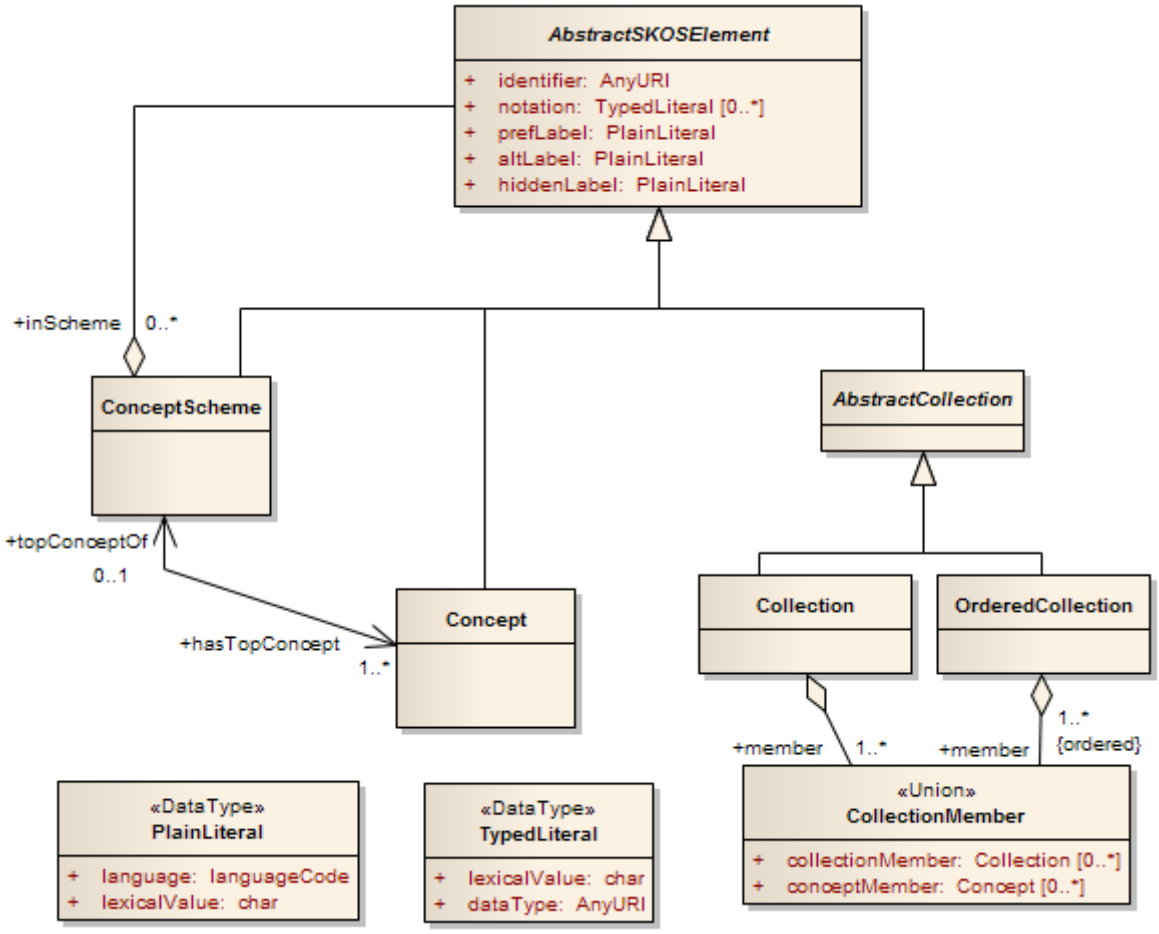


Figure 3. UML model for context of SKOS collection.

Labels

The SKOS core model provides three annotation properties for labeling SKOS elements--skos:prefLabel, skos:altLabel and skos:hiddenLabel. These are each instances of owl:AnnotationProperty, and also are each sub-properties of rdfs:label. As such, for SKOS vocabulary documents, they should be used instead of rdfs:label for assigning labels to elements. Each instance of a SKOS label has an RDF plain literal (string, plus language tag).

Skos:prefLabel assigns a preferred label/name for a concept specific to each language that the vocabulary uses. A resource may have no more than one value of skos:prefLabel per language tag. If a prefLabel is included for a concept in a particular language, labels in that language should be available for all concepts, and the language indicated by a dc:language element for the skos:ConceptScheme. As annotation properties for other SKOS elements, these labels do not have identity, and may not have other properties and relations assigned to them. This does not conflict with the GeoSciML GeologicVocabulary model, but such relationships are allowed by the BS8723 model.

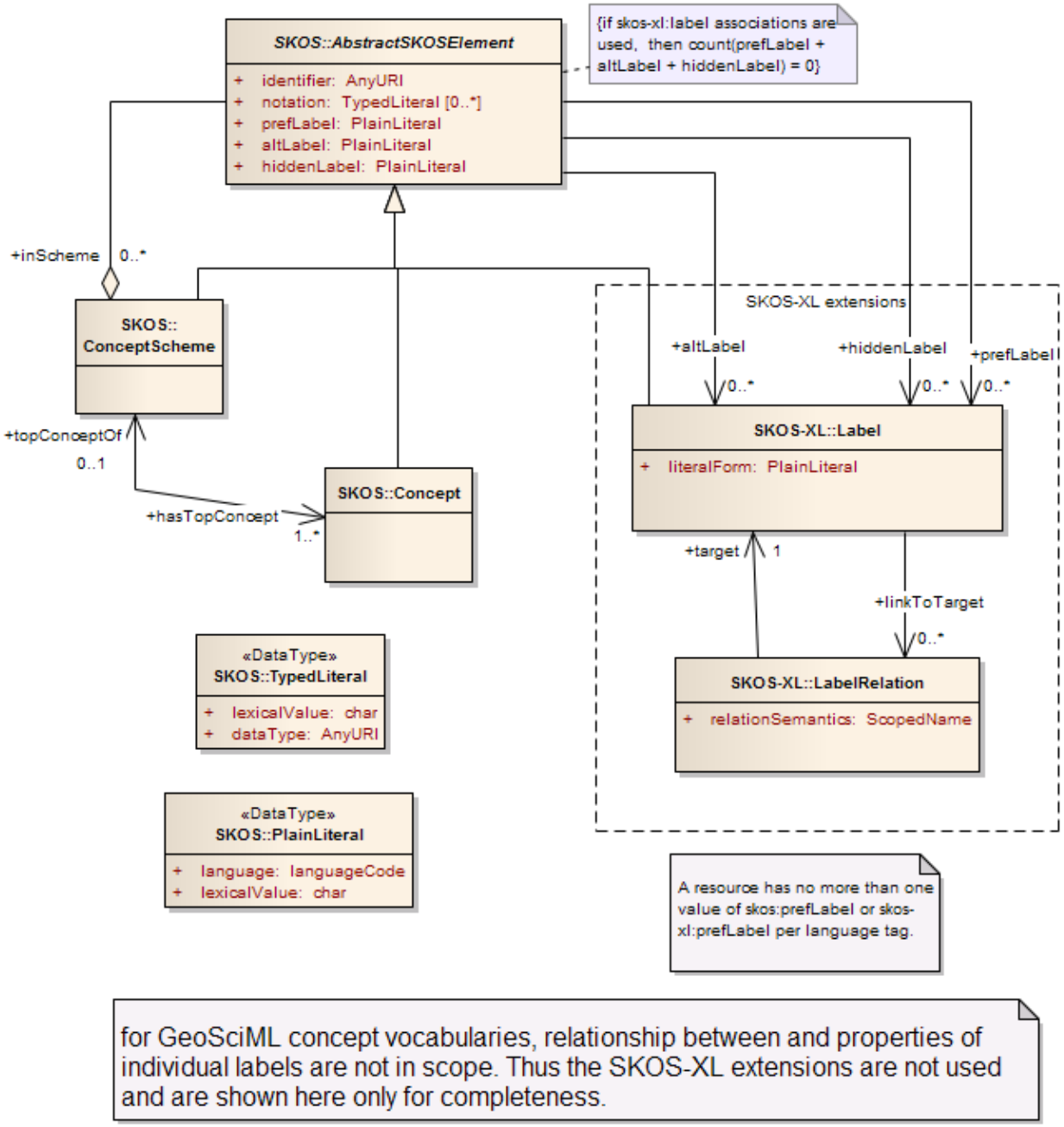


Figure 4. UML model for SKOS annoation. GeoSciML concept vocabularies do not used the SKOS-XL extension, but these are included in the diagram for completeness.

For representation of relationships (e.g. Use, Use for) between and properties (e.g. history and scope notes) of concept labels (ThesaurusTerms), the SKOS-XL extension has been defined. Because these thesaurus type use cases are out of scope for GeoSciML concept vocabularies, the SKOS-XL extensions are not used.

AltLabel

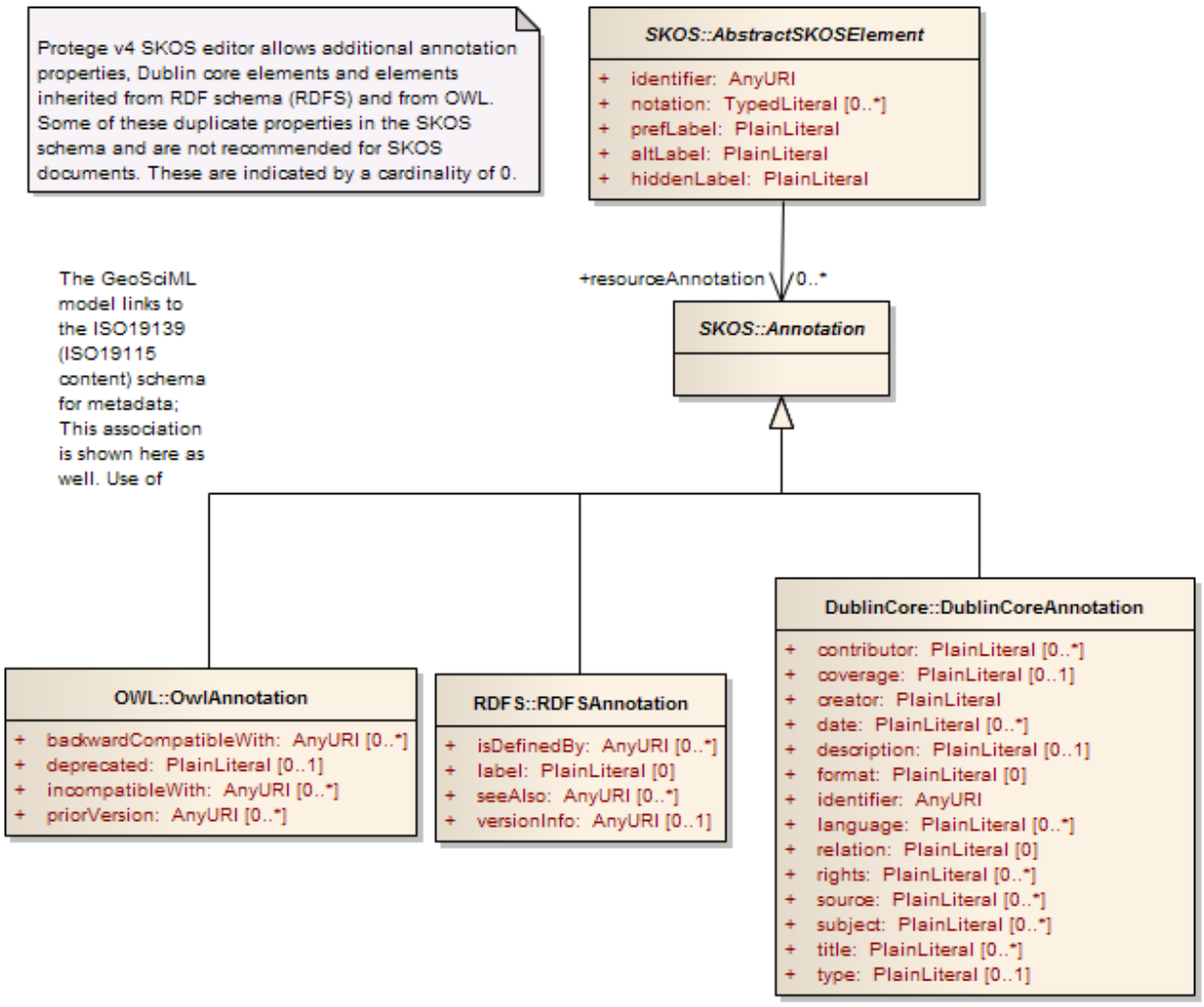
Use skos:altLabel for terms that are synonyms for the concept. Zero to many altLabels may be included in any language.

HiddenLabel

Use skos:hiddenLabel to include labels that should not be displayed, but might be used in search criteria, including common misspellings, deprecated, and non-recommended labels.

Documentation

The SKOS model includes a variety of different **Note** classes that may be used to document any resource. Of particular interest for the purposes of encoding vocabularies is the use of **Note** to define and document individual Concepts. Because SKOS is an RDF schema, it inherits some annotation properties from the RDF schema. Protégé, one of the standard Free, Open-Source tools used to browse and edit SKOS documents maps the SKOS elements into OWL, and then treats the RDF vocabulary document as an OWL document. This has the side effect of bringing in various annotation properties included in the OWL implementation.



OWL ontology header properties

backwardCompatibleWith

- DataType:AnyURI 0 to * owl:backwardCompatibleWith is a built-in OWL property with the class owl:Ontology as its domain and range. If a ConceptScheme is considered as an ontology, then this property could apply to ConceptSchemes. The owl:backwardCompatibleWith statement contains a reference to another ontology that is thus identified as a prior version of the containing ontology, and that all identifiers from the previous version have the same intended interpretations in the new version. Thus, it is a hint to document authors that they can safely change their documents to commit to the new version (by simply updating namespace declarations and owl:imports statements to refer to the URL of the new version). If owl:backwardCompatibleWith is not declared for two versions, then compatibility should not be assumed. NOTE: owl:backwardCompatibleWith is an instance of owl:OntologyProperty.

priorVersion

- DataType: AnyURI Range:0 to * An owl:priorVersion statement contains a reference to another ontology that is thus identified as a prior version of the containing ontology. This may be used by software to organize ontologies by versions. owl:priorVersion is a built-in OWL property with the class owl:Ontology as its domain and range.

incompatibleWith

- DataType: AnyURI 0 to * An owl:incompatibleWith statement contains a reference to another ontology that is thereby identified as an earlier version of the ontology that is not compatible with the referencing, containing ontology. Essentially, this is for use by ontology authors who want to be explicit that documents cannot upgrade to use the new version without checking whether changes are required. owl:incompatibleWith is a built-in OWL property with the class owl:Ontology as its domain and range.

Deprecated

- DataType: PlainLiteral 0 to 1 Deprecation is a feature commonly used in versioning software (for example, see the Java programming language) to indicate that a particular feature is preserved for backward-compatibility purposes, but may be phased out in the future. In the context of SKOS, a skos:Concept may be declared to be deprecated by assigning a value of 'true' to this property. Any other value should be ignored. Documentation for use of this annotation property as presented by the Protégé SKOS extension is lacking, and it is not obvious how this relates to the OWL version 1 deprecatedClass. By deprecating a term, it means that the term should not be used in new documents that commit to the ontology. This allows a ConceptScheme to maintain backward-compatibility while phasing out an old vocabulary (thus, it only makes sense to use deprecation in combination with backward compatibility). The idea is to make it easier for old data and applications to migrate to a new version, and increasing the level of adoption of the new version.

RDF schema annotation properties

isDefinedBy

- DataType: AnyURI Range:0 to * URI of a resource that defines the concept; may be image, text, drawing, OWL file, a physical prototype... rdfs:isDefinedBy is an instance of rdf:Property that is used to indicate a resource defining the subject resource. This property may be used to indicate an RDF vocabulary in which a resource is described. A triple of the form: "S rdfs:isDefinedBy O" states that the resource O defines S. It may be possible to retrieve representations of O from the Web, but this is not required. When such representations may be retrieved, no constraints are placed on the format of those representations. rdfs:isDefinedBy is a subproperty of rdfs:seeAlso. The rdfs:domain of rdfs:isDefinedBy is rdfs:Resource. The rdfs:range of rdfs:isDefinedBy is rdfs:Resource. Rdfs:isDefinedBy corresponds to the GeoSciML prototype association from a gsml:ControlledConcept to a resource that defines the concept.

label

- DataType: PlainLiteral Range:0 to 0 rdfs:label is an instance of rdf:Property that may be used to provide a human-readable version of a resource's name. Do not use for SKOS vo-

cabularies, use SKOS labels instead for consistency. SKOS labels (prefLabel, altLabel, hiddenLabel) are sub-properties of rdfs:label.

versionInfo

- DataType: AnyURI Range: 0 to 1 An owl:versionInfo statement generally has as its object a string giving information about this version, for example RCS/CVS keywords. Although this property is typically used to make statements about ontologies, it may be applied to any OWL construct. For example, one could attach a owl:versionInfo statement to an individual skos:Concept since it is an instance of owl:class. The current thinking for CGI CDTG vocabularies is that vocabularies will be versioned at the ConceptScheme level, with the intention that any change in the definition of an individual concept within the vocabulary is likely to have side effects on usage of other concepts in the vocabulary. Updates to wording, addition of new labels or other changes that do not effect the semantics of the concept do not require assignment to a new version. Thus, the owl:versionInfo is optional, and if included should specify the URI for the version of the ConceptScheme.

seeAlso

- DataType: AnyURI Range: 0 to * URI for a resource that provides some additional information, or association with related terms that are not included in the vocabulary. rdfs:seeAlso is an instance of rdf:Property that is used to indicate a resource that might provide additional information about the subject resource. A triple of the form "S rdfs:seeAlso O" states that the resource O may provide additional information about S. It may be possible to retrieve representations of O from the Web, but this is not required. When such representations may be retrieved, no constraints are placed on the format of those representations. The rdfs:domain of rdfs:seeAlso is rdfs:Resource. The rdfs:range of rdfs:seeAlso is rdfs:Resource. In the context of vocabularies, seeAlso links should be to images or documents that provide additional resources for understanding a concept, but are not normative (in which case the rdfs:isDefinedBy element should be used). Semantic associations should be indicated by skos:related links.
  In the GeoSciML vocabularies, rdfs:seeAlso is also used to list informal terms that are related to the skos:Concept in common usage, but the term itself does not label a concept in the vocabulary, so skos:related is not appropriate. Rdfs:seeAlso is used for terms that are not exactly synonymous with the concept. For synonyms, use skos:altLabel. Rdfs:seeAlso should be used to associate common-usage geologic terms with the more formal labels assigned in the vocabulary when those terms are not exactly equivalent to the vocabulary concept.

SKOS note types

These note types are introduced in the SKOS reference

EditorialNote

The SKOS Primer provides the following guidance: "supplies information that is an aid to administrative housekeeping, such as reminders of editorial work still to be done, or warnings in the event that future editorial changes might be made:

ex:doubleclick skos:editorialNote 'Review this term after company merger complete'@en.
ex:folksonomy skos:editorialNote 'Check spelling with Thomas Vander Wal'@en."

Webster defines editorial thus: 1. “of or relating to an editor or editing <an editorial office>”, or 2 : “being or resembling an editorial <an editorial statement>”.

It is apparent that use cases for EditorialNote have to do with development of a vocabulary, for internal use in the development workgroup; these notes should be removed in release versions. No mapping to GeoSciML GeologicVocabulary element.

ScopeNote

Scope notes are used to explain and provide advice on usage of a term (rdfs:label, BS8723:ThesaurusTerm) to indexers or searchers. Its usefulness in GeoSciML concept vocabularies is limited, particularly because the SKOS-XL extensions are not used. The meaning of a term is given by the definition of the concept that the term labels. In order to avoid confusion about where to put content, any clarification of the distinction of a concept from other related concepts should be included in the definition element. This SKOS element would best be mapped to ControlledConcept.description in the GeoSciML GeologicVocabulary model.

HistoryNote

According to NISO Z39-19-2005, ‘A History Note is used to track the development of terms over time’. Note that this use case applies to what are rdfs:labels or BS8723 ‘ThesaurusTerms’. In library indexing and classification, history notes provide guidance for terms used over a long period of time, during which the associated topic (concept) may have changed. NISO Z39-19-2005 recommends use of HistoryNotes for recording date of creation and updates to terms. Because the current management system for GeoSciML concept vocabularies assigns version identifiers to an entire vocabulary, which all concepts in the vocabulary inherit; the only modifications that are made once a vocabulary has been released and assigned a version identifier are bug fixes (spelling or grammatical corrections, repair of erroneous or missing relationships). For applications conforming to this profile, the dc:date value is assumed to be the creation date for a resource. HistoryNotes for a skos:ConceptScheme should be used to record other events in the lifecycle of the vocabulary (see Table 2)

Table 2. Dublin Core qualifiers for dates (http://dublincore.org/documents/2005/11/07/usageguide/-qualifiers.shtml) that may be applicable to GeoSciML concept vocabularies.

Event	Scope notes
Valid	If the resource is only valid or relevant for a particular date or range of dates, the term Valid may be used to express those dates. This may be particularly important if the resource will be retained over time but its use is valid only during a particular period or until a particular date.
Available	In general, the term Available should be used in the case of a resource for which the date of availability may be distinct from the date of creation, and the date of availability is relevant to the use of the resource.
Issued	The term Issued should be applied when a formal date of issuance or publication is relevant to the resource, and is distinct from other dates that may be used with the resource.
dateAccepted	Date of acceptance of the resource (e.g. of thesis by university department, of article by journal, etc.). If, in the lifecycle of a resource, the date of acceptance by a formal body or entity is relevant to the use of the resource, dateAccepted may be used.

dateSubmitted	If, in the lifecycle of a resource, the date of submission to a body or entity is relevant to the use of the resource, dateSubmitted may be used.

Bug fixes to individual concepts in a ConceptScheme should be recorded using skos:ChangeNote (see below). Any such information pertaining to specific labels for a concept would have to be included in the change note for the concept, because the SKOS-XL extensions are not used in this profile.

HistoryNotes may also be used during the development of a vocabulary to document modifications to a concept or other element in a vocabulary, but these should be removed or summarized in the release version. In the gsml:GeologicVocabulary, these would map to the ISO19139 MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep

ChangeNote

Change notes are dated annotations that describe bug fixes to an element in a vocabulary. See discussion above for HistoryNote. In the gsml:GeologicVocabulary, these would map to the ISO19139 MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage/LI_Lineage/processStep

Definition

A complete explanation of the intended meaning of a concept, including any clarification of what is or is not included within the meaning of a concept. The definition establishes the identity of a concept by distinguishing it from other concepts in a vocabulary. Ultimately, equivalence between concepts is established by logical equivalence of their definitions, irrespective of the labels associated with the concept or language used to express the definition. This SKOS element is equivalent to ControlledConcept.description in the GeoSciML GeologicVocabulary model.

Example

Example notes may be used to list instances or subconcepts of a concept to help understand the meaning. There is no 1:1 mapping to an element in the gsml:GeologicVocabulary model. Example type content would best be placed in the ControlledConcept.description.

Metadata elements

Tools implemented to manage SKOS vocabularies commonly allow for association of Dublin Core metadata elements with vocabulary elements. This section includes some recommendations for use of the Dublin Core elements in the context of GeoSciML Concept Vocabularies. Note that the GeoSciML model associates ISO19115 metadata (gmd:MD_Metadata)with concepts. The Dublin Core elements are a subset of the content in the ISO19115 model; corresponding ISO19115 elements are indicated in the following discussion.

Metadata elements that are associated with SKOS elements are meant to provide information detailing the origin of the content of the vocabulary. This content pertains more to the evaluation of the content for fitness for a users requirements that to the discovery use cases to which some of the Dublin Core elements are more applicable.

Contributor

An agent responsible for making contributions to the resource, may be a person, an organization, or a service. For CGI vocabularies, the Contributor element on the ConceptSchema may be used to identify the members of the task group that fills the Creator role, but it would be re-

dundant to include this information with each concept. Contributor is the most general of the elements used to specify "agents" responsible for the resource, so may be used to identify other agents that participated in vocabulary development if desired.

ISO19139 path: MD_Metadata.identificationInfo. MD_DataIdentification.-pointOfContact.CI_ResponsibleParty.*$name$*[role.CI_RoleCode@codeListValue='*any value except originator or author*'], where *$name$* may be individualName, organizationName, or positionName. Note that in general, the informality of the dc:Contributor field content will make it impossible to automatically map from dc:Contributor to the equivalent ISO19139 xml.

Coverage

The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. In general this element will not be applicable to a GeoSciML concept vocabulary, unless it is a stratigraphic lexicon.

ISO19139 path for spatial coverage by bounding box: compound mapping; MD_Metadata.identificationInfo.MD_DataIdentification.extent.EX_Extent.geographicElement.-EX_GeographicBoundingBox.westBoundLongitude, same prefix path for southBoundLatitude, eastBoundLongitude, northBoundLatitude

Creator

The agent or agents primarily responsible for making the resource, may be a person, an organization, or a service. For CGI concept vocabularies, the creator role will typically be assigned to a task group that produced the vocabulary. Currently, concept vocabularies are produced by the Interoperability Working Group Concept Definition Task Group. The CGI Multilingual Thesaurus Group is developing thesaurus type vocabularies for which that group would fill the Creator role. A creator should be specified.

MD_Metadata.identificationInfo. MD_DataIdentification.-pointOfContact.CI_ResponsibleParty.*$name$*[role.CI_RoleCode@codeListValue='author'] or -[role.CI_RoleCode@codeListValue='originator'], where *$name$* may be individualName, organizationName, or positionName. Note that in general, the informality of the dc:Creator field content will make it impossible to automatically map from dc:Creator to the equivalent ISO19139 xml.

Date

For GeoSciML Concept Vocabularies, the dc:date should be the release date of the vocabulary (ConceptScheme). Dates of other significant events in the development of the ConceptScheme should be recorded in skos:historyNote. The date should specify the year, month and day of release. All concepts in a ConceptScheme will have the same dc:date. If some concepts have been updated with bug fixes, the date is noted in a skos:changeNote. Date for this profile should use xs:date data type, defined thus "date uses the date/timeSevenPropertyModel, with *·hour·*, *·minute·*, and *·second·* required to be ***absent***. *·timezoneOffset·* remains ·optional" (http://www.w3.org/TR/xmlschema11-2). Example date encoding: 2000-12-12+13:00, 2006-10-01. If the month or day is not known, encode as '00', for example '2006-00-00'.

ISO19139 path: Since this date that applies to the described resource (not the metadata record) use MD_Metadata. identificationInfo/ MD_DataIdentification/ citation/CI_Citation/ date/CI_Date/date/ with dateType code='publication'.

Description

Only used for Collection or ConceptScheme to describe the content and intention of the collection or ConceptScheme.

ISO19139 path: MD_Metadata.identificationInfo.MD_DataIdentification.abstract

Format

The file format of the resource, which in the case of this application is always 'application/rdf+xml' (http://www.ietf.org/rfc/rfc3870.txt), and should only be included for the ConceptScheme element, optional elsewhere.

ISO19139 path: MD_Metadata.distributionInfo.MD_Distribution.distributionFormat.-MD_Format.name

Identifier

An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string conforming to a formal identification system. The GeoSciML Workgroup has been using URN's in the CGI namespace (http://www.faqs.org/rfcs/rfc5138.html)

ISO19139 path: MD_Metadata.datasetURI

Language

Language of the intellectual content of the catalog record. Individual, language-localized SKOS content (PlainLiteral data type in the UML model) uses the RDF @xx suffix to denote the language. One or more language elements should be included in the ConceptScheme element, listing all languages supported by the concept scheme. All PlainLiteral content in the document should be available in all languages listed in this element. Use ISO639-2/T language code (lower case) and a three character ISO3166-1 country code string (upper case) to indicate the language and country localization. Syntax is <language code>;<blank space><country code>. Example "fra; CAN".

ISO19139 path: MD_Metadata.identificationInfo. MD_DataIdentification.language

Relation

Not used in this profile. Use skos:related.

Rights

Information about rights held in and over the resource. Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.

ISO19139 path: MD_Metadata.identificationInfo.MD_DataIdentification.-resourceConstraints.useLimitation

Source

A related resource from which the described resource is derived. The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.

ISO19139 path: MD_Metadata.identificationInfo.MD_DataIdentification.citation.-CI_Citation.title. Note that the informality of the dc:Source element precludes accurate mapping to ISO19139 elements.

Subject

Subject key words may be included to assist discovery of a vocabulary document, but in general this use case is more appropriate to a separate metadata document that describes the vocabulary. Optional.

ISO19139 path: MD_Metadata.identificationInfo. MD_DataIdentification.-descriptiveKeywords.MD_Keywords.keyword. The MD_Metadata.identificationInfo.-MD_DataIdentification.topicCategory that would go with the MD_Keyword element is not formally included in dc:Subject, and would probably need to be indicated as 'missing'.

Title

The name by which the resource is formally known. Redundant with skos:prefLabel for concepts and is optional. Mandatory for ConceptScheme, used to indicate the name by which the ConceptScheme should be known

MD_Metadata.identificationInfo.AbstractMD_Identification.citation.CI_Citation.title

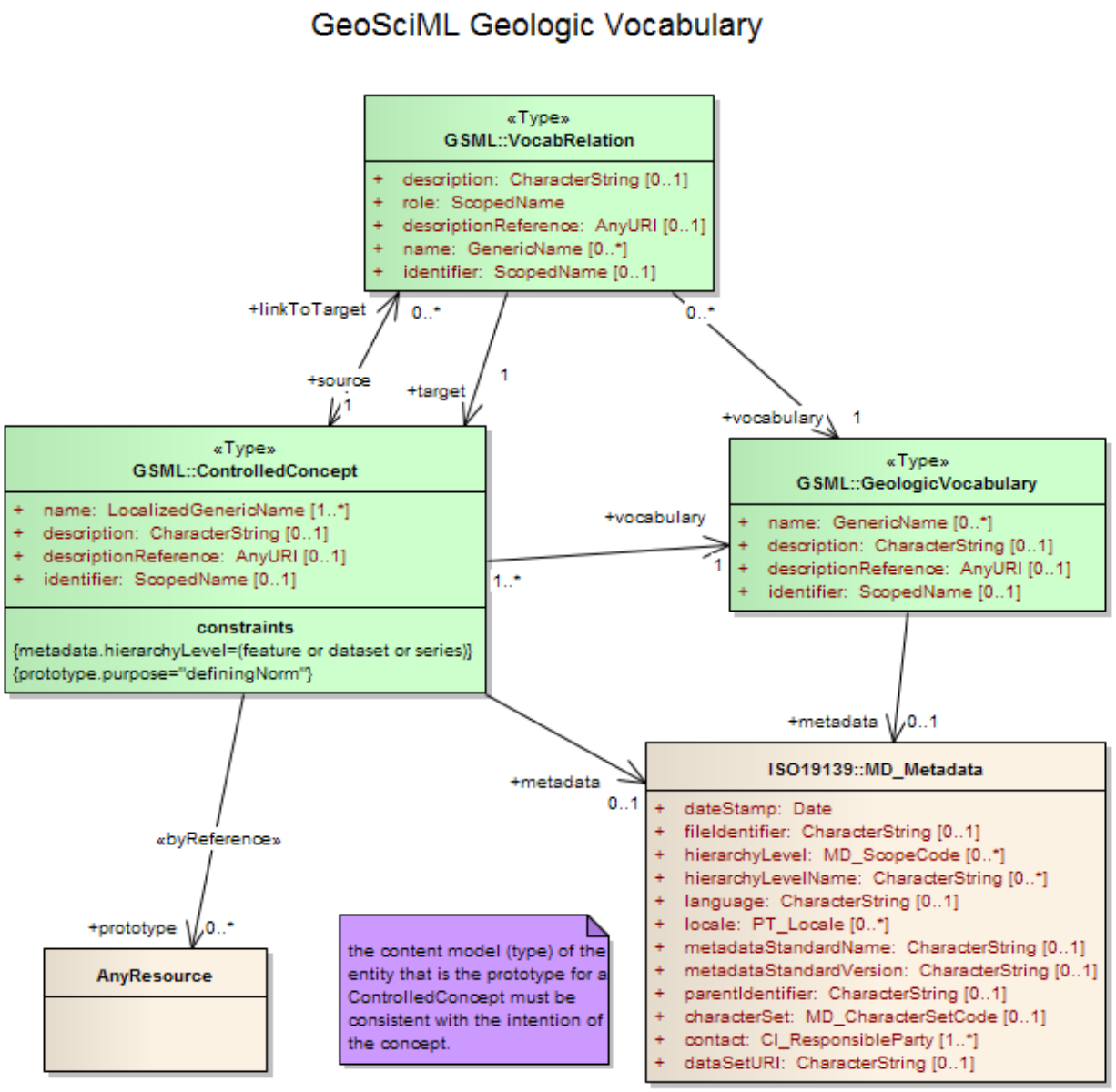
Type

The nature or genre of the resource. DCMI recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary, but all vocabulary elements will be of type dcmi:Text, so there is little point including this element. It is considered optional.

Mapping to GeoSciML GeologicVocabulary conceptual model

The GeoSciML Geologic Vocabulary model includes three GML object types—ControlledConcept, VocabRelation, and GeologicVocabulary. ControlledConcept represents a domain concept. A ControlledConcept has at least one name, and may have many names; these are used to convey the concept to people. A ControlledConcept may have a description, which is intended to be a human readable definition of the concept. A descripitonReference property is also inherited from AbstractGML, but is not used in GeoSciML. The GeoSciML model adds an identifier property for a concept that is intended to be a globally unique, language neutral identifier for the concept. The intention of the design group is that this property will be superseded by gml:identifier when the application migrated to GML v. 3.2. The data type for the name property is LocalizedGenericName, which is a string with a language code. A GeologicVocabulary is a collection of ControlledConcepts and VocabRelations. A VocabRelation is a semantic relationship between two concepts, with the semantics of the relationship encoded in the role property. GeologicVocabulary and VocabRelation inherit name (GenericName, not language localized), description, and descriptionReference from abstractGML. A ControlledConcept has an associa-

tion with one or more resources in a prototype role that serve to define the concept; these resources may be of any type—text documents, drawings, images, other xml, rdf, owl files, etc. ISO 19115 metadata may be associated with GeologicVocabulary and ControlledConcept through the metadata role.



SKOS element (see Figure 2**)**	**GeoSciML GeologicVocabulary element**	**Comment**
Concept	ControlledConcept	

SKOS element (see Figure 2**)**	**GeoSciML GeologicVocabulary element**	**Comment**
Concept.linkToTarget	ControlledConcept.linkToTarget	skos:ConceptRelation (Figure 2) is equivalent to gsml:VocabRelation; gsml:role becomes a skos rdf predicate linking a source concept to a related target concept.
Concept.inScheme	ControlledConcept.vocabulary	
Concept.topConceptOf	No equivalent	Association from concept to containing ConceptScheme specifying the root (top level) concepts in the scheme, not present in GeoSciML model.
Concept.resourceAnnotation.definition	ControlledConcept.description	
Concept.resourceAnnotation.ChangeNote, HistoryNote	ControlledConcept.metadata.MD_Metadata/../lineage/../processStep	Currently know GeoSciML GeologicVocabularies do not include this kind of content.
Concept.isDefinedBy	ControlledConcept.prototype	
ConceptRelation	VocabRelation	GeoSciML soft-types relationships semantics using the VocabularyRelationship vocabulary. SKOS hard-types relationship using rdf predicates.
AbstractSKOSElement.identifier	ControlledConcept.identifier GeologicVocabulary.identifier	
No equivalent	VocabRelation.identifier	SKOS does not reify concept relationships

SKOS element (see Figure 2**)**	**GeoSciML GeologicVocabulary element**	**Comment**
AbstractSKOSElement.altLabel	ControlledConcept.name GeologicVocabulary.name	GeoSciML does not identify preferred labels. Mapping rule would be if a single name is included in a particular language, it is assumed to be the preferred label; if multiple names in a particular language are included, they are all treated as altLabels.
AbstractSKOSElement.prefLabel	No equivalent	
AbstractSKOSElement.hiddenLabel	No equivalent	
AbstractSKOSElement.metadata	ControlledConcept.metadata GeologicVocabulary.-metadata	
AbstractSKOSElement.resource-Annotation.OwlAnnotation, RDFS-Annotation, DublinCoreAnnotation	ControlledConcept.metadata GeologicVocabulary.metadata	Mapping of Dublin Core elements to ISO is discussed in text. Similar mapping could be developed for OWL and RDFS annotation properties.

References

ANSI/NISO, 2005, Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies: ANSI/NISO Z39.19-2005, ISBN 1-880123-65-3, available at http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a.

Issues:

Use of altLabel vs. seeAlso, vs. creating related terms?? (e.g. marl, marlstone)

Use search thesaurus

How to find dikes